Predicting Electric Vehicle Charging Station Occupancy: Model Evaluation and Usability Insights

By

Adel Ali Mohamed Dadaa

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor.

of Philosophy in Computer Science

Department of Computer and Information Science

University of Strathclyde

2025

Declaration

‘This thesis is the result of the author’s original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.’

Signed:

Date: 05/09/2025

Acknowledgement

In the name of Allah, the Most Gracious, the Most Merciful.

First and foremost, I thank Allah for granting me the strength, determination, and patience to complete this thesis. Without His blessings, none of this would have been possible. I extend my deepest gratitude to my supervisor, Dr. Mark Dunlop, for his unwavering guidance, patience, and dedication throughout the years of my PhD journey. Your insightful advice and continuous support have been instrumental in helping me navigate the challenges of this research and achieve my goals. I also wish to express my sincere thanks to Dr. John Welson, my second supervisor, for the valuable support and encouragement provided during the early stages of my PhD.

I am profoundly grateful to my beloved family (my wife and my children) for their boundless love, patience, and support throughout this journey. Your understanding and sacrifices have been a source of strength and inspiration for me. I also extend my heartfelt thanks to my family in Libya for their prayers, encouragement, and constant belief in me.

I am deeply thankful to my friends and colleagues, both here in the UK and back home in Libya, for their companionship, guidance, and shared wisdom during this academic pursuit. Your presence has been a vital part of my personal and professional growth.

A special thanks goes to the Ministry of Education of Libya and Mesellata High Institution for granting me this invaluable opportunity to pursue my PhD. Your support has been crucial in enabling me to advance my education and contribute to the academic and professional development of our community.

Abstract

Electric vehicles are increasingly adopted as an environmental complement to internal combustion vehicles through incentives provided by governments worldwide. However, despite rapid growth in certain market segments, concerns over range and the availability of charging infrastructure thwart further acceptance among the general public. A key challenge is the uncertainty around the unavailability of public charging stations, which in turn increases "range anxiety" for EV drivers. This thesis addresses the challenge of deep learning models' performance in EV charging stations' availability forecasting.

A preliminary analysis of the charging behaviours by historical data samples and the quantitative survey with EV drivers was performed in order to extract the essential temporal and environmental factors affecting the station occupancy. This insight was used in investigating performances of several deep learning architectures for classification and regression tasks, particularly Long-Short Term Memory (LSTM), Gated Recurrent Unit (GRU), Temporal Convolutional Network (TCN), 1-Dimensional Convolutional Network (1D-CNN), and an Ensemble model (comprising 1D-CNN and TCN).

The experimental results revealed that classification models can offer superior occupancy status prediction performance compared to regression models, and the optimum performance is obtained using 1D-CNN and TCN. A new hybrid architecture combining TCN and bidirectional GRU (BiGRU) achieved enhanced classification accuracy and generalized the performance of the baseline models. A complementary quantitative and then qualitative investigation were also conducted to establish the preferences of electric vehicle owners concerning charging forecasting, and their level of trust in model forecasting.

This research enhances the management of EV charging infrastructure by delivering precise and user-centric occupancy predictions. By mitigating range anxiety and increasing confidence in predictive systems, it facilitates informed decision-making, enhances the charging experience, and encourages broader worldwide use of electric vehicles.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

In recent years, the global transition from conventional internal combustion engine vehicles to electric vehicles (EVs) has brought pressing challenges and opportunities in designing, implementing, and managing charging infrastructure. This chapter provides the foundation for understanding these issues by discussing the evolving landscape of Electric Vehicle Infrastructure (EVI), the prevalence of "range anxiety" among EV owners, and the importance of robust modelling techniques in optimizing charging processes. It then outlines the research objectives, ranging from model comparisons to usability insights, and states the key questions that drive this investigation. Finally, it presents the organization of the thesis, providing a roadmap of how subsequent chapters reach deeper into addressing the critical concerns surrounding EV adoption and infrastructure planning.

1.1. Study Context

The whole automotive industry is seeing a swift transition towards the adoption of EVs. The convergence of environmental concerns, technological advancements, and government policies has catalysed the transition from traditional internal combustion vehicles to their electric counterparts. This transformation is crucial for combating climate change, diminishing greenhouse gas emissions, and facilitating sustainable urban development. There has been increasingly substantial global support for EVs in recent years. Recent estimates indicate that over 10 million electric vehicles were on the roads worldwide by the end of 2020, representing a (43%) rise from the preceding year. This increase may continue to persist, as numerous countries endeavour to diminish carbon emissions and invest in electric vehicle infrastructure.

The United Kingdom is swiftly advancing its electric vehicle industry. As of the conclusion of November 2024, according to Zapmap[1], there are over 1,300,000 totally

[1] *https://www.zap-map.com/ev-stats/ev-market*

electric vehicles on UK roads, constituting around 3.94% of the total 34 million cars. This signifies substantial growth since the conclusion of 2020, when there were merely 205,770 electric vehicles, or only 0.6% of the total vehicle population. During this era, the quantity of electric vehicles has risen by more than five times. Consequently, heightened consumer acceptance is being bolstered by financial incentives for the acquisition of electric vehicles and the expansion of EV charging infrastructure (EVCI). EVs will supplant the reduction of environmental emissions, particularly in the transportation sector, which is a significant contributor to air and greenhouse gas pollution. A transition from conventional fuel-powered vehicles to electric ones will aid in diminishing nitrogen oxides (NOx) and particulate matter in metropolitan areas, thereby enhancing air quality and safeguarding public health (El Hafdaoui *et al.,* 2023).

Furthermore, EVs play a crucial role in the advancement of intelligent urban environments in the context of emerging smart technologies and the Internet of Things (IoT). EVs can engage effortlessly with smart grid systems, facilitating effective energy management and the incorporation of renewable energy resources. The two-way energy exchange that occurs between EVs and the grid, referred to as vehicle-to-grid (V2G) technology, enhances grid stability and enables improved energy allocation (Casolaro et al., 2023). In the UK, V2G technology is still emerging but has seen active experimentation through several pilot projects such as Project Sciurus[2], one of the largest residential V2G trials in the world, which involved over 300 homes using Nissan EVs and compatible chargers. Although not yet mainstream, these developments demonstrate growing national interest and investment in V2G integration within the UK's smart energy infrastructure.

1.2. Electric Vehicle Infrastructure

Greater EV adoption (EVA) requires improved infrastructure; this means more charging stations have to be established to increase demand and further reduce range anxiety among the owners of such vehicles. The UK government has hence been

[2] https://www.current-news.co.uk/worlds-largest-domestic-v2g-trial-finds-hardware-costs-still-too-high-despite-significant-financial-rewards/

proactive in investing in the installation of public charging points across the country. As of January 2024, there were more than (35,000) public charging devices have been installed in the UK, including rapid and ultra-rapid chargers. Distribution, however, remains unequal and continues to be focused on more urban areas, therefore the challenge is likely to be the rural-area EVA, and how such infrastructure deployment is strategically planned. EVCI management will involve considerations such as station occupation, station maintenance, and energy load management. Variability in station occupation leads to situations where certain stations are congested whereas others remain underutilized (Dastpak *et al.,* 2023). Efficient management systems that support better utilization of the current infrastructure are needed, and which enable planning for future expansion should be implemented.

1.3. EV Owners and Range Anxiety

Range anxiety describes the fear that an electric vehicle does not have enough battery capacity to arrive at its planned destination and is thus a critical psychological barrier to EVA. This fear is further complicated by concerns related to the availability and quality of charging infrastructure. More recent scholarship suggests that early EV adopters tend to have high levels of range anxiety, which decreases as they gain experience and become more knowledgeable about the technology (Rauh et al., 2015). The factors that instigate this anxiety include, most notably, limited battery range, insufficient EVCI, and a lack of real-time data about the location of charging stations for end-users of electric vehicles in particular. All the above-mentioned challenges are addressed by strongly directing research development efforts towards the improvement of battery technology and range prediction systems. Advancements in battery design, such as increased energy density and fast-charging capabilities, have increased the range of EVs, making them more attractive to consumers. Range prediction algorithms now take into account real-time data such as topography, traffic conditions, and weather, giving users more accurate and reliable estimates of their remaining range. These innovations reduce range anxiety and increase user confidence in, and trust of, the technology.

Range anxiety has a great influence on charging behavior, such as frequent recharging of facilities and preferences for home overnight charging. These trends create peak periods of demand, which strain the electric grid, thus underlining the need for effective demand management strategies (Viswanathan et al., 2018). Complementing increased ranges and improved range prediction capabilities, various solutions are being realized to match demand and supply smart charging and vehicle-to-grid (V2G) technologies, to name just a few. The question of charging station security and reliability is very important in building users' confidence. Challenges such as faulty stations, occupied by vehicles not charging, or even cybersecurity threats discourage the use of EVs and slow down their wide adoption. Physical security ensures that charging stations are accessible and functional, while cybersecurity issues arise from the integration of charging infrastructure into digital networks, rendering it vulnerable to hacking and data breaches (Ostermann et al., 2022). Robust security protocols, including regular maintenance, continuous monitoring systems, and user authentication procedures, are critical to safeguarding the integrity of infrastructure as well as the data of users. The adoption of standardized security frameworks could mitigate risks and improve the overall reliability of the charging network(El Hafdaoui et al., 2023).

1.4. Role of Modelling in Managing Charging Processes

Progress in artificial intelligence (AI), especially deep learning (DL), provides novel options for overseeing EV charging operations. DL algorithms can evaluate extensive datasets to forecast charging station occupancy, optimize energy allocation, and improve user experience. Occupancy prediction at charging stations, with good accuracy, has also been made using DL methodologies such as long short-term memory (LSTM) networks and temporal convolutional networks (TCNs) (Amara-Ouali *et al.,* 2023; Fan *et al.,* 2023). Such frameworks embed both temporal and spatial variables, and thus can identify complex patterns related to charging behaviours. For instance, Amara-Ouali *et al*. (2023) have developed models intended to predict the occupancy level for the purpose of aiding users in finding an available charging station and helping operators distribute resources. Accurate predictions reduce waiting times and enhance user satisfaction, while optimizing EVCI use effectiveness.

The deep learning model (DLM) helps in energy management through its demand pattern prediction to ensure that grid operators balance efficiently, avoiding grid overload because of peak charging hours or integration of renewable energy. Su *et al*. (2023) state that predictive models also contribute towards enhancing user experience by offering real-time information, and customized suggestions pertaining to the identified needs. Applications can guide drivers to available charging stations, considering factors such as location, wait times, and charging speeds. This reduces range anxiety and promotes efficient charging habits.

Traffic congestion and emissions are critical issues in metropolitan environments. Predictive models are essential in alleviating these problems by enhancing the allocation of EVs throughout the EVCI. For instance, predictive occupancy models can notify drivers of the state of particular charging stations, and thus could help reduce unnecessary trips and time use in searching for places to charge. This reduces traffic congestion and the resultant emissions. The efficient management of EVCI supports the broader goals of smart cities by enhancing mobility, reducing environmental impacts, and improving the quality of urban life. Moreover, integrating predictive models into transportation systems could contribute to sustainable urban development.

Understanding EV owners' charging habits is essential for designing effective EVCI and management systems. Charging habits are influenced by factors such as daily travel patterns, access to home or workplace charging, electricity tariffs, and range anxiety. Many owners prefer home charging overnight due to convenience and lower off-peak electricity rates (Viswanathan *et al.,* 2018). Recognizing these habits allows for better infrastructure planning, ensuring that charging stations are strategically located and equipped to handle peak demand periods. Predictive models can analyse usage patterns to forecast demand and guide infrastructure development.

1.5. Research Objectives

This research aims to explore the application of DLMs in predicting EV charging station (EVCS) occupancy, and to provide insights into related user behaviour and preferences.

1.5.1. Model Comparison

DLMs have different learning abilities, and therefore produce predictions with varying degrees of accuracy. Therefore, this research undertakes a comparative analysis of different DLMs, such as LSTM, TCN, and models ensembling techniques, in order to evaluate their effectiveness in occupancy prediction. The comparison includes several considerations, such as prediction accuracy, generalization, and appropriate task identification.

1.5.2. Usability Insights

This research fundamentally seeks to understand how predictive models can improve the user experience, reduce range anxiety, and affect charging habits. Another complementary aspect of this is how different user interface designs and information presentation facilitate effective decision-making by EV owners.

1.6. Research Questions

In order to achieve the objectives explained above, this research seeks to answer the following research questions:

1. How accurately can DLMs predict the availability of EVCSs based on historical data?
2. Which factors most significantly impact the accuracy of DLMs in predicting EVCS occupancy?
    2.1. Which specific features improve the quality and relevance of the training data for predicting EVCS occupancy?

2.2. How does the use of aggregated data from multiple locations versus location-specific data impact the generalization and accuracy of DLMs in predicting EVCS occupancy?

3. How does indicating the prediction of the EVCS occupancy as a classification versus regression task affect the performance and accuracy of DLMs?
4. What display expectations do EV owners prefer to see for EVCS occupancy prediction?
5. What are the key factors that influence EV owners' trust in predictive models as a source to manage their charging times?

1.7. Thesis Organization

Following this introductory chapter, the rest of this thesis is organized into the following chapters:

Chapter 2: Presents the general research background pertaining to EVCI, EVA, EV user behaviour and deep learning modelling.

Chapter 3: Reviews literature on the application of DL in occupancy prediction and summarizes related work on EV occupancy prediction.

Chapter 4: Details and justifies the methodology of this study, including the data collection and analysis, model development, evaluation metrics used and the conducted user studies.

Chapter 5: Analyses the resultant data to provide an understanding of its implications for existing research and practice.

Chapter 6: Develops the studied models, comparing the models' outputs to analyse their performance outcomes.

Chapter 7: Provides an introduction to the proposed Bidirectional-GRU and TCN (BiGTCN) architecture for occupancy classification and the results achieved,

Chapter 8: This chapter dedicated to show the User Study A, where a quantitative analysis conducted to evaluate the user experience.

Chapter 9: This chapter is devoted to presenting User Study B, which involved a qualitative user analysis aimed at assessing usability and trust.

Chapter 10: Summarizes the outcomes of this thesis, highlighting its major results and implications, and proposing future research directions to extend on the results of this study.

Chapter 2: Background

This Chapter provides the background necessary for both the historical and contemporary understanding of electric vehicles, as well as the technological underpinning for deep learning approaches for forecasting. First, it provides a review on the historical evolution of EVs in order to put their development and rise to prominence into perspective. Then, it moves on to discuss the state of the present day in terms of adoption, with particular attention to the landscape within the United Kingdom, considering global trends along with Enhanced EVCI. Further, it presents the key deep learning techniques for sequence modelling, namely LSTM, GRU, TCN, 1D-CNN, and BiGRU. Lastly, the advantages of the approaches in time series forecasting are demonstrated in order to provide a basis for presenting a theoretical framework that encompasses all the important themes, including the dynamics of spatial-temporal charging needs and users' behavior. This extensive overview sets the stage for the subsequent analytical chapters.

2.1. Historical Evolution of EVs

EVs are conventionally thought of in the context of very recent technological breakthroughs and increased concern for eco-friendly transportation in recent decades, electric cars were in fact first designed in the 19th century, when the first electric carriages were piloted alongside steam- and horse-drawn carriages. However, internal combustion engines dominated the 20th century, with the result that active development of EVs practically came to an end, though the concept still lingered among some engineering specialists. However, powered by the progress in the fields of battery technologies, government subsidies, and ecological awareness, the structure of international EVA is growing very fast nowadays. It has been characterized by increased uptake across countries, driven by pioneer markets such as China, the US, and Europe. The adoptions have been supported through policies aimed at reduced emissions in the UK and backed by huge investments in EVCI. Today's environment depicts a world of rapid international movement toward electric mobility, with ambitious goals set to reduce carbon emissions and address climate change with sustainable means of transportation.

Historically, there have been several phases associated with the concept of EVs. In any case, there is a history of significant resurgence, further intertwined with other historical challenges. Although EVs have experienced a notable resurgence in the recent past, there are various overarching concerns that prevail. This section will try to provide a concise overview of the historical development associated with EVs.

Beginning in the latter part of the 19th century, preliminary investigations were undertaken regarding the notion of EVs in their primitive form. The beginning of EVs is attributed to Gaston Plante, a French inventor who created the lead-acid battery in 1859. In 1881, another French scientist, Gustave Trouve, marked a fundamental innovation by effectively utilizing an electric motor to power the Starley Tricycle (Taylor 2022), and in the following year the professors William Edward Ayrton and John Perry demonstrated an electrically powered tricycle (Wakefield 1993). In 1884, Thomas Parker, a British inventor, introduced an EV operated by a distinctively designed high-capacity battery (Faraz et al. 2021). In 1888 Andreas Flocken built the first German EV (Taylor 2022).

During the late 19th and early 20th centuries, electric cars gained significant popularity due to their comfortable and easy-to-use features (Genofer et al. 2018). EVs were first commercially utilized in the US in 1897. A project funded partially by the Electric Storage Battery Company of Philadelphia utilized a fleet of twelve cabs and one brougham in New York City, based on the concept of the Electrobat II (Genofer et al. 2018). At that time, electric cars were more advanced than combustion cars from the late 19th century (Situ 2009). Constructed in 1899, the Lohner-Porsche Mixte Hybrid marked the inception of plug-in hybrid EVs (PHEVs). The term “hybrid” indicates that these vehicles could be charged from a wall socket, in addition to being powered by other sources (Singh et al. 2021).

Despite this promising start, at the beginning of the 20th century, people’s adoption of EVs has stagnated for a long time due to a variety of factors, including the influence of the automobile and oil industries and social and political dynamics (Genofer et al. 2018). Given the absence of *de facto* use and demand, there was a commensurate dearth of EVI development, as well as limited possible travel distances in the rare cases where some efforts where made (Situ 2009). This had to negative impacts on the electric car industry's development and public acceptance, and the automobile industry developed with total reliance on oil (e.g., diesel- and petrol-fuelled internal combustion engines).

Investment and interest in developing the electric car industry, however, come to the fore in the search for environmentally friendly sources and sustainable energy instead of relying on fossil fuels as well as coping with the notable spike in oil prices (Yu et al. 2011). Since then, the weight, storage capacity, and recharging time of the car battery have been the main concerns in the development of electric cars.

During the mid-1960s, GM introduced their initial prototype EV, the "Electrovair", powered by a silver-zinc battery pack capable of delivering up to 532 volts (Faraz et al. 2021). However, this idea failed to achieve mass production. GM's EV1, developed in the mid-1990s, was their inaugural electric car to be produced on a large scale (Yu et al. 2011). However, GM ceased production of the EV1 in 1999 due to concerns about its profitability, despite its initial popularity (Yu et al. 2011). California implemented a policy mandating producers to manufacture automobiles with zero emissions, thus revitalising its planning and manufacturing efforts (Faraz et al. 2021). Subsequently, Ford, Honda, Toyota, and Nissan introduced comparable vehicles, all of which were made available to customers through leasing agreements.

Despite recent advancements in technology, there are still lingering challenges from the past. Nevertheless, electromobility is receiving significant media and policy focus, as prominent automakers are actively creating and introducing electric and hybrid vehicles. The market is growing but the customers are still not comfortable to fully migrate to electric motor technology, still using hybrids as a substitute to bridge the gap (Yu et al. 2011).

Electromobility represents a significant opportunity for the automotive industry to address sustainability challenges; however, it also introduces substantial difficulties.

The adoption of EVs would require a comprehensive transformation of the automobile manufacturing sector itself, and public transport networks, affecting production processes, energy supply infrastructures, business frameworks, marketing approaches, and regulatory policies. This shift involves considerable alterations that extend beyond the simple replacement of vehicle powertrains (Yu et al. 2011).

2.2. Current Adoption Rates

In many countries, there is a great interest and desire to develop and encourage the adoption of electric cars rather than internal combustion engine cars. Perhaps the most prominent motive for this is to reduce dependence on oil due to greenhouse gas emissions, and localized environmental pollution (e.g., poor air quality in urban environments) (Situ 2009). This section reviews the current global status of EVA, with a particular emphasis on the UK, giving a general overview of some aspects that reflect the extent of integration and acceptance of electric cars in societies. Among these aspects are the public's influence, market orientation, government policy, and the extent of their seriousness in providing and developing the necessary infrastructure for these projects.

2.2.1. Global EV Landscape

The worldwide EV market has undergone substantial expansion in recent years, propelled by breakthroughs in battery technology, decreasing costs, and robust government incentives (Yu et al. 2011). According to the International Energy Agency (IEA, 2024), electric car registrations exhibited a substantial and consistent increase over the course of the five years to 2023. The number of registrants experienced a substantial surge, particularly in the year 2022. These are numerous factors that contribute to increasing attention towards the worldwide need for EVs to make up a transitional shift, compelled by the fast development of technology, motivating policies, and a rise in consumer awareness about the environment.

The Chinese government and manufacturers are major players in the general balance of the EV world, and the Chinese market continues to take leading positions in EV registrations compared to other regions. In 2022, this grew notably to a registration volume of more than 9 million units. Strong supportive politics from the government's side, huge investments in EVI itself, and fast development of the manufacturing sector related to electric mobility are the basis for such fast growth in the Chinese EV sector. This further indicates the high market penetration, as each share of sales continues to reinforce the leading position of China in the global EV market.

In Europe, EV registrations have been recording serious growth since 2020. It would seem that such a trend also continued to see a soar in 2022, given the stiff environmental laws within the region, and ample purchases of electric cars. The leading countries within the European region, including Norway, Germany, and the Netherlands, have formed the frontier in EV adoption in highly mature conventional markets, thereby upgrading the regional landscape. This is reflected in sales statistics indicating an increase in the customer acceptability and policy-driven movement towards EVs in Europe (IEA, 2024).

Conversely, the US exhibited a steady rise in electric car registration. Although there has been consistent expansion, the total figures remain comparatively modest in relation to China and Europe. The erratic surge in 2022 reflects consumer demand acknowledging the maturity of EVs, bolstered by government incentives at both the federal and state levels, while the pace of market adoption is still developing. The sales share ratio emphasizes the gradual and steady increase in the adoption of EVs in the US market.

Figure 2-1 illustrates the dynamic and geographically distinct growth trends in the worldwide electric car market. China and Europe are in the forefront of these initiatives, with substantial annual growth in electric car registrations and capturing considerable portions of the market. Conversely, the US has a favourable, if more gradual, pattern of expansion. These trends demonstrate differing levels of market development and the efficiency of policies in various locations, offering vital insights into the worldwide scenario and the rates at which EVs are being adopted. This analysis strengthens the worldwide trend towards electric mobility, which is propelled by a mix of legal frameworks, technology advancements, and shifting consumer preferences. It is crucial to comprehend the regional dynamics in order to develop effective policies and strategies for promoting sustainable mobility globally, as adoption rates continue to increase (IEA, 2024).

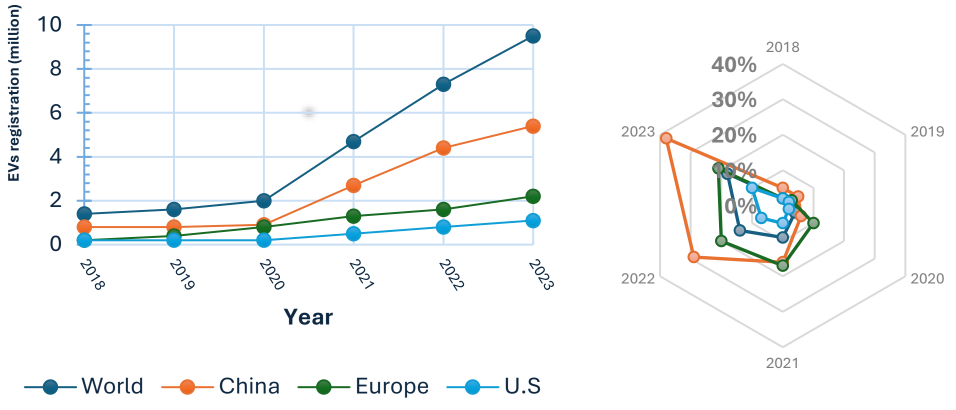


Figure 2-1 Electric Car Registrations (a) and Sales Share (b) in China, US, and Europe, 2018-2023
Source: IEA (IEA, 2024)

2.2.2. UK EV Landscape

The EV transition is very important for the UK to reduce greenhouse gas emissions and decrease dependence on non-renewable sources of energy. Incentives and favourable policies the UK government has implemented for EVs include subsidy and tax exemption policies, among many others. According to previous studies, the increasing price competitiveness of EVs has significantly contributed to their uptake among UK consumers (Chen et al. 2020), but key challenges remain in terms of EVCI and consumer acceptance (Agbro et al. 2021).

Despite these advancements, substantial regional inequalities persist in the UK's EV charging infrastructure. According to the Department for Transport (2025)[3], London significantly outpaces other UK regions with approximately 250 public charging devices per 100,000 residents, which is more than double the UK average of 108 per 100,000. In stark contrast, regions such as Northern Ireland have only 36 devices per 100,000 residents, while the North West and Yorkshire and the Humber both have merely 66. This notable disparity underscores significant unevenness in charging infrastructure deployment, highlighting potential barriers to EV adoption outside of

[3] https://www.gov.uk/government/statistics/electric-vehicle-public-charging-infrastructure-statistics-january-2025

the more equipped regions. Addressing these disparities is therefore critical for promoting equitable access to charging infrastructure, thereby facilitating broader and more uniform adoption of electric vehicles across all regions of the UK.

In this respect, the construction of EVCI is treated as one of the main driving factors for EVA in the UK. Alongside this, the network of EV charging points in the country is growing very rapidly (Chen *et al*., 2020). Notwithstanding the advancements achieved, substantial regional inequalities endure, with Southeast England and Greater London exhibiting a markedly elevated concentration of public EVCI relative to other regions across the nation (Bhagavathy and McCulloch 2020). Addressing these inequalities is crucial to ensure fair access to EVCSs across the country, thus promoting the broader adoption and utilization of EVs.

EVs represent a valid means of decreasing the UK's reliance on non-renewable resources. For example, Raugei *et al*. (2018) modelled small Battery Electric Vehicle (BEVs) and realized a 34% reduction in non-renewable cumulative energy demand when compared against the use of an Internal Combustion Engine Vehicles (ICEV). It follows that the potential of EVs in decreasing dependence on fossil fuels increases with a national grid ever more based on renewable energy. This could further be improved with better technology for BEVs with the inclusion of renewable energy.

EVs are imperative in achieving the UK's carbon reduction targets. For instance, Küfeoğlu and Hong (2020) expound that heavy penetration into the market may eventually dramatically lower emissions; however, the pace of the prevailing adoption rates does not appear realistic to meet targets by 2050. According to Hill *et al*. (2019), the UK will be able to stick to its carbon budget in cases of faster diffusion of EVs and transformation in an electricity grid that supplied these. For this, immediate action has to be taken to enable the fast diffusion of EVs with improved environmental quality of the electric fuel supplied.

However, major barriers to more widespread EVA remain, including high up-front costs and limited vehicle range. As Steinhilber *et al*. (2013) said, such socio-technical inertia powered by consumer resistance and an inadequate infrastructure is one critical factor for transition. Nowadays, most customers show their concern for the high up-front costs and disadvantages of charging. Besides, inadequate EVCI in many regions adds to the problem; thus, removal of economic and infrastructural barriers is a must for faster penetration.

The preference for knowledge of consumers is regarded as very important to accelerate electric mobility. Early adopters in the UK, according to Mandys (2021), are comparably younger, better educated, and living in the south. A few key drivers for purchase include high costs, driving experience, the range of the vehicle, and environmental benefits. By offering these, manufacturers and policymakers alike will know they have to design their products or policies to suit these in order to increase the level of market penetration. However, while demand for EVs has been higher in the Southeast of the UK, Scotland has exhibited a more proactive approach in terms of government policy.

In September 2013, Transport Scotland (2017) released the “Switched-on Scotland Roadmap”, which included a comprehensive plan and strategic strategy to promote the wider use of EVs. The “Roadmap” presents a detailed and extensive long-term plan to direct Transport Scotland’s endeavours in expediting the shift towards EVs. Its main objective is to rid Scotland’s cities, towns, and communities of what it perceives as the noxious emissions of petrol and diesel cars before the year 2050. Other elements that can be included in this are deep cuts in greenhouse gas emissions, better air quality, less noise, and improved health. Scotland is also interested in the associated ambitions of increasing energy security as well as unlocking new economic opportunities by taking the lead in low-carbon transport and energy technologies.

According to national targets, by the year 2040, nearly all of the newly sold cars should emit very low tailpipe emissions. Moreover, it is expected that by the year 2030, half of the number of fossil fuel-driven vehicles currently on Scotland’s roads will have vanished from the urban areas of Scotland, and EVs, with Scotland’s rich renewable energy source, will play an important role in reaching this goal. It calls for the wide-scale introduction of EVs and electric hybrids within the context of a sustainable transport system. Furthermore, this would give fresh impetus to the course towards greener, up-to-date energy-based infrastructure.

A supportive government policy and improving access to EVCI are what make the UK perform quite well in switching to EVs, but their prohibitive expense, driving range limitations, and inconsistent EVCI remain the three major areas to be overcome for the support of equitable nationwide adoption. Fast growth of EVA, combined with a cleaner electricity grid, supports the achievement of long-term reduction goals. Overcoming these socio-technical barriers and improving our knowledge on the preference of consumers are, in fact, vital to their successful integration into the UK market.

2.2.3. Global EVA

Numerous factors can influence the general public's acceptance and proliferation of electric cars, either positively or negatively. For instance, financial capabilities, local government support, and the availability of adequate infrastructure for electric cars all play a role. All of these factors and more can assist in accepting electric cars as an alternative means. Mekky and Collins (2024) suggest that environmental policies have a larger impact on encouraging EVA compared to financial incentives, indicating the significance of prioritizing policies that focus on reducing emissions and promoting cleaner energy sources.

Morton *et al*. (2017) investigated the various elements that impact the acceptance and usage of EVs and hybrid EVs (HEVs). They found that the main elements that influence or motivate anything include financial incentives, technical progress, and socio-economic factors like income and education levels. Cultural values and risk perceptions are key factors that differ among countries. Implementing effective policy measures, such as marketing and product development strategies that reduce perceived risks, as well as financial incentives like freebies, are essential for boosting adoption rates. The findings of Morton *et al*. (2017) highlight the significance of adopting a comprehensive approach that considers both factors that encourage and hinder the adoption of low-emission vehicles in order to facilitate the transition and accomplish sustainability objectives.

According to Morton *et al*. (2017) a scarcity of charging facilities, exorbitant costs, and worries about how far a vehicle can travel before needing a recharge can be major barriers to the adoption of EVs.

2.2.4. Enhanced EVCI

The expected development in EVCI should focus on the creation of really efficient, sustainable, and accessible EVCI that will drive the fast growth of EVs globally. With growing markets comes a growing demand for models that can resolve different logistical and operational problems. The main issues are optimal charging station distribution, real-time prediction of charging demand, and occupancy monitoring of the charging station for reliability and convenience in the EV charging networks. DL approaches have been put forward for modelling these challenges, and promising results have been achieved to date, by which different stakeholders can make forecasts of the demand for EVs, enhance user experience, and efficiently manage resources. Notable success stories in this field, that will be demonstrated in the following chapter, include projects that have leveraged machine learning models to predict and balance charging loads dynamically, improving operational efficiency and user satisfaction (Viswanathan et al. 2018a; Zhu et al. 2019; Hecht et al. 2021; Qiao and Lin 2021). Furthermore, the use of predictive models for occupancy forecasting has been instrumental in minimizing wait times and improving accessibility, especially in densely populated urban areas where charging availability is essential. These are indeed success cases of how DL can be transformative in tackling the challenges of EVI, but many more advanced data-driven solutions are foreseen in the coming years.

While these advancements underscore the potential of predictive models, more research is needed to address several open challenges. For instance, existing studies often rely on generalized datasets, which may overlook location-specific patterns and user behaviours. In this research, occupancy prediction was performed using historical data from three distinct charging stations to understand the variability in usage patterns and improve location-specific accuracy. Furthermore, this study incorporated user evaluation of prediction results, providing insights into how EV owners perceive and interact with predictive tools. Another contribution of this research is the comparison between predicting occupancy levels as discrete categories and predicting total occupancy as continuous numerical values, offering insights into the strengths and limitations of each approach. These contributions aim to bridge gaps in existing literature, providing a more user-centred and context-sensitive understanding of EV charging behaviours.

2.3. Deep Learning for Sequence Modelling

DL has become an innovative tool in multiple disciplines, driven by the expansion of sophisticated neural network (NN) topologies. Noteworthy among these are LSTM, Gated Recurrent Unit (GRU), TCN, 1-Dimensional Convolutional Network (1D-CNN), and Bidirectional GRU (BiGRU), recognized for their distinctive abilities to manage sequential data and temporal dependencies.

2.3.1. Long-Short Term Memory (LSTM) Network

LSTM networks (Figure 2-2) are a type of recurrent NN (RNN) designed mainly to overcome the vanishing gradient problem, which often hampers the training of traditional RNNs. The solution is achieved by incorporating memory cells that can maintain information over long periods, allowing them to learn from sequences of data effectively. LSTMs are particularly useful in applications such as speech recognition, language modelling, and time series forecasting, as in this research, wherein understanding long-range dependencies is crucial (Lipton et al. 2015). The architecture of LSTMs includes input, output, and forget gates that regulate the flow of information, enabling the network to retain or discard information as needed (Lipton *et al*., 2015; Marei and Li, 2022).

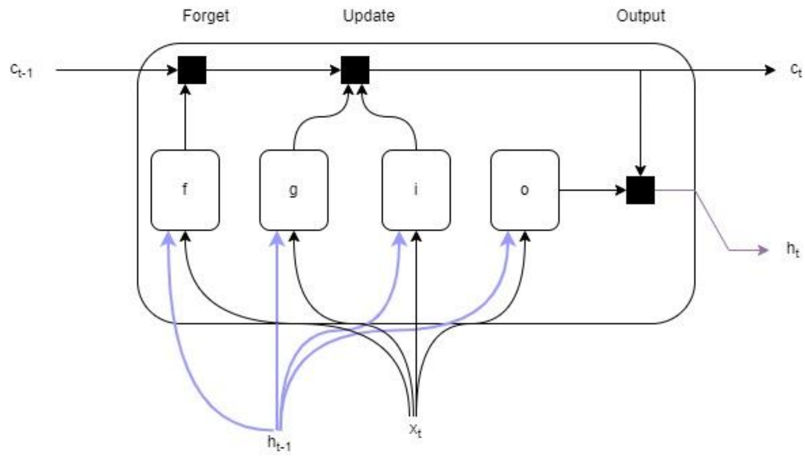


Figure 2-2 LSTM Architecture

Source: Marei and Li (2022)

$$i_t = \sigma_g(W_i x_t + R_i h_{t-1} + b_i)$$

Equation 2-1

$$f_t = \sigma_g\left(W_f x_t + R_f h_{t-1} + b_f\right)$$

Equation 2-2

$$g_t = \sigma_c\left(W_g x_t + R_g h_{t-1} + b_g\right)$$

Equation 2-3

$$o_t = \sigma_g(W_o x_t + R_o h_{t-1} + b_o)$$

Equation 2-4

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$

Equation 2-5

$$h_t = o_t \odot \sigma_c(c_t)$$

Equation 2-6

2.3.2. Gated Recurrent Unit (GRU)

GRUs (Figure 2-3), a simplified variant of LSTMs, also address the vanishing gradient problem, but with fewer parameters, making them more computationally efficient. GRUs combine the forget and input gates into a single update gate, which simplifies the architecture while still capturing essential temporal dependencies (Lipton et al. 2015). This makes GRUs particularly advantageous in scenarios where computational resources are limited or where rapid inference is required. Studies have shown that GRUs can perform comparably to LSTMs in various tasks, such as natural language processing and time series prediction (Lipton *et al.*, 2015; Marei and Li, 2022).

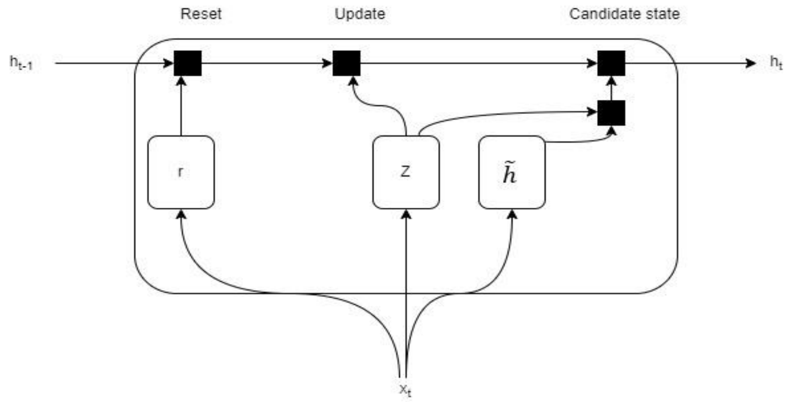


Figure 2-3 GRU Architecture

Source: Marei and Li (2022)

$$r_t = \sigma_g\left(W_r x_t + b_{W_r} + R_r h_{t-1}\right)$$

Equation 2-7

$$Z_t = \sigma_g\left(W_Z x_t + b_{W_Z} + R_Z h_{t-1}\right)$$

Equation 2-8

$$\tilde{h}_t = \sigma_s\left(W_{\tilde{h}} x_t + b_{W_{\tilde{h}}} + r_t \odot (R_{\tilde{h}_t} h_{t-1})\right)$$

Equation 2-9

$$h_t = (1 - Z_t) \odot \tilde{\tilde{h}}_t + Z_t \odot h_{t-1}$$

Equation 2-10

$$y_t = \sigma_g(W_y(h_t) + b_y)$$

Equation 2-11

2.3.3. Temporal Convolutional Network (TCN)

TCNs (Figure 2-4) represent another innovative approach to sequence modelling. Unlike RNNs, TCNs utilize convolutional layers to capture temporal dependencies, allowing for parallel processing of input sequences. This architecture employs dilated convolutions, which enable the network to expand its receptive field without significantly increasing the number of parameters. Research indicates that TCNs can outperform LSTMs in certain applications, particularly in tasks requiring long-range temporal dependencies, such as plasma disruption prediction. The ability of TCNs to handle varying input lengths and their robustness against overfitting make them a compelling choice for many sequence-based tasks (Lu et al. 2020).

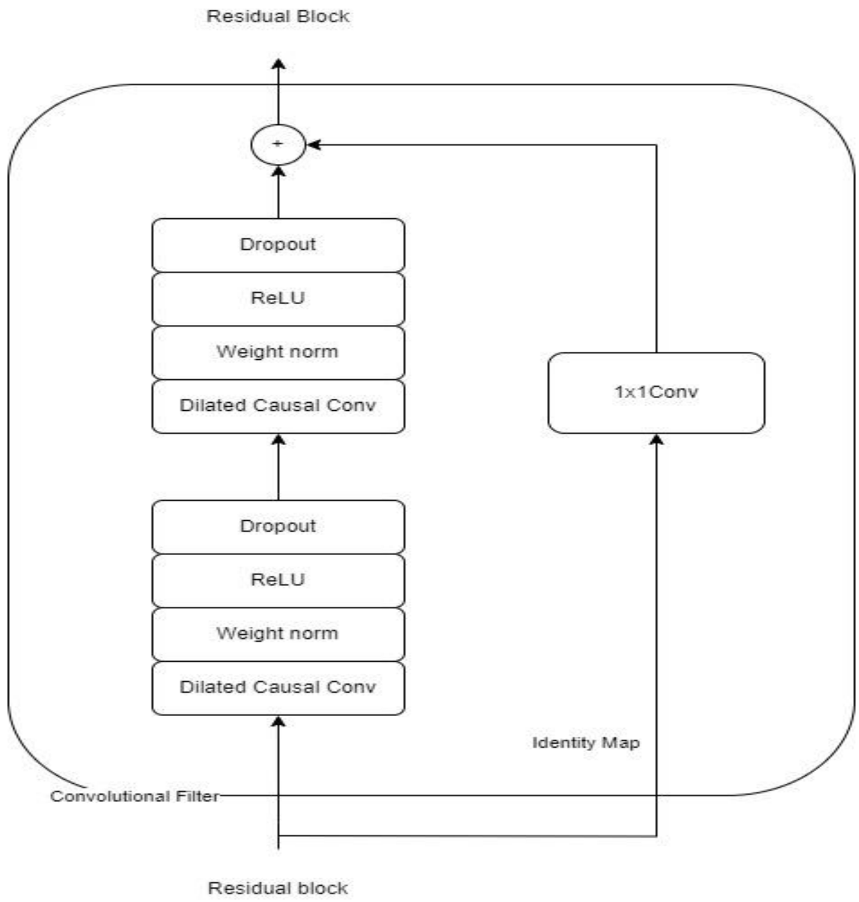


Figure 2-4 TCN Architecture
Source: Lu *et al*. (2020)

As can be seen in Figure 2-4, each residual block has been implemented as two layers with dilated convolutions, followed by weight normalization, Rectified Linear Unit (ReLU) activation, and Dropout. Weight normalization is applied at every convolutional filter in the process of a forward pass. The result of dilated convolutions is added to the input using element-wise summing. If there is a dimension mismatch between the input and output, an additional 1×1 convolution can be used. Otherwise, identity mapping is performed (Lee et al. 2021).

2.3.4. 1D-Convolutional Neural Network (1D-CNN)

1D-CNNs are specifically designed for processing one-dimensional data, such as time series or sequential signals. These networks leverage convolutional layers to extract features from the input data, enabling them to learn hierarchical representations effectively. The architecture typically consists of multiple convolutional layers followed by pooling layers, which help reduce dimensionality while preserving essential features (Rai and Mitra, 2021). Other studies have demonstrated the efficacy of 1D-CNNs in various applications, including Electrocardiogram (ECG) signal classification and remote sensing data analysis, where they have shown superior performance compared to traditional methods (Zhao et al. 2019a; Feyisa et al. 2022; Salah et al. 2023).

1D-CNNs have lately gained much attention regarding the task of classifying time series data due to their ability to automatically extract features from raw data without depending on a great amount of manual feature engineering. This is the quality that makes the 1D-CNNs a very useful tool in many fields, such as medicine, environmental monitoring, and condition monitoring of structures. 1D-CNNs are designed to process sequential data; hence, it is suitable for time series classification problems. They use convolutional layers that are instrumental in capturing the local pattern and dependency within the data in order to understand the temporal dynamics effectively. For instance, Yuan *et al*. (2021) have shown real-time classification capability with 1D-CNN, showing its efficiency in handling the time series efficiently. In similar thinking, Feyisa *et al*. (2022) emphasize the issue of 1D-CNNs that can grasp directly from the raw time series data without any requirement of domain expertise feature extraction. This can thus be an innovative representation learning from the data itself, whereby models are built to adapt to a variety of time series with a minimum of pre-processing.

Besides, 1D-CNN architectures have shown strong performance in specific classification tasks, such as time-series analysis and sequential data classification, where spatial patterns in the input are critical. This could include something like the kernel size, since that would affect the capability to capture relevant features at any given scale of a model. Tang *et al*. (2020) discussed how selecting appropriate kernel sizes can significantly affect the performance of 1D-CNNs in time series classification tasks, suggesting that a systematic approach to kernel size configuration can enhance model efficacy. Adding a number of convolutional layers enhances the capability of the model to find complicated patterns of the data, as supported by a study on DLMs for time series classification conducted by Zhao *et al*. (2019).

2.3.5. Bidirectional Gated Recurrent Unit (BiGRU)

BiGRUs extend the capabilities of standard GRUs by processing input sequences in both forward and backward directions. This bidirectional approach allows the network to capture context from both past and future states, enhancing its ability to understand complex temporal patterns. BiGRUs have been successfully applied in tasks such as sentiment analysis and machine translation, where context from both directions is crucial for accurate predictions (Lipton et al. 2015). The combination of forward and backward processing enables BiGRUs to achieve state-of-the-art performance in various sequence learning tasks.

The integration of these architectures into hybrid models has further enhanced their capabilities. For instance, combining 1D-CNNs with LSTM layers has been shown to improve performance in tasks such as arrhythmia detection, where both spatial and temporal features are critical (Picon et al. 2019). This hybrid approach allows the model to leverage the strengths of both convolutional and recurrent architectures, resulting in improved accuracy and robustness.

The diverse architectures of LSTM, GRU, TCN, 1D-CNN, and BiGRU each offer unique advantages for DL applications, particularly in the realm of sequence modelling. LSTMs and GRUs excel in capturing long-term dependencies, while TCNs provide efficient parallel processing capabilities. 1D-CNNs are adept at feature extraction from sequential data, and BiGRUs enhance context understanding by processing sequences bi-directionally. The ongoing research and development in these architectures continue to push the boundaries of what is possible with DL, making them invaluable tools in various fields, including healthcare, finance, and natural language processing.

2.4. Deep Learning Advantages for Time Series Forecasting Problems

DL has emerged as a powerful tool for time series forecasting, offering numerous advantages over traditional statistical methods. One of the most significant benefits of DLMs, particularly those based on RNNs such as LSTM networks, is their ability to capture complex temporal dependencies in data. The LSTM networks are uniquely built for sequential data processing, making them particularly proficient for tasks related to time series forecasting, where the relationships among observations may span extensive time frames. This is very important in a range of practical applications, such as air pollution level prediction, since the aftermath of events that happen today can be felt over very long periods of time (Sun *et al*., 2022; Velarde, 2023).

Moreover, DLMs have the great capability of extracting features automatically from the raw data themselves, hence bypassing heavy pre-processing and manual feature engineering. This attribute is particularly useful in the context of time series forecasting, where the underlying patterns are complicated and hence not easy to discern. For example, research has been able to prove that DLMs trained on data such as the M3 competition dataset, which contains 3,003 time series covering a range of domains including **microeconomic**, **macroeconomic**, **financial**, **industry**, **demographic**, and other sectors (Solís and Calvo-Valverde 2022) . These series vary in frequency (yearly, quarterly, monthly, etc.), length, and complexity, performing better than traditional approaches to forecasting, thus giving indications of being able to recognize and use the underlying patterns (Solís and Calvo-Valverde, 2022). The automatic feature extraction capabilities of DL also reduce the time and expertise required to develop forecasting models, making them more accessible to practitioners across various fields (Lara-Benítez et al. 2021).

Another essential advantage of DL in the forecast of time series is the intake of noise and random fluctuations that might occur in the data. Non-stationary time series data, with a possible presence of trends, seasonality, and aberrations, are hard for traditional statistical methods to handle. Using methodologies such as dropout and batch normalization, the DLMs minimize the effect of noise and stabilize the model (Aktas *et al*., 2023; Pavlyshenko, 2022). This robustness is particularly beneficial in domains like finance and healthcare, where data can be volatile and unpredictable (Gunarto *et al*., 2023; Shahid *et al*., 2020).

DLMs have very significant scalability advantages. As the volume of data grows, the traditional methods may not work so well since they rely on fixed parameters and assumptions about the static nature of the underlying distribution of data. DLMs can scale up to meet very large datasets, thus enabling more sophisticated analysis that improves the accuracy of forecasts (Pavlyshenko, 2020). For example, the use of ensemble methods and hybrid models that combine different DL architectures, such as CNNs and LSTMs, has been shown to yield superior performance in multivariate time series forecasting tasks (Salinas *et al*., 2020; Wan *et al*., 2019). This flexibility enables practitioners to tailor their approaches to specific forecasting challenges, further enhancing the utility of DL in this domain.

Besides, interpretability for DLMs has hugely improved with certain recent developments, especially since the introduction of attention mechanisms and "explainability" in AI methodologies. While traditional DLMs often suffered from criticism for their "black box" nature, recent developments have allowed one to look more closely into the processes responsible for their capabilities of prediction. Attention-based models, for instance, are able to identify which of the past observations bear the most relevance for predicting future outcomes, providing relevant context to the decision-makers (Lim et al. 2019).

Another factor that increases the efficiency of DL frameworks for time series forecasting is the integration of transfer learning into them. Transfer learning enables the adaptation of models trained on one dataset for use on another, potentially reducing the amount of training data required, improving performance in related tasks (Solís and Calvo-Valverde, 2022). This is particularly beneficial in scenarios where historical data is limited or costly to obtain, enabling practitioners to leverage existing models and knowledge to enhance their forecasting capabilities (Du et al. 2023). The ability to generalize across different but related datasets is a significant advantage of DL, making it a versatile tool for a wide range of forecasting applications.

In the context of time series forecasting, the advantages of DL are multiple: it can uncover complex temporal dependencies, realize automatic feature extraction, and proved resilient against noise; besides that, it shows scalability, enhanced interpretability, and the possibility of transfer learning. Such properties make DL one of the important approaches to solve the difficulties related to time series forecasting and to provide a practitioner with effective tools enhancing predictive ability in many diverse domains. As the research develops in this area, new architectures and methods will likely lead to even more significant improvements in the precision and relevance of the forecast.

In the field of time series forecasting, Dynamic Linear Models, especially frameworks like Long Short-Term Memory (LSTM) networks and Temporal Convolutional Networks (TCNs), have been found to be remarkably effective. These models are great at capturing complex temporal relationships and extracting features by themselves from raw data, which turns out to be very important for most forecasting tasks in many different fields (Li and Law, 2024). Large Language Models (LLMs), exemplified by OpenAI's GPT series, are principally intended for tasks involving natural language processing. Nonetheless, contemporary studies have examined their potential use in forecasting time series data. For example, research has analysed the reprogramming of LLMs to accommodate time series information by correlating time series modalities with inputs derived from natural language (Jin et al., 2024). And in both cases, methods were also proposed to represent time-series as tokenized sequences of numerical values so that LLMs would have been able to zero-shot extrapolate the time series (Gruver et al., 2024).

Despite these novel approaches, the effectiveness of LLMs in time series forecasting remains a subject of debate. Ablation studies have shown that removing the LLM component or replacing it with a basic attention layer does not degrade forecasting performance; results even improve in some cases (Tan et al. 2024). This might indicate that, while LLMs have strong pattern recognition abilities, the application of these in time series forecasting does not necessarily lead to better results than the specialized DLMs. While LLMs offer some promising avenues for time series forecasting; especially via innovative reprogramming techniques; DLMs such as (LSTM, GRU and

TCN) are by far more established and specialized tools in this regard. The choice between LLM and DLM must consider the characteristics of data, computational resources, and considerations relating to interpretability specific to a given forecasting problem.

2.5. Theoretical Framework

This framework discusses DLMs, CNNs and RNNs specifically, which can be used to predict the availability of EVCSs, including time series forecasting, capturing spatial-temporal dynamics, and user behaviour. These theories underpin the research questions pertaining to the accuracy, performance, and user trust of predictive models in EVCI.

2.5.1. Time Series Predicting with Machine Learning and Deep Learning

CNN and RNN are highly concerned with temporal dependencies and nonlinearity in data. Most prediction applications rely on their historical data to predict the future; hence, such models are also suitable for the prediction of EVCS occupancy. Several studies have demonstrated the effectiveness of DL methods in the prediction of occupancy state of EV charging space. For instance, some studies showed how the occupation status of EVCI can be predicted for the next day using machine learning models (Hecht *et al.,* 2021; Ma and Faye, 2021), while Becerra-Rico *et al*. (2020) found that deep networks (e.g., LSTM and GRU) are feasible for forecasting the non-linearity of data. These models are important in planning the availability of the charging station, which normally varies within a day.

Connection to RQs:

RQ1: This theory thus supports the first research question directly, through detailing the potential of DLMs in predicting with high accuracy the EVCS occupancy from historical data.

RQ2: This theory pertains to the effects of input features, such as time, date, and weather, on prediction accuracy.

2.5.2. Spatial-Temporal Factors Affecting EV Charging Demand

The charging demand depends on the time of day, charger placements, and flow in the neighbourhood. These are key factors that any predictive models would account for to improve predictions' accuracy. This amends the ability to forecast the availability at EVCSs by considering traveling peaks, user behaviour, and geographical distribution in a spatial-temporal analysis. Ma and Faye (2021) stated that various factors, including the time of the day, day of the week, weekday/weekend, charging power, could influence charging occupancy profiles. Recently, Yang *et al*. (2024) enhanced this further by incorporating a time series prediction model with a mechanism of fused attention that captures the fluctuation in demand during different parts of the day to detect congestion at a fast-charging station.

Connection to RQs:

RQ2-2: This theory addresses the trade-off between training models on location-specific data versus using combined data from multiple locations, and how this choice impacts model generalisation and predictive accuracy.

RQ2-1 The theory examines how spatial and temporal characteristics (e.g., time of day, geographic parameters) influence model performance.

2.5.3. User Behaviour and Demand Forecasting

User behaviour is a principal determinant in the demand forecasting of charging stations. For this study, each model should integrate user-related variables of charging habits, preferred location, and frequency of a session of use. Once these behavioural patterns are discerned, the model is able to predict not only overall demands but also user interactions with EVCI at different times and locations. For the EV charging application, Majidpour *et al*. (2016) observed that using customer profile data, comprising charging records, leads to higher accuracy compared to using station-level measurements only. In this line, Nespoli *et al*. (2023) extracted knowledge of user behaviour patterns through clustering techniques that allowed them to have more targeted and precise forecasting models related to the sessions of EV charging.

Connection to RQs:

RQ(4&5): This theory is important to understand the factors that motivate the trust of EV owners in the output of the predictive model.

RQ(4&5): This theory helps in finding the barriers related to the adoption of the predictive model for the availability of the EVCS, as this considers user preferences and behaviours.

Chapter 3: Literature Review

The literature review in this Chapter investigates into the topic of EV charging, initially focusing on the key behavioural aspects that impact EV charging, and subsequently identifying the key factors that influence these behaviours. It then highlights how the demand and energy consumption for EV charging have evolved in different contexts, driven by technological developments and user patterns. The review of the existing literature on occupancy prediction at EV charging stations points to the methods and findings that lie at the heart of this rapidly growing research area. A literature summary at the end of the chapter summarises these studies, highlighting unaddressed gaps and establishing the groundwork for the subsequent investigation and methodological design of the thesis.

The literature review in this chapter was carried out using a structured and iterative process to ensure thorough identification and synthesis of relevant studies across multiple research domains. A combination of systematic and narrative review techniques was adopted to balance breadth and depth. The review process began with an initial scoping phase to define the scope of inquiry, followed by targeted searches using electronic academic databases including IEEE Xplore, ScienceDirect, SpringerLink, and Google Scholar. Search terms included combinations of keywords such as “electric vehicle (EV) charging behaviour,” “charging station occupancy,” “EV infrastructure planning,” “spatiotemporal forecasting,” “machine learning for EVs,” and “deep learning prediction models.”

Inclusion criteria were applied to select peer-reviewed journal articles, conference papers, and reputable technical reports published primarily within the last ten years, with particular emphasis on studies offering empirical results or introducing novel modelling techniques. Relevance was further determined by evaluating the methodological rigor of each study, the clarity of objectives, and their contributions to understanding behavioural trends, power demand forecasting, or occupancy prediction at EVCSs. Studies were categorised thematically under behavioural influences, charging demand and power forecasting, predictive modelling approaches, and user-centric considerations.

The synthesis process involved critical comparison of findings, highlighting methodological strengths and limitations, and identifying conceptual or technological gaps. In particular, special attention was given to works addressing public EV infrastructure (EVCI) in urban contexts, privacy-aware predictive modelling, and real-time data integration. This review informed the framing of the research problem, the identification of current limitations, and the formulation of the methodological design proposed in subsequent chapters.

3.1. EV Charging Behaviour

EV charging behaviour is shaped by a multifaceted interplay of technological, policy, and user-centric factors, which are integral to effective infrastructure planning and energy management. As the EVA continues to expand, gaining a nuanced understanding of these dynamics is essential for optimizing EVCI, reducing idle times, and enhancing the overall efficiency of EV charging networks. Lucas *et al*. (2019) utilized advanced machine learning models to predict idle times at EVCSs, providing key insights into optimizing resource allocation and enhancing station availability. Their work highlighted the necessity of precise forecasting to determine when charging points are likely to be occupied or idle, thereby facilitating better scheduling and improving user satisfaction.

Similarly, Sun *et al*. (2020) proposed a 5G-enabled smart grid architecture to enhance interactions between EVs and the electric grid. Their ensemble-based approach predicted charging behavior and supported dynamic pricing strategies, contributing to improved grid stability and operational efficiency. However, the complexity of implementing a 5G-enabled system and the potential scalability challenges associated with ensemble models limit the immediate practicality of their solution, particularly in regions with varying technological capabilities or infrastructure constraints. Fotouhi *et al*. (2019), on the other hand, developed a stochastic model to analyse EV drivers' behavior, emphasizing the importance of strategic infrastructure planning to meet charging demand effectively. Despite its utility, their model does not account for user demographics, such as age, income, or location, which are critical for tailoring infrastructure to diverse user needs and ensuring equitable access to charging facilities.

Wolbertus *et al*. (2018) explored the influence of policy measures, such as parking incentives and restrictions, on EV charging habits. Their findings highlighted the potential of behavior-focused initiatives to encourage optimal utilization of EVCI, reduce congestion, and minimize wait times. However, the study lacked a clear methodology for addressing data privacy concerns, which is a critical issue in the context of behavior monitoring and policy implementation. Qiao and Lin (2021) conducted a granular analysis of charging behaviours, focusing on temporal and spatial patterns to optimize urban EVCI. While their findings provided valuable data-driven insights for improving energy management and resource allocation, the overly specific nature of their models raises concerns about generalizability to different urban contexts. Akshay *et al*. (2024) emphasized the importance of accurate power forecasting using Seasonal Autoregressive Integrated Moving Average (SARIMA) models to address rising energy demand and prevent grid overload. Although their study demonstrated how precise forecasting can guide strategic decisions related to charging station placement and capacity expansion, SARIMA models may struggle to adapt to dynamic changes in demand, limiting their flexibility in highly variable environments.

Viswanathan *et al*. (2018) identified inefficiencies in the EVCI in San Diego and emphasized the importance of strategic planning to support government objectives for increased EVA and reduced greenhouse gas emissions. Their work underscored the critical need for well-designed charging networks that can accommodate growing demand while mitigating operational challenges. Li and Xiong (2023) employed a genetic algorithm-based backpropagation neural network (GA-BPNN) model to enhance load prediction accuracy for EVCSs, addressing fluctuations in charging demand. Their approach offers valuable insights into optimizing the distribution network's response during peak demand periods. Xu *et al*. (2022) and Zhou *et al*. (2022) contributed to understanding the spatial distribution of charging demand and managing uncertainties, with the former focusing on integrating traffic flow characteristics, and the latter employing Bayesian DL techniques to bolster grid stability.

Reddy *et al*. (2024) introduced a user-friendly EVCS finder application designed to help drivers locate available charging points efficiently. By integrating real-time data, their solution enhances user experience and addresses the increasing demand for effective EVCI. However, the scalability of their application poses a significant challenge, particularly when handling large volumes of big data from diverse sources. Ensuring seamless operation in high-demand urban areas or during peak usage periods may require additional system optimization and resource management strategies. Li *et al*. (2018), Hecht *et al*. (2021), and Bikcora *et al*. (2016) presented models utilizing hybrid, ensemble, and probabilistic approaches to forecast charging demand and enhance grid management, and demonstrated the effectiveness of advanced predictive models in improving the reliability and efficiency of EVCI. Chen *et al*. (2022) introduced the TWAFR-GRU model for real-time occupancy prediction, addressing key challenges in optimizing EVCI and ensuring effective utilization of charging stations.

Moreover, understanding energy consumption patterns in EVs has evolved significantly. De Cauwer *et al*. (2015) emphasized the importance of real-world data in predicting energy consumption for EVs, and demonstrated that factors such as driving styles and the integration of Intelligent Transport System (ITS) can significantly influence energy use, similar to how these factors affect fuel consumption in conventional vehicles. These insights are crucial for developing more efficient charging strategies and informing users about optimal driving behaviours to reduce overall energy consumption.

In addition to forecasting and consumption prediction, the integration of advanced algorithms for power allocation in hybrid systems is also pertinent in the emerging EV-related research topography. Although Wang *et al*. (2021) focused on a dual Proton Exchange Membrane Fuel Cell (PEMFC) and battery hybrid locomotive, their hierarchical power allocation method, based on an online extremum seeking algorithm, provides valuable insights into efficient power distribution in systems integrating batteries and alternative energy sources. This methodology has potential for adaptation in EV charging systems to optimize energy allocation during charging sessions. However, the approach's reliance on extensive data requirements poses a significant limitation, particularly in settings where such data may not be readily available or practical to collect. This restricts its immediate applicability to real-world EV infrastructure, where data constraints are often a key challenge. Addressing this limitation could enhance the scalability and utility of such power allocation strategies in broader contexts.

Chang *et al*. (2021) demonstrated that LSTM NNs can be used to effectively analyse and predict aggregated fast-charging power demand for EVs. Their findings indicated that accurate forecasting could inform optimal scheduling of public EVCSs, thereby minimizing energy costs and maximizing profit potential for operators, particularly in the context of fluctuating energy prices and the dynamic nature of EV deployment. This predictive capability is essential for energy providers to align supply with demand, especially during peak charging times.

These studies highlight the importance of understanding EV charging behaviour for infrastructure planning, policy formulation, and the promotion of sustainable electric mobility. By leveraging machine learning, big data analytics, and smart grid technologies, these approaches contribute to a more efficient and user-friendly EV charging ecosystem.

3.2. Instrumental Factors in EV Charging Behaviour

With increased worldwide interest in EVs, this is an increasingly popular field of research in many disciplines, including with regard to understanding charging behaviour and habits. There are some studies that have investigated the influencing factors in EV charging behaviour, implications for power systems, and the optimization of EVCIs.

The ley challenge in research on EV charging behaviour is the existence of randomness and variability in the users' charging habits. Studies have shown that peak demand for EV charging varies, considering that different drivers have different charging preferences, affected by time, location, and charging facility type (Dong et al. 2018; Qu 2018). Charging behaviour can be influenced by local conditions and user preferences. This variability poses challenges for power grid management, as uncoordinated charging can lead to peak demand spikes, particularly during evening hours when users return home from work (Morstyn et al. 2018).

Aside from this, the very nature and properties of the EVCI itself shape user behaviour. The fact that charging may come in various modes, including fast and slow charges, affects when and how the user prefers to charge. Velychko *et al*. (2022) highlighted that the standardization of charging mode contributes to user convenience, whereas Peng *et al*. (2014) observed that it would be necessary in the future to implement controlled charging behaviour in order to minimize negative impacts on the electricity grid. The coordination of charging schedules and smart charging have been suggested in various research studies for improving grid stability and efficiency (Morstyn et al. 2018).

Other psychosocial and decision-making aspects of EV users also have an influence on charging behaviour. Xing *et al*. (2020) introduced a data-driven model that incorporates human decision-making processes into the forecasting of charging demand. This approach acknowledges that users' choices are not solely based on technical factors, but also on personal preferences and situational contexts. Furthermore, studies have shown that the time and duration of charging events can greatly influence the overall performance of EVs, including their emissions and energy consumption (Wen et al. 2016).

A study by Anderson et al. (2023) about real charging behavior and preference of EV users in Germany illustrates critical factors influencing the habits of users. Home charging dominates due to convenience, while many perceive the development of public charging infrastructure to be underdeveloped. Other important factors are the immediate occupancy rate of charging stations, extended waiting times, and proximity to charging infrastructure. Basically, the study found that users are ready to walk 5-10 minutes up to a charging station. Some knowledge gaps by users in respect to charging technologies have also been exposed, indicating proper education and awareness.

Users with higher annual mileage or driving BEVs have so far shown higher acceptance for paying more money for faster-charging options. Openness toward the use of smart charging solutions exists; however, the ability for spontaneous travel should remain free. These types of findings give insight into how infrastructural availability, charging speed, and user education are all instrumental in setting trends in EV charging behavior and hence give recommendations toward actionable improvements in charging infrastructure and services.

Literature suggests that infrastructure planning will have to take into consideration the evolving user charging habits whatever the future developments. Morrissey *et al*. (2016) analysed charging behaviour in order to inform EVCI planning, and emphasizing holistically understanding user patterns to optimize placing and capacity at the charging station. Morrissey *et al*. (2016) highlighted that the available household data indicates that EV owners predominantly charge their vehicles at home in the evening, coinciding with peak electrical grid demand, suggesting that incentives may be necessary to encourage charging during off-peak periods.

Car park locations emerged as the preferred sites for public charging among EV users, while fast chargers exhibited the greatest utilization rates, suggesting that public fast EVCI is poised for commercial viability in the short to medium term (Morrissey et al. 2016). Power feeding charging stations with renewable sources of energy has also been suggested in order to increase the sustainability of and decrease the environmental impacts associated with EV charging (Lavrenova and Denysiuk 2023).

Overall, EV charging behaviour is a multidimensional phenomenon that considers users' preferences, infrastructures' capabilities, and power systems' implications. These will be important for continued research in the development of strategies that meet users' diverse needs and maintain grid stability and sustainability in continued growth in the EV market.

3.3. EV Charging Demand and Power Consumption

One significant approach to managing EV charging demand involves the use of advanced forecasting techniques. As mentioned previously, Chang *et al*. (2021)

showed how LSTM NNs can undertake the analysis and prediction of EV power demand related to aggregated fast-charging requirements and schedule optimization, reducing energy and financial costs for operators, particularly in light of fluctuating energy prices and the dynamic nature of EV deployment. This predictive capability is vital for energy providers to align supply with demand, especially during peak charging times.

Understanding EV charging demand and power consumption is critical for optimizing infrastructure, ensuring grid stability, and enhancing the overall efficiency of charging networks. Akshay *et al*. (2024) underscore the importance of accurate power consumption forecasting using SARIMA models, in order to prevent grid overload and guides strategic decisions regarding charging station deployment. This aligns with Viswanathan *et al*. (2018), who highlight inefficiencies in existing EVCI, and stress the need for well-designed networks that can meet the growing demand for EVs. Together, these studies underscore the pivotal role of predictive modelling in infrastructure planning and grid management.

Li and Xiong (2023) enhance load prediction accuracy through the use of a GA-BPNN model, providing valuable insights for optimizing the distribution network's response to fluctuating charging demand. Their work complements Zhou *et al*. (2022) utilized Bayesian DL techniques to manage uncertainties in charging demand and maintain grid stability. Integrating these predictive models into energy management systems is essential for ensuring a reliable charging experience and minimizing risks associated with high-demand periods.

Hecht *et al*. (2021) and Bikcora *et al*. (2016) presented probabilistic and ensemble machine learning frameworks that predict charging availability and demand, offering robust solutions for managing the complexities of EV charging networks. By integrating these approaches, energy providers can make informed decisions that enhance the efficiency of charging systems, reduce waiting times, and improve user satisfaction.

3.4. Related Work: Predicting the Occupancy State of EVCSs

Predicting the occupancy state of EVCSs is a crucial aspect of optimizing resource management, improving user experience, and ensuring the efficient utilization of EVCI. Accurate prediction models can address operational challenges, minimize customer wait times, and promote the effective use of charging networks, ultimately contributing to the widespread adoption of electric mobility.

One approach to predicting the occupancy state of EVCSs involves leveraging historical data to inform real-time decision-making. Alface *et al*. (2019) highlighted how an aggregative framework of historic data can be used to develop better predictions of available charging spots. They observed that drivers struggle with finding open charging stations, especially because of a lack of real-time data along the route. This framework tries to alleviate these uncertainties by using a collaborative prediction methodology that is mainly based on historical data, which, in turn, makes the whole charging process more efficient and guides the drivers to appropriate locations. However, a strong reliance on historical data, while ignoring real-time or external sources of data, may limit the predictive accuracy of the model. This reliance on historical data means that sudden changes in the occupancy trend and/or unplanned events cannot be accurately captured; hence, forecasts given to drivers cannot be timely and reliable.

Feng *et al*. (2022) proposed a hybrid DL framework for forecasting vacant parking space availability that can be directly applied in EVCS contexts. Their research emphasized the strong temporal and spatial correlations within the vacant parking space data. With the use of a dual “convolutional LSTM model combined with a dense convolutional network” (ConvLSTM-DCN), they were able to forecast parking availability over both short-term and long-term horizons. The model not only enhances the accuracy of the predictions but also helps reduce the time drivers spend searching for available spots, thereby reducing environmental impacts and generally improving the efficiency of transport systems. However, the proposed model may pose scalability issues across extensive charging networks, as their model is complex and computationally demanding. Models that require significant resources may prove impractical for applications necessitating real-time processing or rapid implementation in real-life automobile applications, which could constrain their efficacy, especially in geographically large or resource-limited settings.

Ma and Faye (2022) identified a hybrid LSTM NN technique to predict the occupancy of EVCSs, which is an essential ingredient of any real-time vehicle-to-charging station assignment system. Predictive models can enable charging service platforms to reduce customer waiting times and render infrastructure usage more efficient, thereby solving operational difficulties for EV fleets and also contributing to developing user-friendly applications for smooth charging. However, their model excludes ancillary data sources pertinent to occupancy, such as concerning traffic or weather conditions, user demographics that might influence charging behaviours, and details concerning how user behaviours impact occupancy. Without relevant external variables and an in-depth study of user behaviours, the overall accuracy and generalization of the prediction could be limited.

A more sophisticated approach was developed by Ostermann *et al*. (2022), who presented an in-depth analysis of charging point occupancy forecasting using various supervised learning algorithms. They used classification and regression methodologies for individual and site occupancy forecasting, hence providing significant insights into optimizing charging station utilization. This dual approach acts to enhance user experience and promote electromobility by underlining the need for real-world adaptable predictive models in charging scenarios.

However, none of the reviewed studies described above explained how their models maintain user data privacy. Since models requiring aggregated centralized data raise serious privacy issues, the lack of a clear methodology for maintaining the data privacy could be a great disadvantage, especially in regions with strict data privacy laws. This could thus be one of the determining factors in user trust and the adoption of the proposed methods.

Soldan *et al*. (2021) emphasized the application of big data streaming methodologies for the prediction of occupancy probabilities at EVCSs, underscoring the critical role of real-time data in enhancing the precision of forecasts. Through the integration of big data analytics, the research illustrates the imperative for prompt and effective energy management, resulting in heightened user satisfaction and optimized operations of charging stations. Specifically, the volume of streaming data in real time is crucial to manage the dynamic charging demand, while EVCI can adapt to dynamically altering conditions. While Soldan *et al*. (2021) has focused on big data streaming, delays in data processing are likely to make real-time implementation challenging. Network capabilities may prohibit the ability of real-time data transmission, and thus affect model performance. Big data streaming might be problematic in cases of poor connectivity, which would lead to latency and lesser accurate forecasts. These limitations underline the possible obstacles that exist vis-a-vis applying their methods universally, especially in regions with inadequate network infrastructure.

Douaidi *et al*. (2023) proposed a federated DL method for charging station occupation forecasting, stressing that the federated learning model with distributed sources can be used to improve the accuracy of the forecast. Federated learning is a decentralised machine learning methodology in which models are trained across numerous devices or servers containing local data samples, without the necessity of data exchange. This approach improves privacy and security while facilitating the synthesis of insights from various sources (Douaidi et al. 2023). The proposed federated learning framework avoids privacy problems, since the local data are processed locally but contribute to a general model; this brings decentralized data into improving predictive power. Such an approach will be especially valuable in the context of increasing EVA, where considerations for privacy and data security play a critical role. The federated model contributes significantly to the creation of intelligent systems that adapt to real-time occupancy patterns and further facilitate better resource management and user experience. However, while federated learning allows overcoming privacy issues, it introduces a set of different problems, such as data heterogeneity and increased communication overhead. Douaidi *et al*. (2023) realized that federated learning can be complex to implement effectively, due to challenges including ensuring consistent model performance across devices with heterogeneous data distributions, and variable compute capabilities, besides handling increasing communication loads imposed by model updates. Such challenges would, therefore, be part of the scalability and viability of using these models of federated learning on real EVCI.

Sao *et al*. (2021) introduced a sophisticated information fusion method to predict the occupancy of EVCSs by integrating multiple information sources: meteorological conditions, road traffic flow, and historical records. This broad approach augments the accuracy in occupancy prediction and presents a systemic framework that would allow real-time forecasting, improving prediction accuracy to facilitate better resource allocation and user experience, which are central for the effective management of EV charging networks. By integrating data from multiple sources, a more holistic view of occupancy-influencing factors is achieved, thereby increasing the reliability of predictions. However, Sao *et al*.'s (2021) application presents DLMs with issues regarding interpretability, whereby such models are commonly referred to as "black box".

Most DLMs are not interpretable, and hence it is difficult to understand how the decision-making process was reached for the predictions. This might cause them to be less trustworthy and reduce stakeholder confidence in understanding how conclusions are drawn. Also, Sao *et al*.'s (2021) deep information fusion approach might be challenging for model verification and operationalization. In real-world environments, complex models pose considerable problems in terms of their validation and trustworthiness because attaining consistent performance throughout may require exhaustive testing of a number of scenarios, possibly necessitating substantial computational resources. Such considerations could restrict the feasibility of deploying these models within operational EV charging networks.

Qiao and Lin (2021) proposed a data-driven approach, emphasizing the granularity of the charging behaviour to improve occupancy forecasting in EVCSs and contributing to more effective management of public EVCIs by increasing the accuracy of occupancy forecasts, and enabling adaptive responses to user demands. Their findings are crucial for developing intelligent systems that can respond in a timely manner to charging patterns and user needs to enhance the efficiency of charging services. However, a focus on the fine-grained aspect of charging behaviour may lead to overly location- or time- bound models. Although such models perform very well on the training data, they may not generalize well across different spatial or temporal settings. In other words, the predictive models face limitations in adapting to broader domains or different time periods with unique charging patterns. This restricts their general applicability and reduces their effectiveness.

Su *et al*. (2023) employed a Spatiotemporal Graph Convolutional Network (SGCN) to forecast charging station occupancy, effectively integrating spatial and temporal data to enhance prediction accuracy. Their model demonstrated the potential to optimize charging station utilization by anticipating occupancy patterns, particularly in high-demand urban areas. However, the high computational cost associated with SGCNs poses a significant limitation, potentially restricting their scalability and practicality in real-world applications. While the study addressed the challenge of spatial-temporal integration effectively, the computational demands may limit broader adoption, especially in resource-constrained settings.

Shaw *et al*. (2022) introduced a neighbour-based optimized logistic regression model for the prediction of charging station occupancy, thereby showing the potential of advanced statistical methods in occupancy detection. Based on a neighbour-based optimization model, local patterns of charging behaviours are modelled and thus help in improving resource allocation and user experience. The results of the study underlined the need to adopt statistical models that are interpretable yet efficient in addressing the complexities involved in charging station occupancy. However, this focus on local optimization alone could lead to their missing the bigger picture of exogenous variables affecting the system. Localized approaches tend to ignore key variables such as special events, changes in policy, or sudden changes in user behaviour, which can have major impacts on occupancy trends. In assuming the continuity of these patterns across time, the model is based on static assumptions that clearly reduce the strength of the model. This can then reduce the prediction accuracy if external conditions change, hence limiting the adaptability and overall applicability of the model in a dynamic setting.

Stein *et al*. (2024) analysed forecasting methodologies, considering data with privacy restrictions for workplace charging settings. They contributed valuable insights when developing models that work with limited data, which is crucial for workplace charging scenarios, while optimizing charging station utilization with privacy concerns. This is particularly important for corporate and institutional settings, where privacy considerations may affect data availability, but accurate occupancy forecasting remains essential. However, while Stein *et al*. (2024) discussed the challenges associated with privacy-restricted data, they did not fully address the technical and logistical hurdles of implementing predictive models in live systems. Deployment challenges such as integrating the models with existing infrastructure, ensuring real-time data processing, and managing user adoption are critical factors that can affect practical application. Besides that, the maintainability and update challenges regarding models being updated according to evolving patterns necessitates additional efforts. If left unattended, these aspects may make the models less accurate after some time and therefore not as effective in managing workplace charging station occupancy.

Motz *et al*. (2021) investigated occupancy patterns within the domain of civil engineering, providing essential insights that can be utilized for forecasting occupancy at EVCSs. Through an examination of occupancy behaviours and trends, their results can be employed to refine predictive models for EVCI, thereby aiding in achieving greater accuracy and more effective resource management. The application of civil engineering to the study of occupancy of charging stations brings a different perspective to the issue at hand, thereby enriching the area of EVI planning. However, models derived from their findings may require extensive data, which poses problems in regions where data infrastructure is still relatively limited. Thus, while Motz *et al*. (2021) may have valid findings, they may not be directly applicable in data-poor environments, hence limiting the general applicability of their findings. Another criticism of their work is that the prediction given by their models may considerably be changed by biased or noisy data. Errors and inconsistencies in data result in low reliability of occupancy predictions, rendering them invalid for resource management and planning purposes. Such limitations underscore the importance of ensuring data quality and availability when applying insights from previous studies for EVCS occupancy prediction.

3.5. Summary of the Literature

Taken together, the reviewed works present various ways to estimate EVCS occupancy, ranging from hybrid NNs and big data streaming techniques to federated learning and spatiotemporal modelling. Each of these methods has its own merits, which can contribute to improved prediction accuracy, handling real-time data, and addressing privacy concerns, for overall efficient charging network management and improved user experiences. By combining these approaches, charging service providers and policymakers have an opportunity to create more robust and flexible hybrid systems that serve evolving EV user needs while supporting broader goals of sustainable transportation. Table 3-1 summarizes the reviewed studies.

Table 3-1 Summary of Reviewed Literature

Study	Focus	Key Insights
Akshay *et al.* (2024)	Power forecasting using SARIMA models	SARIMA model guides station placement and capacity expansion. Forecasts may not adapt well to dynamic changes
Alface *et al.* (2019)	Framework leveraging historical data for charging spot prediction	Historical data limits real-time accuracy. Does not consider sudden occupancy changes
Bikcora *et al.* (2016)	Models using probabilistic approaches for grid management	Hybrid models enhance grid management. Privacy concerns not fully addressed
Cauwer *et al.* (2015)	Real-world data for predicting EV energy consumption	Driving styles and ITS affect energy use. Need for high-quality data for accuracy
Chang et al. (2021)	LSTM NNs for fast-charging power demand	Forecasting aligns supply with demand. Real-time adaptability not addressed
Chen *et al.* (2022)	TWAFR-GRU model for real-time occupancy prediction	Real-time prediction for optimizing station utilization. High computational demand for real-time predictions
Douaidi *et al.* (2023)	Federated DL for charging station occupation prediction	Federated approach improves prediction while maintaining privacy. Implementation challenges in live systems
Feng *et al.* (2022)	Hybrid DLM for vacant parking space availability	Spatial-temporal correlations improve prediction accuracy. Complex model verification in real-world settings
Fotouhi *et al.* (2019)	Stochastic model for analysing EV drivers' behaviour	Strategic infrastructure planning essential for charging demand. Lacks consideration of user demographics
Hecht *et al.* (2021)	Probabilistic and ensemble frameworks for charging availability prediction	Probabilistic frameworks manage charging network complexities. Complexity in model interpretability
Li and Xiong (2023)	GA-BPNN model for load prediction accuracy	Enhanced load prediction optimizes network response. Lack of interpretability of DLMs
Li *et al.* (2018)	Hybrid, ensemble, and probabilistic models for charging demand forecasting	Ensemble models improve EVCI reliability. Federated learning introduces communication overhead
Lucas *et al.* (2019)	Machine learning to predict idle times at EVCSs	Precise forecasting helps optimize resource allocation and user satisfaction. Focused only on historical data, not real-time sources
Ma and Faye (2022)	Hybrid LSTM networks for EV station occupancy prediction	Real-time prediction for seamless charging experience. Limited generalizability across different conditions
Motz *et al.* (2021)	Civil engineering insights applied to EV charging occupancy	Civil engineering concepts enrich occupancy prediction. Biased data may reduce model reliability
Ostermann *et al.* (2022)	Charging point occupancy forecasting using supervised learning algorithms	Dual approach for classification and regression enhances utilization. Scalability of federated model is challenging
Qiao and Lin (2021)	Granular analysis of charging behaviours for urban infrastructure optimization	Data-driven insights improve efficiency and user experience. Overly specific models may not generalize well
Qiao and Lin (2021)	Granular charging behaviour for occupancy forecasting	Granular behaviour data enhances adaptive response. Reduced accuracy in broader domains

Study	Focus	Key Insights
Reddy *et al*. (2024)	User-friendly EVCS finder application	Real-time data aids in locating available stations. Scalability issues with big data.
Sao *et al*. (2021)	Deep information fusion for EVCS occupancy	Fusion of diverse data sources increases prediction reliability. Limited application in data-poor environments
Shaw *et al*. (2022)	Neighbour-based optimized logistic regression for occupancy prediction	Local optimization improves resource allocation. Static nature reduces prediction reliability
Soldan *et al*. (2021)	Big data streaming for predicting occupancy probabilities	Real-time data improves forecast accuracy. Static assumptions reduce adaptability
Stein *et al*. (2024)	Methodologies for forecasting workplace charging with privacy restrictions	Privacy-restricted data use in workplace settings. Technical and logistical challenges in practical deployment
Su *et al*. (2023)	Spatiotemporal graph convolutional network for occupancy prediction	Spatial and temporal integration optimizes station availability. High computational cost
Sun *et al*. (2020)	5G-enabled smart grid architecture for predicting EV charging behaviour	Hybrid AI model enhances grid stability and pricing strategies. Complex implementation, potential scalability issues
Viswanathan *et al*. (2018)	Identifying inefficiencies in EVI in San Diego	Strategic planning to accommodate growing EV demand. May not capture real-time occupancy changes
Wang *et al*. (2021)	Hierarchical power allocation for hybrid locomotives	Adaptation potential for efficient power distribution in EV systems. Extensive data requirement limits applicability
Wolbertus *et al*. (2018)	Policy measures impacting EV charging habits	Behaviour-focused policy initiatives reduce congestion. No clear data privacy methodology
Xu *et al*. (2022)	Spatial distribution of charging demand	Traffic flow characteristics crucial for spatial distribution. Does not account for privacy concerns
Zhou *et al*. (2022)	Bayesian DL for managing uncertainties in charging demand	Bayesian techniques bolster grid stability. Network requirements for real-time data processing

3.6. Identified Research Gap

Despite the considerable body of research undertaken in the domain of EV charging, as discussed previously, several significant deficiencies persist. The majority of current investigations primarily emphasize the forecasting of charging demand and the enhancement of infrastructure planning (Chang et al. 2021; Li and Xiong 2023). The presented research provides substantial insight into forecasting power consumption and the accuracy of load predictions; however, it often neglects the real-time assessment of occupancy at the charging points, especially those in urban areas. Furthermore, studies such as those by Ma and Faye (2022) and Ostermann *et al*. (2022) focused on the prediction of charging station occupancy, and such analysis tends to focus on private sites or limited samples, whereby it does not give ample attention to public EVCI.

A further remarkable deficiency is the lack of inclusion of viewpoints of all critical stakeholders, like EV owners, while assessing and formulating predictive models. While Douaidi *et al*. (2023) emphasize the importance of privacy issues linked with federated learning, there is a comparative lack of attempts to address EV users themselves as contributors to the examination of the effectiveness and usability of these predictions. User preferences must be understood and then integrated into the assessment of predictive systems in order for the solutions to actually cater to user needs and be more widely adopted.

A number of recent studies have made substantial progress in EV charging station occupancy forecasting. For instance, Ma and Faye (2022) achieved prediction accuracies of 99.99% for 10-minute ahead forecasts and 81.87% for one-hour ahead forecasts using a hybrid LSTM model on data from Dundee, UK, demonstrating the value of feature separation in LSTM architectures. Similarly, Douaidi et al. (2023) applied federated deep learning techniques across distributed datasets and reported accuracy of 86.21% and F1-score of 91.49%, while maintaining user data privacy through decentralized model training.

Other studies, such as Ostermann et al. (2022), highlighted the effectiveness of using ensemble models such as XGBoost and Random Forests, reporting up to 40.8% improvement over naïve baselines when predicting 15-minute interval occupancy. Amara-Ouali et al. (2023) emphasized the benefits of hierarchical forecasting over multiple aggregation levels (station, regional, global), finding improvements in predictive accuracy when using this approach over traditional flat models.

Despite these advancements, direct comparisons remain challenging due to differing experimental conditions. The literature employs diverse datasets—some proprietary (e.g., ACN dataset with workplace focus) and others open-access (e.g., Belib dataset from Paris) which vary in data granularity, plug types, user behaviours, and temporal coverage. Metrics also differ: some works use accuracy and F1-score for classification (e.g., predicting station occupancy state), while others rely on regression metrics like MAE or RMSE for predicting total occupancy or charging duration.

In contrast, this research employs real-world data collected from three geographically distinct urban charging stations, focusing on comparative modelling of both classification (occupancy categories) and regression (total occupied slots). It introduces a multi-model deep learning framework including CNN, TCN, LSTM, GRU, and ensemble variants, tested under consistent conditions across all locations. Furthermore, it is one of the few studies to integrate user feedback on prediction usability, thereby bridging the gap between technical performance and perceived trustworthiness of the predictions. This dual methodological and user-centric focus helps position this study as a more holistic contribution in the field, while also highlighting the limitations of past studies that either overlook spatial variability or exclude user-involved evaluations.

In this respect, the aim of the present study is to overcome these shortcomings by focusing on the estimation of occupancy at public charging stations in cities, while incorporating the perceptions of EV users at the same time. This approach aims at creating a more user-oriented, inclusive, and flexible predictive model that will enhance the efficiency of public EVCIs and support the continued development of electric mobility.

Chapter 4: Research Design

Chapter 4 elaborates on the multi-layered research design that will guide this research toward its core objectives. The chapter provides a high-level overview of the research strategy, hence setting the methodological framework. Further into the chapter, Level 1 elaborates on the understanding and analysis of data, which describes the dataset in detail and analyses seasonality and other relevant characteristics. Next, Level 2 focus on initial modelling and quantitative user study in the form of a two-tiered process of model development and evaluation coupled with the investigation of user patterns. Finally, Level 3 presents the proposed BiGTCN model and a qualitative user study to interpret the findings obtained from the quantitative analysis and put them into context. Such an approach is needed to establish Chapter 4 as the basis for a rigorous, iterative investigation into the issues of occupancy and user behavior at EV charging stations.

4.1. Overview of the Research Strategy

Understanding the behaviour of EV users during charging sessions is critical for creating a predictive system that can reliably forecast the availability of charging places at public EVCSs. However, the varied and unpredictable character of EV users themselves makes it difficult to determine coherent patterns of charging behaviour required for optimal charging station scheduling (Chung *et al*., 2019). This study looks into the use of DL algorithms to provide an efficient mechanism that can help EV users manage their car charging schedules.

Previous research has found a low correlation between the required charge and actual charging by owners of small-battery PHEVs, owing to drivers reporting a scarcity of on-road charging facilities (Tal *et al.,* 2014). Furthermore, "range anxiety"—the worry of draining the vehicle's battery before reaching a charging station—contributes to overcharging and inefficient use of existing infrastructure (Cahour *et al.,* 2012). To solve these concerns, this study uses DLMs to analyse billing patterns, taking into account characteristics like time of day, location, and user preferences. By incorporating real-time data on charging station availability and traffic conditions, these models hope to deliver more accurate predictions and suggestions for ideal charging periods, lowering user anxiety and increasing effective EVCI utilization.

This study takes an inductive method, starting with the collecting and analysis of empirical data to identify underlying patterns and linkages. This methodology is well suited to the study’s exploratory nature, allowing for the development of new insights and theories based on the observed data. It encompasses three levels, as shown in Figure 4-1 and described below.

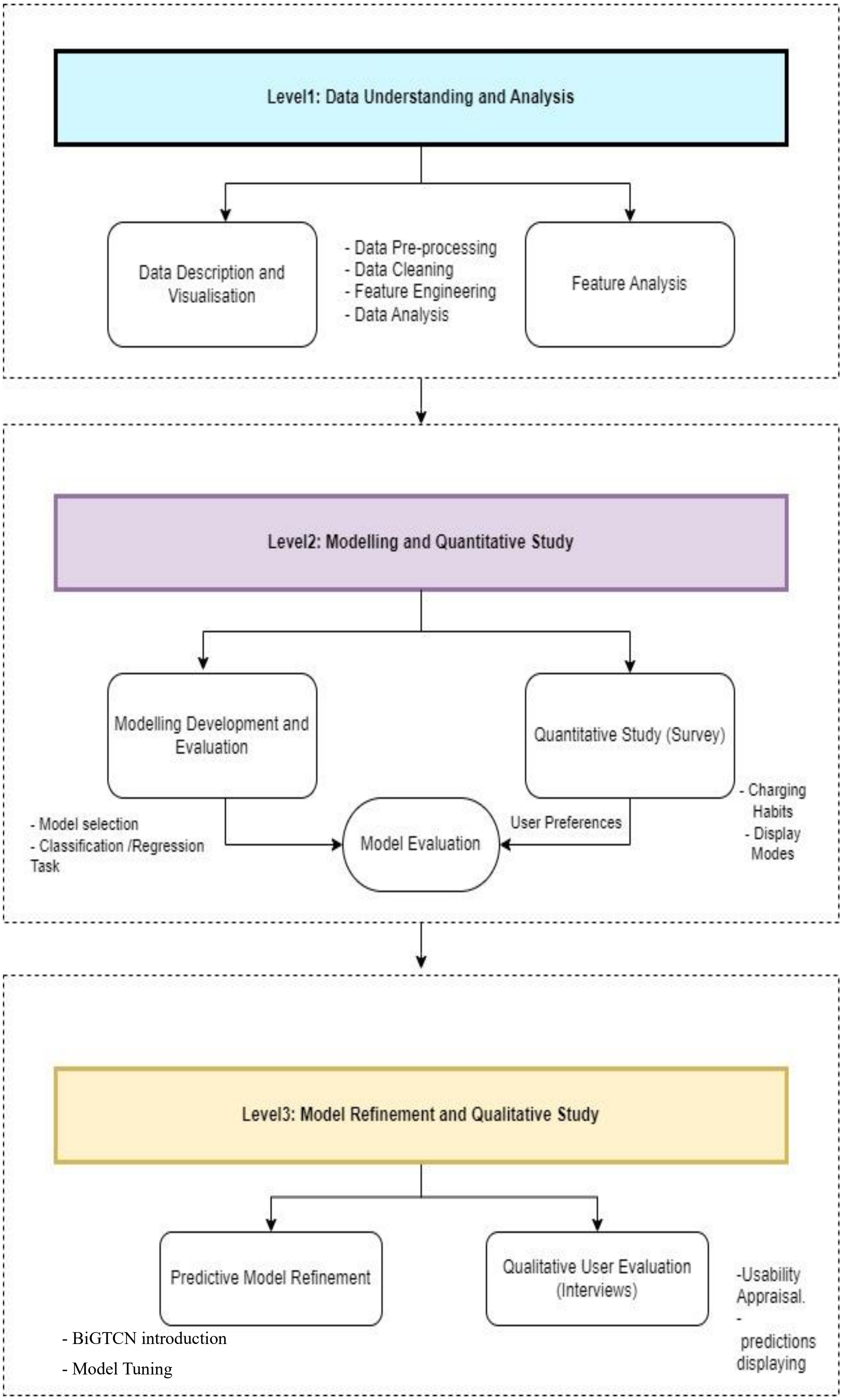


Figure 4-1 Research Design

Level 1: Data Understanding and Analysis

This stage consists of acquiring historical data from three charging stations, followed by data description and visualization. It contains feature analysis (including feature selection and importance) and investigates the effect of seasonal quarters on charging station utilization. The purpose of this level is to identify significant factors impacting occupancy rates and efficiently prepare the data for modelling.

Level 2: Modelling and Quantitative User Analysis

In this stage, two parallel processes are conducted:

- *Model Development and Evaluation:* DLMs, specifically LSTM, GRU, and 1D-CNN, are developed and trained to forecast EVCS occupancy. Classification and regression models are created to predict occupancy states and values, respectively. Model performance is rigorously evaluated to determine the optimum classification and regression models for the investigation.

- *Quantitative Study:* An online survey is conducted to obtain information about EV users' charging patterns and preferences for predictive model outputs. This quantitative analysis provides insights into user behaviour and preferences for displaying prediction results.

This level results in the selection of the best-performing classification and regression models, as well as a better understanding of user behaviour and output display preferences based on survey data.

Level 3: Qualitative User Evaluation and Model Refining

This level includes two parallel processes:

- *Qualitative User Assessment:* Qualitative research was performed via interviews with EV owners to evaluate the usability and trust in the predictive models. The interviews were formulated based on insights derived from the preliminary modelling and quantitative user study in Level 2. The qualitative research offers profound insights into user trust, usability challenges, and other obstacles to the adoption of predictive algorithms.

- *Model Refining:* Simultaneously, the best models identified at Level 2, further refined for improvement in generalization and accuracy. Improved classification and regression models were further evaluated and optimized.

Rationale

The sequential methodology of these stages capitalises on the advantages of each research strategy, hence augmenting the validity and reliability of the study findings. This study guarantees a thorough investigation of EVCS occupancy prediction by initially comprehending and arranging the data, subsequently constructing and assessing models in conjunction with preliminary user insights, and ultimately refining models through qualitative user input.

The following sections will elaborate on each research level, illustrating their cumulative contribution to the overarching research aims.

4.2. Level 1: Data Understanding and Analysis

Data understanding and analysis (DUA) research is highly valuable for gathering data, discovering patterns and associations in user behaviour (Mackey and Gass, 2015). In this research, the DUA study was initialized by gathering the relevant datasets related to EV charging operations that could be used during the research experiments. One of the major challenges encountered at this stage was assessing the suitability and completeness of the available data, particularly in terms of temporal coverage, data quality, and relevance to the research objectives. The primary dataset used in this research was obtained from Leeds City Council's open data repository via Data Mill North, which provides historical records of charging sessions from five public EV charging stations in Leeds.

Since databases about electric car charging operations were somewhat limited and mostly were incomplete, if they existed. DUA allows for the collection of real-world data on EV charging patterns, including monitoring charging station usage, peak demand times, and user behaviours (Schuemie *et al.*, 2014). This was achieved in this study by pre-processing the data, determining the components and its definitions, and performing the required analysis and statistical description and testing.

This comprehensive approach provided valuable insights into the current state of EV charging processes and helped plan the development of the future predictions. The DUA study also involved analysing charging station locations, types of chargers available, and their utilization rates. This information was crucial for identifying potential bottlenecks in the EVCI and areas where charging demand is consistently high or varies significantly, making them critical for understanding usage trends and informing infrastructure planning decisions. Additionally, the study examined the impact of various factors such as weather conditions, time of day, and special events on charging patterns. By gathering this diverse range of data, researcher was able to create a more holistic understanding of the EV charging ecosystem and its challenges, before commencing with modelling and simulation.

4.2.1. Data Description

One of the challenges faced in this study was providing historical records of the EV charging process to analyse EV charging behaviour and train the proposed DLMs for predicting the availability state of public EVCSs. Most of the available data either lacked sufficient coverage, or targeted highly specific users. Additionally, at the time of data collection, it was difficult to record live charging processes due to the country's lockdown strategy and the lack of public charging usage during the COVID-19 pandemic, which forced the researcher to rely on historical data. The final source for the study was the online repository maintained by Leeds City Council. Figure 4-2 shows the geographical location of the selected data sites of the five charging stations in the area of Leeds. Leeds City Council originally made the data available through a collaborative web portal Data Mill North[4] under the Open Government License v3.0. It provides data files quarterly, ranging from the second quarter of 2014 to the first quarter of 2021. The names and types of plugs used by each charging station are shown in Table 4-1.

[4] https://datamillnorth.org/

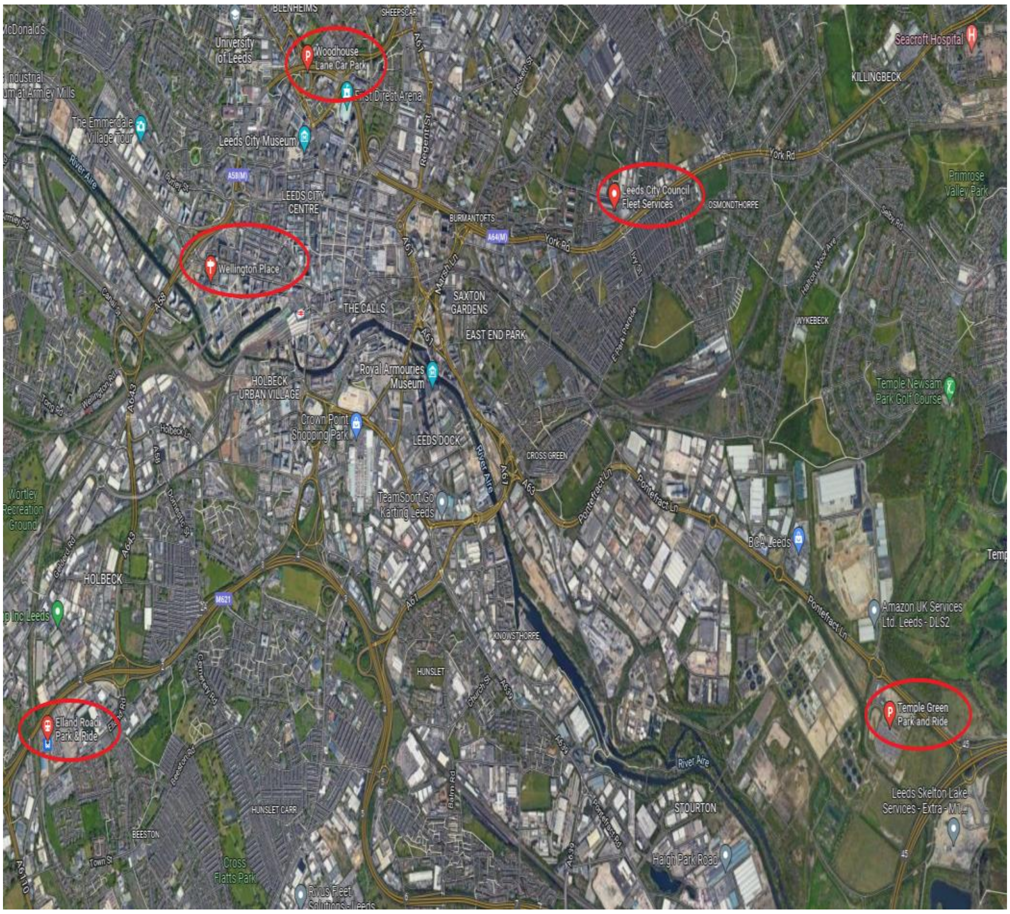


Figure 4-2 Geographical Locations of Selected Data Records

Source: Google Maps

Table 4-1 Names and Plug Models Used in Each Location

Site Name	Plug Model
Woodhouse Lane Car Park	APT 7kW Dual Outlet
Elland Road Park and Ride	APT 7kW Dual Outlet & APT Triple Rapid Charger
Wellington Place	APT 7kW Dual Outlet
Torre Road Council Depot	APT 7kW (PowerShare)
Temple Green Park and Ride	APT 7kW Dual Outlet

The three charging stations selected for this study were picked based on several critical criteria that enhance the credibility and relevance of the analysis. First, the stations have datasets with historical records from 2014 to 2021. This large data ensures sufficient representation to identify long-term trends in EV charging behavior. It also enables training and testing of DLMs under a wide variety of conditions.

Second, the selection of spatially heterogeneous charging stations within Leeds ensures the comprehensiveness of views in observing the EVs' usage patterns across different settings of an urban area. Those charging stations represent various public charging facilities, from busy urban city centres to residential areas, so that the research has broad coverage of charging scenarios and different user behaviours. This diversity is important because it would mean developing predictive models with enhanced generalizability.

Furthermore, the sample chosen has pragmatic advantages in relation to the data collection period. Given the challenges presented by the COVID-19 pandemic, such as reduced access to public charging points and limitations in collecting real-time data, the availability of reliable historical data became a major consideration. The data set provided by Leeds City Council through its Data Mill North portal proved instrumental in this respect. It provided comprehensive, quarterly documentation of electric vehicle charging sessions over several years, which gave the investigator an opportunity to review charging patterns and create models despite the limitations caused by the pandemic.

Lastly, by being open access, the dataset is licensed under the Open Government License v3.0, which guarantees the ethical and transparent use of data. Publicly available data is used in this study, meeting the reproducibility and transparency principles and therefore encouraging further research in this field. All these facts together validate the selection of the three charging stations and indicate their suitability to meet the objectives of the study.

For consistency, the three studied locations of Temple Green Park and Ride, Elland Road Park and Ride, and Woodhouse Lane Car Park are referred to as Location 1 (Loc-1), Location 2 (Loc-2), and Location 3 (Loc-3) (respectively) during the remainder of this research. The metadata of each quarter contains ten columns, as listed below:

- **Charging event**: This column is a seven-digit serial number to distinguish each charging event from the other.
- **User ID**: each user (EV driver) is identified by a user id and linked with its charging sessions.

- **Connector**: The connector number for the charging site of the charging session.
- **Start Date**: The start date of the current charging session.
- **Start Time**: The start time of the current charging session.
- **End Date**: The end date of the current charging session.
- **End Time**: The end time of the current charging session.
- **Total kWh**: The total Kilo Watt spent in the current charging session.
- **Site**: The site name and location of the charging point.
- **Model**: The model of the charging point.

4.2.1.1. Data Validity and Reliability

From the very outset of data gathering, or rather since data analysis itself, it became clear that observations up to the end of 2018 were incomplete and not representative for the sample in study. Validity, according to the quantitative paradigm, relates to the assurance that any given assessment will measure what it purports to measure (Sadık, 2019). Being that EVs had limited availability and EVCI was developing inconsistently during the duration, such had probably impacted the reliability of the data. For the integrity of the study and to avoid any likely bias, such data gathered within this period were excluded.

Similarly, data starting from the first quarter of 2020 were excluded to maintain the study's reliability. It was observed that there was a significant reduction in the number of EVs charging operations beginning at the end of the first quarter of 2020. These decreases coincided with the onset of the COVID-19 pandemic and subsequent lockdown measures, which significantly altered normal usage patterns. Given that data reliability is essential for producing consistent and dependable results (Elton-Chalcraft *et al.,* 2008), including this period would have introduced extraneous variables, thereby threatening the internal validity of the study.

Consequently, the analysis was restricted to data recorded in the year 2019 for three of the five charging stations initially considered, as displayed in Table 4-2. This selection was carefully made to ensure both the validity and reliability of the findings. The chosen data reflects consistent patterns of behaviour under typical conditions, accounting for both the number of recorded samples and the varying environments surrounding the selected stations.

Table 4-2 The Location Description of the Three Charging Locations

Location	Name	Distance to Leeds City Centre and Vital Roads	Neighbourhood Environment
Loc-1	Temple Green Park and Ride	13 mins./ 5.63 km to Centre 0.8 km to M1 0.32 km to an Amazon Warehouse	Industrial site.
Loc-2	Elland Road Park and Ride	9 mins./ 4.18 km to Centre 0.48 km to football stadium 0.48 km to M621	Mixture of industrial and residential areas.
Loc-3	Woodhouse Lane Car Park	3-5 mins./ 1.13 km to Centre 0.16 km to Leeds Beckett University 0.64 km to University of Leeds	A livelier area adjacent to a mixture of educational institutions and commercial centres, as well as residential buildings and a hospital.

4.2.1.2. Data Pre-Processing

4.2.1.2.1. Data Cleaning

During the data pre-processing stage, two primary issues were identified in the original dataset. The first issue involved the presence of null values, and the second related to the use of temporary, invalid start and end dates when the system failed to register certain observations. Given that the amount of missing data was minimal compared to the total number of successfully recorded charging events, the temporary invalid dates were first marked as null values. Subsequently, all invalid data entries were removed from the dataset to ensure the integrity of the analysis.

4.2.1.2.2. Data Transformation

To align with the research methodology employed in this study, the original dataset was transformed into a time series format. Although the original dataset provides some insights into the charging process, its structure requires a reorganization process to show a more meaningful structure for the purpose of this research. The main aim of the data transformation process was that transforming the records from event-based to time-based records. This transformation was necessary to facilitate the analysis and better support the time-series methods applied in this research. As a result, a new dataset was derived from the original data, ensuring it adhered to a time-series format with maintaining the original records.

4.2.1.3. Transformed Dataset Features

All processes of data pre-processing were performed in Python using the available libraries, such as Pandas, NumPy, and Datetime. The transformed dataset contains the following features:

"date": This field represents the observation time or the time series step, captured as a timestamp (Date + Time). This column was generated by initiating a sequence of dates, beginning and ending with the specific start and end dates identified in the original dataset. The time series data is sampled at an hourly frequency, meaning each row corresponds to one hour of observation.

"NewOccupied": A Python function was developed to calculate the number of EVs connected to a charging plug at a particular site at this timestamp.

"NewFreed": In contrast to the previous column, another Python script was used to determine the number of EVs that disconnected from the charging station during the current timestamp.

"TotalOccupied" (Target Value): This column captures the total number of occupied charging plugs at the station during the current timestamp. A Python script was used to count the number of active connectors at a specific charging site within the designated time frame.

Feature Engineering: Feature engineering is the process of obtaining features from raw data and converting them into a structure that fits into machine learning models. It is an important stage in the machine learning pipeline, since the appropriate features can alleviate the challenges of modelling (Zheng and Casari, 2018). Apart from the complex hidden layer in DL networks, feature engineering also can be useful approach in this field. Bengio *et* al. (2013) demonstrated that feature engineering is beneficial for speech recognition, computer vision, classification, and signal processing. Within the context of this research, feature engineering was employed to modify the initial dataset, facilitating a more significant analysis and augmenting the precision of occupancy forecasts for EVCSs.

The transformed data as described above when used as an input-data would contain “Date” as a predictor, and “TotalOccupied” as a target value. This might be too insufficient to capture the underlying temporal dependency that predicts “TotalOccupied’’ the total number of the occupied charging points. Therefore, the date was used to generate more features that would capture the time-related information which could enhance the predictive model in capturing relations and trends. The new extracted features were defined as follows:

“Day of the Week”: Extracted from the date for capturing weekly periodicity in charging behaviour.

“Day”, “Month”, “Week of Year”: All three of these features were created in order to capture daily, monthly, and seasonal variation in EVCS usage.

“Is Weekend”: This is a binary feature since whether the day falls on a weekend may indicate that users’ behaviour is different.

“hour”: Time of day.

“sin_hour”, “cos_hour” “sin_day_of_week”, “cos_day_of_week”: sine and cosine created trigonometric transformations of the former already express better the cyclic behaviour inherent in time, especially in daily and weekly regularities.

Besides these temporal features, the following derived features were used in enhancing the dataset:

“OccCate”: This feature is derived categorically to denote various occupancy classifications. The “OccCate” was used later as a target value for the classification task. The categorical coding of the occupancy states referred as an ordinal value from 0 to 3 for the categories (Empty, Partially Empty, partially full, full).

“AverageOccupied” and “AverageOccupied1”: These are average occupancy during similar periods in past occurrences and thus help the model find the repeating patterns in the charging practice.

“LocC1”, “LocC2”, “LocC3”: these binary variables define the location in which the charging station is located. The model can then find various charging stations, perhaps showing different patterns in usage.

Table 4-3 The Transformed Dataset Feilds

Column Name	Description
date	The full date and time of the record in format
TotalOccupied	Refers to the total occupied spaces in the station in the current stamp
Day of Week	Contains day number (0 to 6) for (Monday to Sunday)
OccCate	Refers to the occupancy category where (0-3) for (Empty, Partially Empty, partially full, full)
day	Extracting the day from the date
month	Extracting the month from the date
week_of_year	Extracting the week of the year from the date
Is Weekend	Extracting (0/1) for (weekday/weekend)
hour	Extracting the time:hour from the date
sin_hour	The periodic sine transformation of the hour
cos_hour	The periodic cosine transformation of the hour
sin_day_of_week	The periodic sine transformation of the day of the week
cos_day_of_week	The periodic cosine transformation of the day of the week
AverageOccupied	The average occupancy value in day of week and time of the day (Ex: The average occupancy value on Mondays at 7:00)

AverageOccupied1	The average occupancy category in day of week and time of the day (Ex: The average occupancy category on Mondays at 7:00)
LocC1	Sign flag for location 1
LocC2	Sign flag for location 2
LocC3	Sign flag for location 3
Temp	Weather data: Temperature
Prec	Weather data: precipitation
Code	Weather data: weather code
WindS	Weather data: Wind speed

Another important aspect of this analysis was the necessity to turn the input data from a univariate perspective to a multivariate one, including several features not limited to just the use of the date field itself. Moving to multivariate data brings great advantages with it: first, the possibility of discovering complex relationships among multiple variables and improving its generalization to different contexts (Zheng and Casari, 2018). By including multiple relevant features, the model gains a broader context, allowing it to better understand patterns in the data and improve predictive performance. Generating features in this process would input into the model far more meaningfully than just a date, thus enabling the model to capture the real temporal trends, location-specific variability, and patterns in user behaviour essential to accurate forecasting of the target values at EVCSs.

Table 4-4 and Table 4-5 summarize the descriptions of the target columns and the corresponding dataset averages (hereinafter referred to as the “CS2019” dataset).

Table 4-4 Statistical Description of the Combined Dataset (CS2019)

	TotalOccupied	AverageOccupied	OccCate	AverageOccupied1
count	26190.00	26190.00	26190.00	26190.00
mean	2.95	2.95	0.58	0.58
SD	3.61	3.25	0.86	0.71
min	0.00	0.00	0.00	0.00
25%	0.00	0.06	0.00	0.00
50%	1.00	1.60	0.00	0.13
75%	6.00	6.85	1.00	1.40
max	14.00	9.27	3.00	2.10

Table 4-5 Statistical Description of the Each Location (L1, L2, L3) in the Data

	TotalOccupied			**AverageOccupied**			**OccCate**			**AverageOccupied1**		
	L1	L2	L3	L1	L2	L3	L1	L2	L3	L1	L2	L3
count	8730	8730	8730	8730	8730	8730	8730	8730	8730	8730	8730	8730
mean	0.80	0.75	7.29	0.80	0.75	7.29	0.08	0.13	1.53	0.08	0.13	1.53
std	1.45	1.29	2.63	1.19	1.02	1.01	0.28	0.36	0.80	0.15	0.23	0.28
min	0.00	0.00	0.00	0.00	0.02	3.21	0.00	0.00	0.00	0.00	0.00	0.37
25%	0.00	0.00	5.25	0.00	0.06	6.85	0.00	0.00	1.00	0.00	0.00	1.40
50%	0.00	0.00	7.00	0.02	0.15	7.40	0.00	0.00	1.00	0.00	0.00	1.56
75%	1.00	1.00	9.00	1.81	1.46	8.06	0.00	0.00	2.00	0.12	0.17	1.73
max	8.00	9.00	14.00	3.44	3.48	9.27	2.00	3.00	3.00	0.50	0.87	2.10

In this respect, the dataset encompasses a large number of records, with 26,190 charging sessions spread over three locations (Loc-1, Loc-2, and Loc-3) hence providing a strong basis for future time-series analysis. The inclusion of such a large dataset ensures reliability and generalizability of the results over an extended period.

Table 4-5 shows considerable differences between the three sites, particularly in the “TotalOccupied” column, where Loc-3 has a rather high mean value (7.29) compared to Loc-1 and Loc-2, both having means less than 1. The latter would be interpreted to mean that Loc-3 has rather high occupancy rates, which may indicate some variance of charging demand, infrastructure, or user behavior in the compared sites. Moreover, the standard deviation (std) of “TotalOccupied” is much higher at Loc-3 (2.63), compared to Loc-1 (1.45) or Loc-2 (1.29), which would suggest that the range of occupancy behaviours observed at this location is wider.

Moreover, the quartile statistics bring out these differences. The 75th percentile figure for "TotalOccupied" for Loc-1 and Loc-2 is constantly 1, showing very low demand, while it spikes at 9 for Loc-3, emphasizing its role as a high-demand charging station. The differences like these show that analysis should be done with specific locations in mind to capture the unique charging behaviours and to build tailored predictive models.

By reorganizing the dataset to emphasize said variances and further combine the data in a meaningful way, CS2019 gives a good starting point for the goals of the study. The large differences in location and the size and width of the dataset ensure that the analysis is both comprehensive and relevant to real-world electric vehicle charging scenarios.

Furthermore, historical weather data was incorporated into the dataset to capture the potential effects of weather on charging behaviour. The weather data was sourced from the Weather Forecast API (https://open-meteo.com/en/docs) for Leeds City Centre, matching the date and hour of each record in the EV charging dataset. The features extracted from the weather data included:

Temp: Temperature in degrees Celsius.

Prec: Precipitation in millimetres, which could indicate rain or snow affecting driving and charging patterns.

Code: Weather condition codes as defined by the WMO Weather interpretation codes (WW). These codes describe weather conditions such as clear sky (0), rain (codes 61-67), snow (codes 71-86), and thunderstorms (codes 95-99).

WindS: Wind speed in meters per second, which might influence driving behaviour.

Including weather features allows the model to capture how different weather conditions may impact charging behaviour, such as reduced travel during heavy rain or snow. This integration provides a richer context for understanding fluctuations in the target values at charging stations.

4.2.2. Impact of Seasonal Quarters on Occupancy

Time series data often exhibits temporal patterns that can provide useful insights for predictive modelling, particularly in relation to user behaviour and system usage. While these patterns may include daily, weekly, or seasonal fluctuations, the ability to detect and quantify them meaningfully depends on the length and granularity of the dataset. In this study, the dataset spans approximately one year, which limits the ability to make conclusive statements about long-term seasonal trends. Nonetheless, an exploratory analysis was conducted to assess whether any preliminary seasonal effects might be visible within the year.

Specifically, statistical tests were used to explore variations in occupancy levels ("TotalOccupied") and occupancy categories ("OccCate") across four calendar quarters (Q1: Jan–Mar, Q2: Apr–Jun, Q3: Jul–Sep, Q4: Oct–Dec). This division was chosen to enable a broad-level view of changes across the year, although it is acknowledged that this form of aggregation may not fully capture the true seasonal or cyclic behaviour of EV charging. The aim was not to claim seasonality definitively, but to highlight any patterns that may warrant deeper investigation in future research with multi-year data.

Tests were chosen based on their ability to handle the characteristics of the dataset, such as non-normality or unequal variances, and included ANOVA and Kruskal-Wallis tests. The findings provide an initial indication of potential temporal variation, though they are interpreted with caution due to the single-year data constraint.

4.2.2.1. Normality Test (Kolmogorov-Smirnov Test)

The Kolmogorov-Smirnov Test was employed to verify whether the data followed a normal distribution. This test often appears to be a more effective test than the chi-square test for any sample size. Also, the Kolmogorov-Smirnov test is well-suited for large datasets, making it a robust choice given the substantial size of the dataset used in this study (Lilliefors, 1967). Due to the non-normal distribution detected across all locations and measures, non-parametric testing was deemed appropriate.

In addition to assessing the data distribution using the Kolmogorov–Smirnov test, this study also evaluated the stationarity of the "TotalOccupied" time series at each EV charging station. Stationarity is a critical property in time series modelling, as many forecasting techniques assume a constant mean and variance over time. To examine this, the Augmented Dickey-Fuller (ADF) test was applied, where the null hypothesis posits the presence of a unit root (i.e., non-stationarity).

4.2.2.2. Kruskal-Wallis H-test

Given that the Occupancy data did not meet normality requirements, the Kruskal-Wallis H-test was used to assess whether there were significant differences in occupancy (“TotalOccupied”) across different seasonal quarters (Q1, Q2, Q3, Q4). The Kruskal-Wallis H-test is a non-parametric test that compares medians across multiple groups, making it appropriate when data do not conform to normality and when testing multiple groups (Sheskin, 2003). This test was chosen because it does not require the assumption of normal distribution, is suitable for both ordinal and continuous data, and effectively identifies differences between multiple independent groups.

Hypotheses:

Null Hypothesis (H_0): There is no significant difference in occupancy (“TotalOccupied”) across the seasonal quarters.

Alternative Hypothesis (H_A): There is a significant difference in occupancy across at least one pair of seasonal quarters.

As will be seen in the following chapter, Kruskal-Wallis H-test suggests that there is a statistically significant difference in the distribution of “TotalOccupied” across the different categories of “quarter”. This test indicates whether there is a clear difference between at least one of the groups specified by the seasons of the year, but it does not indicate where exactly this difference between the groups is. Therefore, this test was followed by another analysis (Pairwise comparison), which highlights specifically where the difference between the groups is.

4.2.2.3. Pairwise Comparison of Quarters

Pairwise comparisons are performed to identify which of these quarters are significantly different in their distribution of occupancy. P-values are adjusted for multiple comparisons using a Bonferroni adjustment. One-way analysis of variance (ANOVA) was carried out for each site "Grouped into quarters of the year Q1, Q2, Q3, and Q4" based on "TotalOccupied". The Bonferroni adjustment entails multiplication of the original p-value of each test by the total number of comparisons made. This adjustment helps to protect the overall Type I error rate, which is the probability of making the wrong decision in rejecting the null hypothesis; that is, obtaining a false positive. In the context of multiple paired comparisons, these adjustments have to be made because when more comparisons are made, the greater the chance of observing at least one significant difference by chance alone. Modified p-value may be defined as follows:

$$p_{adjusted} = p_{original} \times n$$

Equation 4-1

$p_{adjusted}$: the adjusted significance value.
$p_{original}$: the original p-value for a given pairwise comparison.
n: the total number of pairwise comparisons.
- If the resulting value is greater than 1, it is set to 1.
- If the resulting value is less than or equal to 1, it is used as the adjusted p-value (Adj. Sig.).

The use of the Kolmogorov-Smirnov Test for normality, the Kruskal-Wallis H-test for comparing multiple groups, and pairwise comparisons for post-hoc analysis formed a comprehensive approach for assessing seasonal effects on occupancy at EVCSs. These methods provided robust, statistically validated insights into seasonal patterns, enabling a deeper understanding of how EV charging behaviour varies throughout the year.

4.2.3. Features Analysis

4.2.3.1. Features Selection

Feature selection is a critical step in building efficient and interpretable machine learning models. It involves selecting a subset of relevant features that contribute the most to the predictive power of the model. Among the various methods, embedded feature selection integrates feature selection within the model training process (Guyon and Elisseeff, 2003). One popular technique for embedded feature selection is Lasso regression.

Embedded Feature Selection Using Lasso

Feature Selection Lasso (Least Absolute Shrinkage and Selection Operator) regression is a linear model that uses L1 regularization to enhance the prediction accuracy and interpretability of the statistical model it produces (Tibshirani, 1996). The L1 penalty term encourages the coefficients of less important features to become exactly zero, effectively performing feature selection. The objective function of Lasso regression is:

$$\min_{\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}(y_i - X_i\beta)^2 + \alpha\|\beta\|_1\right\}$$

Equation 4-2

y_i: target variable.
X_i : feature vectors.
β: coefficient vector.
α: regularization parameter controlling strength of penalty.
$\|\beta\|_1$: L1 norm of coefficient vector.

The number of the selected features can be controlled by adjusting α, whereby small values of α provide less regularization and more features. This property makes Lasso particularly useful for feature selection in high-dimensional datasets (Zou and Hastie, 2005). The main advantage of using lasso for feature selection is that apart from reducing the model complexity by eliminating irrelevant features, the method allows integrating feature selection within the model training process. Also, the feature selection enhances the predictive models by making the trained model more interpretable (Tibshirani, 1996).

4.2.3.2. *Feature Importance* ***(Weather Data)***

Integrating weather features with charging data can potentially enhance predictive models by providing additional context that influences charging behaviours. While these inclusions are going to enhance the resolution of any predictive capability, they add to data complexity and can heighten the risk of model overfitting. Feature importance evaluation methods will be used to decide whether the weather features should be kept or excluded from the input dataset.

Granger Causality Test (GCT)

Amongst various potential methodologies, GCT was selected to establish if the meteorological variables predictively cause charging behaviours. This statistical tool assists in the selection of cases where past values of weather parameters yield the forecasting of future values of the charging data, and is useful in the decision-making process with respect to inclusion of weather-related information. GCT (Granger, 1969) seek to determine whether and at what significance level times series is useful in forecasting another.

GCT was employed in this study to check the historical data of the variable (meteorological variables like temperature, precipitation, weather code, and wind speed) provide predictive power to the response variable "TotalOccupied" which represents the occupancy rate of EVCSs. Such identification of causal relationships among those variables should help in improving the feature selection methodology to predict the availability of charging stations. GCT works by checking whether the past values of selected features "weather features" add significantly more information to the enhancement of the forecast of another series "occupancy" after accounting for the information provided by the historical values of the target variable itself. In case the relationship is significant, then one can conveniently conclude that the variable of prediction Granger-causes the target variable and hence contains information important to enhancing the capability of the model.

GCT was selected in this study because it allows the identification of which directional influence happens among time-series variables, something important for the current research, targeting the prediction of occupancy with regard to weather conditions at public electric car-charging stations. In such a way, all the included causal variables enhance the general accuracy of the predictive model by accounting for substantial interdependence. Specifically, several external factors, including weather features like temperature (Temp), precipitation (Prec), wind speed (WindS), and weather codes (Code), were considered as potential causal predictors. For instance:

- User decisions with respect to recharging their vehicle may be affected based on whether they prioritize comfort or driving efficiency. In particular, low temperatures can reduce battery performance.
- Precipitation and wind can therefore affect the use of the charging facilities, as bad weather will eliminate many people from engaging in outdoor activities who would otherwise do so, thus reducing demand for charging.

To validate these assumptions, the GCT allows for testing whether historical values of these weather variables can effectively explain the variability in occupancy over time. If these variables are found to Granger-cause “Total Occupied”, this indicates that they should be included as features in the predictive models to enhance their performance. This ensures that the explained variability by the Granger-causality relationship with “Total Occupied” is revealing, hence it can form the foundation for feature selection and formulation that might be used toward further improvement of the efficacy of the prediction models.

In this study, GCT was performed using the Python function *grangercausalitytests* from the *statsmodels.tsa.stattools* library. Testing was done on a statistical basis to determine whether one or more time series is useful in forecasting another Specifically, it was applied to examine whether the weather features (temperature, precipitation, weather codes, and wind speed) could Granger-cause the occupancy levels of EVCSs. Hence, based on the subsequent p-values from such analyses, the study attempted to empirically deduce if these variables were of any predictive power relative to the state of occupation. The following is a pseudocode rendering of the methodology used in executing the GCT using Python, as shown below:

4.2.3.3. Feature Importance Analysis: Light Gradient Boosting Machine (LightGBM) and Shapley Additive Explanations (SHAP)

To better understand the factors influencing occupancy predictions at EVCSs, feature importance analysis was conducted using LightGBM and SHAP values. This dual approach enabled an in-depth examination of the significance of various features and how they contribute to the model’s predictive power.

Feature Importance Using LightGBM

LightGBM, a gradient boosting framework, was employed to identify the most influential features contributing to occupancy prediction. LightGBM was chosen for its speed and efficiency in handling large datasets, making it ideal for evaluating feature importance in complex models (Ke *et al.,* 2017). The dataset contained various features, including time-related data (e.g., day, month, hour), weather-related data (e.g., temperature), and lagged temperature values. The LightGBM feature importance scores allow to determine which features had the greatest influence on occupancy levels, providing initial insights into the relationships within the data.

Comparing Results Using SHAP Values

While LightGBM's feature importance provides an overview of feature relevance, it does not offer a detailed understanding of how each feature impacts individual predictions. To address this limitation, SHAP values were utilized, which offer a more granular view, providing insights into both the magnitude and direction of the influence of each feature on model output (Lundberg and Lee, 2017). By using SHAP, it was possible to visualize not only which features were most influential overall but also how their values influenced specific predictions.

For instance, SHAP summary plots revealed how "AverageOccupied" consistently had a high influence across most predictions, with high values typically increasing the predicted occupancy level. Similarly, features like day and month were analysed to determine how their impacts varied across different scenarios. The colour distribution in the SHAP summary plot provided further insights: for features such as day, there were varied impacts depending on the specific day of the week, with certain days leading to higher occupancy predictions. The SHAP values also allowed the examination of complex interactions between features, providing a more comprehensive understanding compared to LightGBM's feature importance alone.

Justification for Combined Analysis

The combination of LightGBM feature importance and SHAP values enabled a robust analysis of feature significance from both a global (model-level) and local (individual-prediction) perspective. LightGBM offered an efficient and straightforward view of feature contributions, while SHAP provided interpretability at the individual prediction level, helping to build trust and transparency for stakeholders, such as EV users and charging station managers, by explaining model behaviour. Together, these methods facilitated a deeper understanding of the factors affecting charging station occupancy and allowed for targeted model improvements and enhanced interpretability.

Pseudocode (4-1) Feature Importance Analysis Using LightGBM and SHAP

```
BEGIN
1. IMPORT necessary libraries:
  - Import LightGBM as lgb,  Import SHAP library,  Import pandas as pd, Import numpy as np Import
        matplotlib.pyplot as plt
2. LOAD and PREPROCESS the dataset:
 - Load data into a DataFrame called 'data'
 - Perform necessary preprocessing steps (e.g., handle missing values, encode categorical variables)
3. DEFINE features and target variable:
 - Set 'target' as 'TotalOccupied'
 - Define 'features' as all columns in 'data' except the target variable
4. SPLIT data into features (X) and target (y):
 - X = data[features]
 - y = data[target]
5. INITIALIZE the LightGBM regressor model:
 - model = lgb.LGBMRegressor(n_estimators=100, random_state=42)
6. TRAIN the model using the data:
 - model.fit(X, y)
7. CALCULATE model performance metrics (optional):
 - Predict y_pred = model.predict(X)
 - Compute Mean Squared Error (MSE) and Mean Absolute Error (MAE)
8. EXTRACT feature importance from the model:
 - importance_scores = model.feature_importances_
 - Create a DataFrame 'feature_importance_df' with 'Feature' names and 'Importance' scores
 - Sort 'feature_importance_df' by 'Importance' in descending order
9. INITIALIZE SHAP explainer:
 - explainer = shap.Explainer(model, X)
10. COMPUTE SHAP values:
  - shap_values = explainer(X)
11. VISUALIZE feature importance using SHAP (optional):
  - Generate SHAP summary plot: shap.summary_plot(shap_values, X)
  - Generate SHAP dependence plots for important features (if needed)
12. INTERPRET results:
  - Analyze the most important features based on LightGBM importance scores
  - Use SHAP values to understand the impact of each feature on model predictions
END
```

4.3. Level 2: Initial Modelling and Quantitative User Analysis

Following the DUA study in Level 1, researcher focused on simulation and modelling processes in the Stage 1 of this level, as well as analysing the user charging behaviour and preferences in Stage 2 of this level. In Stage 1, primarily, the researcher investigated the utilization of diverse DL techniques to create prediction models. The decision to adopt this DL approach was made following the successful history of this method in dealing with providing future predictions for many similar fields in previous studies (Almaghrebi *et al.,* 2020; Becerra-Rico *et al.,* 2020; Frendo *et al.,* 2021; Hecht *et al.,* 2021; Ma and Zhang, 2018). Moreover, in Stage 2, as a parallel approach, the researcher conducted an online quantitative survey to collect data on the charging patterns of EV users and their preferences about predictive model outputs. This quantitative analysis offers insights into user behaviour and preferences, mainly, for the presentation of prediction findings.

4.3.1. Stage 1: Modelling Development and Evaluation

Simulations and modelling are highly valuable tools for comprehending intricate correlations and forecasting future observations. Simulation and modelling can effectively leverage improvements in the volume of accessible computations and data. Emerging learning methods and concepts for deep NNs further expedite this advancement (LeCun *et al.,* 2015).

During this stage, the researcher employed DL techniques to explore its capability to produce future predictions of EVCS occupancy states, thereby transitioning the study from descriptive analysis to predictive analytics. Specifically, the models were trained to forecast occupancy for up to six hours ahead (i.e., a forecast horizon of six time steps, with each step representing one hour).

To evaluate model performance over this forecast window, standard regression and classification metrics (including MAE, MSE, and R^2) were computed based on the average error across all six predicted steps for each time window. This approach captures the overall performance of the model across short-term horizons rather than focusing solely on the final prediction step. It reflects the average predictive reliability within the forecast range and provides a more robust estimate of model behaviour for operational planning at EV charging stations.

Stage 1 was mainly organized in order to answer research questions (**RQ1**, **RQ2** (specifically **RQ2-2**), and **RQ3**). For **RQ3**, the researcher divided Stage 1 into two phases, whereby Phase 1 focused on the classification task, and Phase 2 was organized to manage the regression task. The reason for split this stage into two phases was to assist the ability of prediction models in both tasks separately. For **RQ1**, a comparative experiment on the chosen five predictive models in each phase was conducted to assess their respective performances. The outcomes from this stage were defining the best-performed model in each phase.

4.3.1.1. Phase 1: Classification Models for the Occupancy State Prediction

In the context of predicting the occupancy state of EVCSs, the selection of classification models is critical for achieving high accuracy and reliability. Focusing on contributions to this respect, five different classifier architectures were targeted: 1D-CNN, TCN, LSTM, GRU, and an Ensemble model combining 1D-CNN and TCN. Each model was chosen for its robust performance in the classification tasks, particularly on time series data important for accurate predictions of the state of occupancy.

Research has shown that 1D-CNNs are able to yield high prediction accuracy with minimal or no complex pre-processing of the input data, as applied in numerous fields such as environmental monitoring and health diagnostics (Li and Li, 2022; Nannavecchia *et al.,* 2021). The architecture in general allows for computational efficiency even with low processing power devices, thus also making them suitable for real-time applications like occupancy prediction at an EVCS (Mathew *et al.,* 2023). Similarly, it was noted that the TCN model utilizes dilated convolutions to capture the long-range dependencies of the temporal data. This therefore avoids the problem of vanishing gradients found in most RNNs, which boosts the efficiency of the model in the consideration of sequence modelling tasks (Reza *et al*., 2022; Zhao *et al*., 2019). TCNs are also responsible for parallel processing; this, in turn, reduces computation time rather noticeably when compared with traditional recurrent models, suitability for tasks where decisions need to be quick-such as predicting the state of occupancy at charging stations (Reza *et al.,* 2022).

LSTM and GRU models represent the baseline in this comparison. Both models are well-established within the domain of sequence modelling because of their capacity to maintain information over extended periods. This feature is crucial in many applications relying on temporal dependencies (Reza *et al.,* 2022). LSTMs are highly complex models, having memory cells, which have given better performances for many time-variant classification tasks. On the other hand, GRUs are simpler, with fewer parameters compared to LSTMs, rendering them easier to train (Reza *et al.,* 2022). The inclusion of both models as benchmarks in this study allowed for a thorough evaluation of the performance of the more complex models (e.g., 1D-CNNs and TCNs).

The "Ensemble" model used in this study combines the strengths of both the 1D-CNN and the TCN in its attempt to further improve the efficacy in classification by availing the feature extraction capability of the 1D-CNN with the temporal dependency learning of the TCN. This hybrid will arguably yield a more robust model for both hierarchical feature extraction and handling of long-term temporal relationships. The rationale for this model combination strategy is based on findings in the literature on ensemble, where it is indicated that different architectures combined have better results for different applications in predicting the flow of traffic (Hecht *et al.,* 2021).

4.3.1.2. Phase 2: Regression Models for the Occupancy Value Prediction

In Phase 2, the focus shifts to predicting the actual occupancy values of EVCSs, treating this as a regression problem. Initially, this phase concentrated on two models, namely LSTM and GRU, due to their effectiveness in handling sequential data and capturing strong temporal relationships. Both architectures have demonstrated outstanding performance in learning complex temporal dependencies for regression tasks, as observed in the literature LSTMs have an architecture based on a memory cell, where long-term information can be kept, thus being able to capture long-range dependencies in time series. GRUs have a simpler structure, while rendering similar performance and completing training faster.

For benchmarking purposes, TCN, 1D-CNN, and an Ensemble model combining TCN and LSTM were used as baseline models. The TCN model, with its dilated convolutional structure, is particularly effective at capturing long-range temporal dependencies, making it suitable for comparative studies. TCN has shown strong performance in both classification and regression tasks due to its ability to model temporal dynamics effectively. The 1D-CNN model was added as a baseline since it is effective in feature extraction and computationally efficient even for contexts in which handling complex sequential data is not needed (Kiranyaz *et al.,* 2019).

The proposed model is an ensemble of TCN and LSTM, designed as a hybrid framework to leverage the strengths of both architectures. This approach combines TCN's capacity for handling temporal dependencies with LSTM's memory retention capabilities, creating a robust and effective ensemble for occupancy value prediction. A key assumption in this study is that the Ensemble model would outperform the baseline models by successfully capturing the temporal strengths of TCN, with LSTM providing a solid foundation for sequence learning. This integrated model is expected to capture occupancy data trends, both immediate and long-term, enhancing accuracy and generalizability in regression tasks.

In this phase, the Ensemble model serves as the primary proposed model, with LSTM, GRU, TCN, and 1D-CNN functioning as baseline models. This framework enables a comprehensive evaluation of various DL architectures in the regression task of forecasting occupancy values at each time step for EVCSs. Additionally, this facilitates a fair comparison to assess whether the ensemble strategy outperforms single-model architectures in effectively understanding and predicting occupancy behaviour.

For each of the phases described above; to address the sub-research question (**RQ2-2**), the experimental design of each phase was divided into two scenarios:

- **Scenario 1 (Scen-1): Combined Input Data**, where all models were trained on one combined dataset for all locations.
- **Scenario 2 (Scen-2): Location-Specific Input Data,** in which each model was trained on a specific location dataset.

Each scenario provides insight into the models' ability to produce predictions when trained on different datasets.

4.3.1.3. Evaluation and Selection of Models

Model performance evaluation has significant importance in identifying how well the selected DL architectures will do the forecasting of the availability of EVCSs. To be more specific, different metrics and evaluation techniques were proposed by researchers depending on whether the nature of the problem was classification or regression. Metrics used in each phase of the research are shown in detail herein, together with formulae and interpretations. All metrics calculations were obtained using the relevant libraries provided in Python. Specifically, libraries such as **scikit-learn** and **TensorFlow/Keras** were used to calculate metrics like accuracy, precision, recall, F1 score, MSE, and MAE, ensuring reliable and standardized evaluations across all models.

4.3.1.3.1. Phase 1: Classification Model Evaluation

The metrics used to evaluate the performance of the classification models in Phase 1 are described below: *accuracy*, *precision*, *recall*, *F1 score*, and *support*, together with the *confusion matrix*. All these metrics will be able to provide a wide view-especially the strengths and weaknesses of the models in making different states of EVCS occupancy (Powers, 2020).

Accuracy: Accuracy represents the ratio of correctly predicted instances to the total instances. The Accuracy calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Equation 4-3

Where: *TP* = True Positive; *TN*= True Negative; *FP*= False Positive; *FN* = False Negative.

Accuracy can be used as an overall metric that reflects how often the model makes correct predictions.

Precision: The ratio of correctly predicted positive observations to the total predicted positive observations; it is also known as the positive predictive value.

$$Precision = \frac{TP}{TP + FP}$$

Equation 4-4

Precision is usually particularly helpful where the associated cost of false positives is high since it tells how precisely a model can predict a class.

Recall (sensitive): Recall is the true positive rate, the ratio of correctly predicted positive observations against all actual positive observations.

$$Recall(Sensitive) = \frac{TP}{TP + FN}$$

Equation 4-5

Recall carries information when the associated cost of failure to detect, or cost of False Negatives, is very high, since it then is a measure of the capability of the model to find all relevant cases.

F1 Score: F1 Score is the harmonic mean of precision and recall. This immediately allows balancing both measures. The F1 score will turn out to be most helpful in the context of imbalanced classes as it considers both false negatives and false positives.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Equation 4-6

Support: Support is the total count of actual positives in each class present in the set; it reflects the distribution of the instances of the set among the different classes.

Confusion Matrix: A confusion matrix is a summary of prediction results where the number of true positives, true negatives, false positives, and false negatives are described inside of its matrix It offers a comprehensive assessment of the model’s efficacy across various categories, emphasising areas of strength and weakness.

4.3.1.3.2. Phase 2: Regression Model Evaluation

The performance for Phase 2 was considered with respect to continuous variables of occupancy at EV charging points. The performance of the models was computed using mean squared error (MSE) and mean absolute error (MAE) metrics, which relate to quantitative measures of deviation between forecasted values and their true observations (Willmott and Matsuura, 2005).

1. MSE considers the average of the square of the differences between the predicted outcome and their actual values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Equation 4-7

Where: y_i = Actual value; $\hat{y}_i$ = Predicted value; *n* = Number of observations.

2. MAE represents the average of the absolute differences between the predicted values and the actual values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Equation 4-8

Since the errors are squared, MSE punishes much larger errors as compared to the small errors. A lower MSE basically means a better fitting of a model. MAE is less sensitive to the outliers as compared to MSE. It intuitively measures the average model prediction error in the units of the original data. A lower MAE represents better predictive accuracy.

3. The R^2 metric make difference in the model performances evaluation as to how well the model explains the variance in the target variable "The coefficient of determination". R^2 is defined as:

$$R^2 = 1 - \frac{\sum_{1=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Equation 4-9

Where $\bar{y}$ is the mean of the observed values. The calculated R^2 shows a value from the range between 0 and 1, where, if closer to 1, it means the model will capture more variance within the target variable with the best fit and good predictability. According to Chicco *et al*. (2021), a high R^2 here signifies that this model can handle variation in the current level of occupancy and demand for EVCSs and helps provide deep insight into the performance of both short- and long-term predictions.

However, R^2 is limited as an indicator of model quality in that it inherently does not penalize model complexity. For example, with such a complex data set, especially when the data show very nonlinear relationships, its use has to be augmented with other metrics such as MAE and RMSE to make an assessment complete (Nagelkerke, 1991). Inclusion of R^2 in this approach provides an accessible and interpretable measure of fit, which ensures that the models chosen for EV demand and occupancy prediction are capable of capturing the underlying patterns in data in a way that can guide practical decision-making.

The different estimation metrics adopted for the two phases give a meaningfully complete basis for strengths regarding correct classification and occupancy predictions. First are the classification metrics, such as accuracy, precision, and recall, the F1 score, and the confusion matrix to identify model efficiencies on the basis of distinguishing states of occupancy. In Phase 2, MSE and MAE were employed in model performance testing with respect to accuracy in numerical value predictions; hence, weighing on the precision of the continuous output prediction. Individually, both of these metrics make sure that the model performance testing and validation are comprehensive, both for classification and regression tasks.

4.3.2. Stage 2: Quantitative User Analysis

4.3.2.1. Methodological Overview

Methodologies from quantitative users have become, especially through surveys, the required tools in studies at various levels in areas related to health, technology, and social sciences. Advantages that are included are generalizability, enrichment of data, and the capability to provide inputs for better decision-making. The establishment of generalizable data probably remains one of the strong points of the quantitative user methodologies in research. The quantitative studies could explain user awareness and attitudes across different demographics, especially within the field of ageing and health research, as noted by Kylén *et al*. (2020). Generalizability provides scope for the researcher to develop findings applicable for wider populations rather than limiting it to individual case studies. Moreover, structured surveys allow for statistical analysis to take place on data collection, which thereby provides valid findings (Schon *et al*., 2020).

This consequently allows for the emergence of patterns and trends that may not be visible in qualitative data, and such insight could help enhance the general understanding of users' experiences and needs. Aside from generalizability, quantitative user methods provide a degree of precision and objectivity which is usually absent in qualitative methods. Using surveys, for instance, allows researchers to quantify users' preferences and behaviour, thus creating an even more accurate representation of user groups. Such accuracy is particularly beneficial in design areas such as User Experience design, which focuses on enhancing customer satisfaction and usability by improving the overall interaction between users and products or services. The survey used in this study was designed to address **RQ4:**

> **RQ4:** "What display expectations do EV owners prefer to see for EVCS occupancy prediction?"

The responses from the quantitative survey allowed identifying general preferences across the population of EV owners. The following will provide the main process of the survey adaption in this research.

4.3.2.2. Using Online Surveys to Understand EV Charging Behaviour

The main aim of the survey was to gauge the general trend in the charging pattern of EV users and also to explore their preference regarding output displays of predictive models' outputs. The survey was designed using the online software Qualtrics. The survey questions and structure were checked by the researcher then have been validated and revised by the supervisor to insure the suitability of the survey structure to this research. After receiving approval from the University of Strathclyde Departmental Ethics Committee (Approval Number: 2018), the researcher conducted a pilot study with a small sample of eight participants to determine the validity of the survey. The pilot study aimed to assess and validate the survey structure, ensuring its suitability before the final version was published. The pilot study was excluded from the subsequent analysis. After implementing the necessary minor changes based on the pilot study feedback, the final version of the survey was published online, specifically targeting relevant EV owners.

Ads were placed in EV owners' groups on *Facebook* and *X* platforms to effectively reach the desired audience. Participant responses were collected anonymously, ensuring that no personal data was gathered. After around 20 days, from August 27th, 2022, to September 15th, 2022, the online survey was deactivated and the responses saved for the analysis. The survey encompassed three major sections to help find answers to certain specific research questions and hypothesis:

1. **Demographic and Operational Information:** This section gathered basic information about the participants, including their demographic data, as well as details about their typical charging habits and vehicle usage.

2. **Charging Habits of EV Users:** This section aimed to extract insights into the charging behaviour of EV drivers, such as their preferred charging times and locations.

3. **Prediction Display Usability:** This section was designed to gauge participant preferences regarding different formats of predictive model outputs.

The survey data analysis was performed using IBM SPSS Statistics software version 29.0.2.0(171), to ensure statistical accuracy and provide in-depth insights. The third section's analysis included an evaluation of the participants' preferences for various prediction formats. The findings from this analysis were intended to help better understand user needs and develop user-centric solutions for EVCI. The following hypotheses were considered with **RQ4** in the survey analysis:

> **H1:** EV users have preferred times of the day to charge their vehicles, which may vary based on the charging location (e.g., home, workplace, or public station).

H2: Participants generally prefer using online platforms to predict charging point availability and have specific preferences regarding prediction display modes.

H3: The likelihood of running out of battery or getting close to it is associated with the duration of EV ownership, as new EV owners often lack well-defined charging plans.

These hypotheses help provide valuable insights into EV user behaviour, and the analysis of these insights contributes to improving the prediction models and the presentation of charging station availability. The full version of the survey questionnaire is included in the appendices of this study.

During the evaluation stage, the results of the predictive models were assessed using both quantitative metrics (e.g., accuracy and precision) and quantitative feedback obtained from survey participants, specifically using Likert scale responses to evaluate views on different display formats. This dual quantitative approach allowed for a thorough evaluation of the model's performance and validation of any underlying assumptions related to user behaviour.

4.3.2.3. Association Between Preferred Charging Location and Preferred Charging Time (Chi-Square Test)

To evaluate the relation between preferred charging location and preferred charging time for EV users, chi-square test of independence was used. The primary aim was to determine whether the survey participants showed a statistically significant relationship between the preferred time of charging and the preferred location for charging. The chi-square test of independence is a non-parametric test that is used when both the variables are categorical. It is particularly effective in examining the association between variables that are not normally distributed or when the data is organized in contingency tables, as in this case. This statistical test was suitable here because both variables, charging location (home, workplace, public) and charging time (early morning, midday, overnight), are categorical. The chi-square test assesses whether observed frequencies in each category differ significantly from what would be expected by chance, providing insights into associations between categorical variables (McHugh, 2013).

The null hypothesis for the chi-square test is that there is no significant association between the preferred charging time and the preferred location for charging. Conversely, the alternative hypothesis suggests that there is a significant association between these variables. The chi-square statistic was calculated along with other symmetric measures, such as Cramér's V, which measures the strength of the association between the two categorical variables. Cramér's V ranges between 0 and 1, where values closer to 1 indicate a stronger association. This additional metric helps quantify the degree of association between the preferred charging times and locations. By using the chi-square test, the researcher was able to determine the significance of the observed relationship between categorical variables, providing insights into the patterns of EV charging habits among participants.

4.3.2.4. Predictions Display Preferences

An adapted version of the attitude scale used by Sözen and Güven (2019) was used to evaluate participants' views on different display formats for predictions, particularly focusing on their clarity and informativeness. The interval length is calculated as follows:

$$Interval\ length = \frac{R_{max} - R_{min}}{ls}$$

Equation 4-10

Where R_{max} is the maximum rating score, R_{min} is the minimum rating score, and *ls* is the Likert scale points.

In this study a five-point Likert scale was used. Therefore, from Equation 4-10 the interval length = 0.8. Table 4-6 shows the rating range used to measure the attitude of participants from the Likert scale in this study.

Table 4-6 Rating Range of Likert Scale of the Study

Dislike a great deal	[1:1.80)
Dislike somewhat	[1.80:2.60)
Neither like nor dislike	[2.60:3.40)
Like somewhat	[3.40:4.20)
Like a great deal	[4.20:5]

4.3.2.5. Battery Running Out Related to EV Ownership Experience

In testing this assumption about the depletion of batteries and their relationship to the ownership experience for EVs, responses given to the question about ownership experience were cross-tabulated. This cross-tabulation classifies the respondents according to how long they have owned an EV in trying to find an associated relationship with the level of frequency for having a problem with the car's battery. This will classify the data in such a way that any concealed pattern between the duration of ownership and the frequency of depleting such batteries is brought out.

4.4. Level 3: Model Refinement (Proposed BiGTCN) and Qualitative Study

At this level of research, work was done in parallel between developing the proposed predictive model from the results of the experiments conducted at the previous level and enhancing the study of the user experience and confirming the results of the questionnaire in the previous stage of this research.

4.4.1. Proposed BiGTCN

Since the results of the comparative experiments conducted in the previous phase of this study favoured the TCN model as the best classification model for predicting EV charging station occupancy, further improvements to its performance were explored in this stage. The objective was to achieve the most accurate possible predictions of charging station occupancy levels by addressing specific challenges identified in earlier experiments.

The proposed BiGTCN architecture enhances the TCN model by integrating a BiGRU layer. This hybrid structure is specifically designed to tackle the challenges of predicting EV charging station occupancy, where capturing both short-term and long-term temporal dependencies is critical for accuracy. While TCNs are efficient at modelling temporal patterns through dilated convolutions, their unidirectional nature may limit their ability to fully capture dependencies spanning both past and future time steps. To overcome this limitation, the BiGTCN incorporates a BiGRU layer that processes data bidirectionally, further improving the model's capacity to handle sequential data effectively.

4.4.2. Qualitative User Study

Through the survey conducted in the previous phase, key charging behaviours of EV Owners were identified, along with their evaluations of various formats for presenting predictive results. To investigate deeper, particularly in addressing the RQ5, a more focused qualitative study was designed to analyse users' confidence in prediction models and their reliance on these models for planning.

Given the need for clarity in participant understanding and the explicit communication of views to the researcher, the study was conducted through personal interviews. This format ensured rich, in-depth data collection while addressing the objectives of the research.

The interview questions were initially drafted by the researcher using *Microsoft Word* and subsequently refined after consultation with the research supervisor. Once finalized, the questions were digitized and prepared for administration using the *Qualtrics software*. Following ethical approval from the Departmental Ethics Committee (Approval Number: 2205), participants were recruited via e-advertisements disseminated across multiple EV communities and groups on Facebook and X platforms. Despite initial interest from fifteen individuals, eleven participants attended their scheduled interviews.

Interview Structure

The interview was structured into three distinct sections, each designed to address specific aspects of the research objectives:

4.4.2.1. Demographic Information

This section collected general demographic data about the participants to provide context for their responses and identify patterns in charging behaviours and perceptions.

4.4.2.2. Perceptions of Charging Accessibility and Acceptance of Digital Solutions

This section explored participants' views on charging accessibility and their acceptance of digital tools designed to enhance the EV charging experience. Participants rated their agreement with several statements on a five-point Likert-type scale. To analyse these responses, measures of central tendency (mean, median, and mode) were calculated using Qualtrics and Microsoft Excel.

Using central tendency measures in Likert-scale analysis, though typically ordinal, is a widely accepted practice in qualitative research to complement thematic findings (Boone and Boone 2012; Joshi et al. 2015). This mixed-methods approach provided a comprehensive understanding of participants' attitudes toward EV charging habits, challenges, and comfort with digital solutions. Quantitative summaries offered valuable insights into general trends, while the qualitative findings captured nuanced perspectives.

4.4.2.3. Testing Usability and Trust

In this section, participants were presented with three scenarios designed to evaluate their thoughts on predictive outputs and test their level of trust in the predictions. This section was analysed using Thematic Analysis, following the six-phase framework outlined by Braun and Clarke (2006). This rigorous approach ensured a systematic and reliable examination of the data, uncovering key themes related to usability and trust in predictive systems.

This section further investigates the results in Section (8.2.4), which shows the result of display modes rating. Each participant in this interview asked to try the following scenarios.

Scenario 1: Participants were shown predictions for the occupancy state in an EVCS for a certain time in the following day. These predictions were produced by the chosen model from the comparison experiments in Chapter 6.

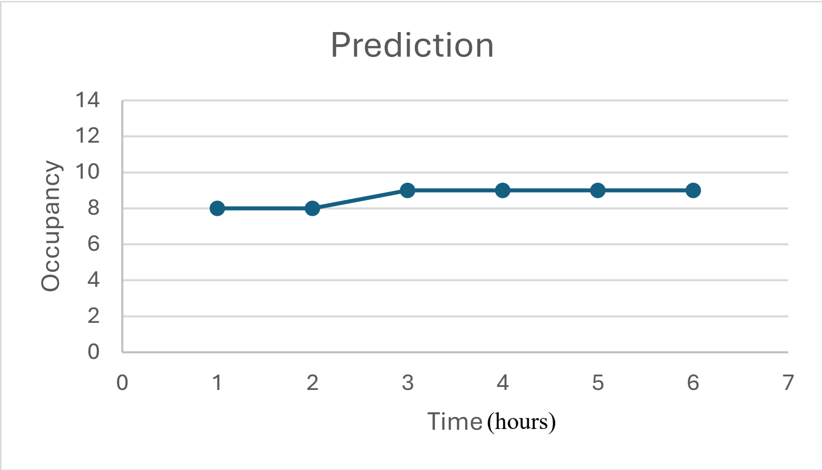


Figure 4-3 Sce1: Prediction of the Occupancy State (Graph-Display) in within A Certain (6) Hours

Scenario 2: Participants were shown a short-term prediction for the occupancy state in a certain charging point within the next six hours. The predictions, in this time, were shown in a form of a text expression.

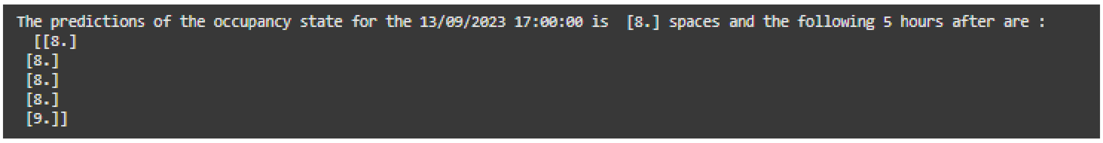


Figure 4-4 Sce2: Prediction of the Occupancy State (Text-Display) in within A Certain (6) Hours

Scenario 3: One of the previous scenarios was randomly chosen to be repeated, but this time the predictions were presented in a categorical description (Full, Moderate Full, Moderate Empty, Empty).

Each participant received the scenarios in a randomised ordering. In each scenario, Researcher asked participants the following questions:

What information you can get from this result, was this way of presenting predictions clear and simple?

Is there any further information that you think of that should be included to make the display clearer and more understandable?

What was your initial decision when you saw these predictions?

4.4.2.4. Trust in Model's Predictions

To assess users' confidence in the model's predictions, participants were instructed to reevaluate their judgements based on previously provided situations. They were presented with the model's predictions in conjunction with the actual recorded values from a previous scenario. This technique sought to determine if access to historical accuracy data could affect participants' trust in predictions and their decision-making processes. Participants expressed varying levels of trust in the predictive model, often contingent on the situation. Trust was higher when the predictions aligned closely with participants' prior experiences or when the situation was not critical.

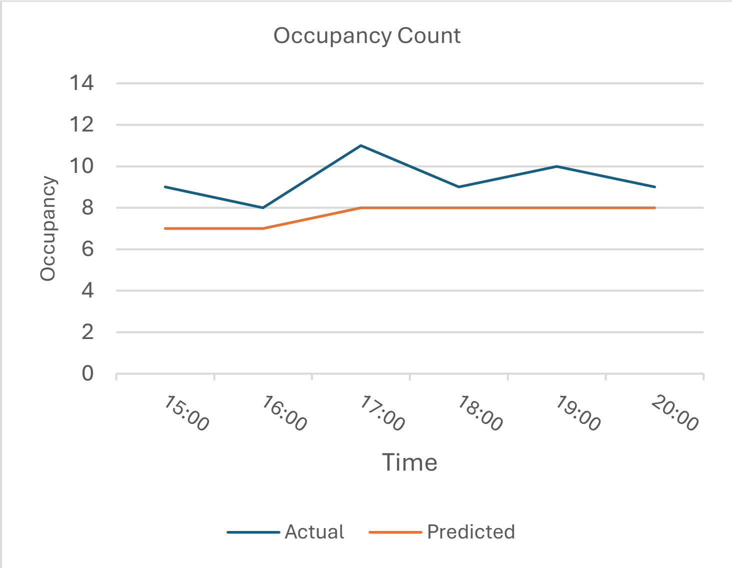


Figure 4-5Comparison of Actual and Predicted Occupancy Counts Over Time: Evaluating Users' Trust

The mixed-methods design of the qualitative study enabled a thorough exploration of EV owners' charging behaviours, attitudes toward predictive tools, and trust in forecasting systems. By employing personal interviews and a combination of quantitative and qualitative analyses, the study provides a rich and nuanced understanding of user perceptions, enhancing the broader implications of the research. The following chapters (5,6,7,8,9) shows the results obtained in each level of this research.

Chapter 5: Exploratory Data Analysis Results

The exploratory data analysis (EDA) stage is a rudimentary but fundamentally significant step in developing any kind of predictive model, whereby the researcher gets to understand the basic structure and key characteristics of data that have been collected. In fact, the main dataset considered for the purpose of this research technically forms the basis on which the predictive modelling efforts were carried out. By thoroughly examining and analysing this data, preliminary insights can be gained into the most significant factors that will influence the model's performance. According to Páez and Boisjoly (2022), for a modeller to obtain a reasonably good predictive model, a profound understanding of the nature of the raw materials is required. EDA allows the researcher to check for patterns, anomalies, test hypotheses, and also verify assumptions; the related information becomes important in making informed decisions on how pre-processing of data, feature selection, and choices of modelling techniques can be done.

This is where, in the preliminary phase, one strikes a base to build up a robust and correct predictive model, correctly matched to the peculiar attributes of a given dataset. Data visualization is one of the most important methods that can be relied upon in the field of data analysis. The human brain utilizes multiple parallel processing channels to efficiently detect patterns through the sense of sight (Franconeri *et al.,* 2021). From a data analysis perspective, this pertains to knowing that the process of collecting data and presenting its results to decision makers also includes the estimation and interpretation of the discrete choice model. Additionally, descriptive statistics serve as a means of condensing a data collection, allowing for a simplified viewpoint (Páez and Boisjoly, 2022). Descriptive statistics and visualization techniques are the primary tools used in EDA (Páez and Boisjoly, 2022). In this chapter the *CS2019 dataset*, including *the weather dataset*, was used to provide the necessary visualization and descriptive statistics, which were then used to reveal the underlying patterns, correlations and trends among the data.

5.1. Data Understanding and Analysis

This DUA study provides an initial understanding of the data's patterns or trends, serving as a precursor to subsequent deeper analyses performed to evaluate these insights.

5.1.1. CS2019 Data Visualization

The first section of the EDA stage focuses on different visualization for the CS2019 Dataset. The main objective of this section is to identify some indicators, such as relationships, trends, and external influences, that may contribute to the forecasting process. The aim of this section to provide insights about the answer to the following sub-research question (**RQ2-1**):

RQ2-1: Which specific temporal and environmental features improve the quality and relevance of the training data for predicting EVCS occupancy?

In this context, "quality" refers to the completeness, consistency, and reliability of the input features, while "relevance" refers to the strength of their association with patterns in the target variable (EVCS occupancy). By exploring how different temporal and environmental attributes align with occupancy trends, this section provides insights into which features enhance the model's ability to make accurate predictions.

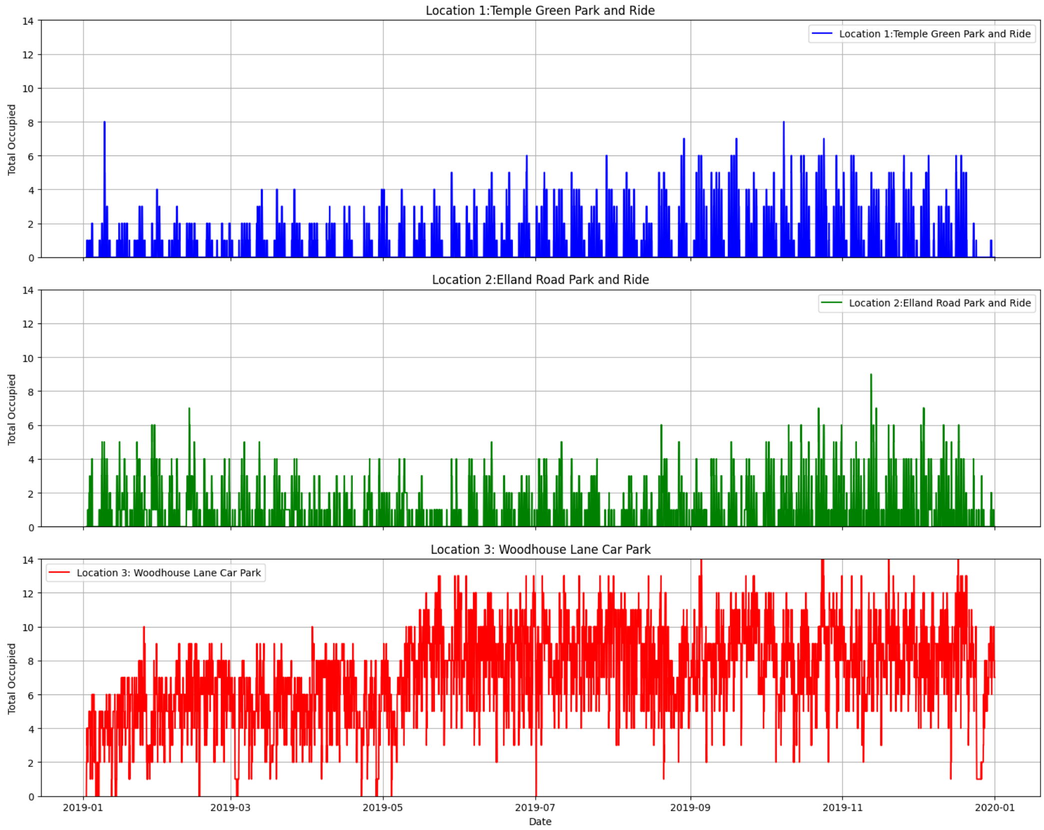
5.1.1.1. Effect of Spatial Features on Occupancy State

The indicator of the number of vacant charging points inside the electric charging station is the first criterion for the extent of the station's occupancy or availability to receive electric cars for the charging process. This field, whether it contains a number, a logical value, or a classification code, is considered a dependent variable or target variable. Since one of the most prominent goals of this study is to predict this value, the first steps in exploring the data are to display the values recorded in the past for this field and link them to some other fields in an attempt to come up with any possible indicators that support the process of building and evaluating a predictive system.

As stated in the data description section (4.2.1), the dataset consists of daily total occupancy data from three separate charging stations: Temple Green Park and Ride, Elland Road Park and Ride, and Woodhouse Lane Car Park. The data was collected during the year 2019.

To illustrate the temporal variation and pattern that characterizes each station, Figure 5-1 distinguishes each station separately.

- ***Temple Green Park and Ride*** displays a consistently stable occupancy with slight variations over time. This implies the presence of a stable and loyal group of users, either because of the dependable service or advantageous location. This location maintains a more constant usage with fewer sharp peaks, suggesting regular daily use by commuters or residents.
- ***Elland Road Park and Ride*** illustrates a decrease in total occupancy levels, which may be attributed to seasonal fluctuations or reduced demand. This location’s lower peaks may indicate underutilization or competing nearby alternatives that capture potential users.
- ***Woodhouse Lane Car Park*** exhibits a significant level of variability with distinct peaks, especially during the latter months of the year. This suggests intermittent utilization, potentially influenced by certain occurrences or seasonal periods of shopping, due to its potentially key position. This area, characterized by sharp and irregular peaks in usage, may be influenced by demand driven by specific events, indicating that this station caters to a user base that is more temporary or fluctuating in nature.



** Total capacity for station 1 and station 2 is (12). Total capacity for station3 is (14).*

** Graphs are line graphs for print clarity. They show, for example, station 1 and 2 regularly pulsing between 0 and 2 cars during each day and emptying at night while station 3 varying more and rarely completely emptying.*

Figure 5-1 Occupancy Records Over Time in 2019 for Each Charging Station

5.1.1.2. The Temporal Effect on the Occupancy Pattern

When analysing the occupancy pattern of EV charging points in the database (as described previously), apart from the possible spatial effect, the temporal effect is also worthy of attention. If it is proven that the time aspect has an effect on the occupancy rate, then it can be relied upon later to predict the occupancy status of charging points. At this stage, the occupancy data is displayed, and the possibility of inferring a direct or indirect relationship between the time variable and the occupancy rate is verified. Since the time variable can be multi-unit, this study will focus on the time as a quarter of the year, days of the week, and hours of day. The aim of this analysis to find out the possible effect of each time unit on the overall occupancy state. The limitations of the dataset, which covers historical data for only one year.

5.1.1.2.1. Days of the Week Observation

Days of the week may have a different impact on a charging station's overall occupancy rate. In order to investigate this, the variation in the occupancy rate of the sites in the study sample with the change in day have been monitored. Also, the occupancy rate between weekdays and weekends has been compared. These comparisons explained whether the difference in days of the week has an effect on the occupancy rate of charging stations.

Figure 5-2 provides a clear view of the average total occupancy for each charging station, broken down by weekdays and weekends, with separate lines representing each day of the week.

Figure 5-3 clearly illustrates a notable difference in occupancy patterns between Station 3 and Stations 1 and 2. This variation is likely attributed to the differences in geographical context and surrounding land use. Station 3 is located in a predominantly residential area, which results in higher occupancy levels during the night and early morning hours, followed by a decline during the day when residents are typically away. In contrast, Stations 1 and 2 are situated in areas characterized by industrial and educational activity, where demand for charging peaks during the daytime hours when employees and students are present. These contextual differences help explain the contrasting occupancy behaviours observed across the three stations.

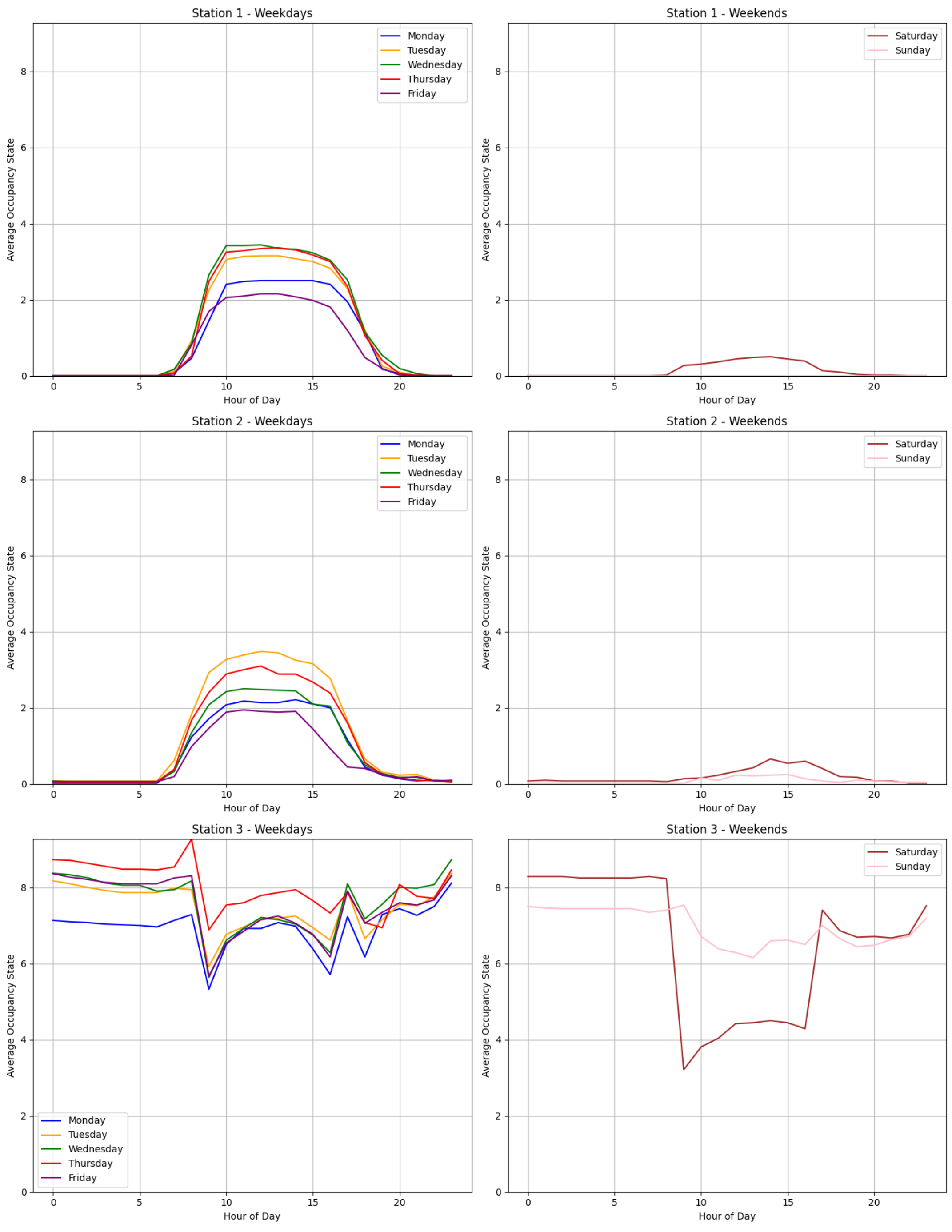


Figure 5-2 Average Occupancy State for Each Station During Weekdays and Weekends

Station 1: Temple Green Park and Ride (Stat-1)

Weekdays: The occupancy rate shows a strong, consistent pattern across all weekdays. The station experiences peak usage around midday, with a clear increase starting from around 8 AM and tapering off after 4 PM. The patterns are quite similar across the weekdays, with some variation in the peak occupancy where Mondays and Fridays show a slightly lower occupancy rate at peak time.

Weekends: The station shows very little activity during the weekends compared with weekdays, with occupancy remaining close to zero throughout the day. This suggests that this station is mainly used by weekday customers.

Station 2: Elland Road Park and Ride (Stat-2)

Weekdays: Similar to Stat-1, this station also shows a consistent pattern during weekdays, with occupancy peaking during typical travelling hours. The pattern across different weekdays is quite uniform, indicating regular use by travellers. Also, there are some variations in the peak occupancy where Mondays and Fridays again show a slightly lower occupancy rate at peak time.

Weekends: Like Stat-1, Stat-2 also shows very low occupancy during weekends compared with weekdays, reinforcing the idea that this station primarily serves weekday customers.

Station 3: Woodhouse Lane Car Park (Stat-3)

Weekdays: The occupancy rate is higher overall, and there are noticeable fluctuations throughout the day. The station appears to be in use throughout the day for all weekdays, with multiple peaks and troughs. While Monday is also relatively less busy than the other weekdays, Thursday, on the other hand, appears to be slightly higher in average specially before around (4 p.m.).

Weekends: Unlike Stat-1 and Stat-2, Stat-3 sees significant usage on weekends. The occupancy rate remains relatively high throughout the day, with some fluctuations. This suggests that the station is located in an area that attracts visitors or activities on weekends as well.

5.1.1.2.2. Quarters of the Year

Quarters of the year may have a different impact on a charging station's overall occupancy rate. In order to investigate this, the variation in the occupancy rate of the sites in the study sample with the change in quarters of the year have been studied.

This may explain whether different quarters of the year have an effect on the charging station occupancy rate. Figure 5-3 presents a comprehensive comparison between the actual occupancy category (OccCate) and the average occupancy category (AverageOccupied1) for three charging stations over various quarters. This facilitates the analysis of how occupancy patterns evolve over time and vary according to location. Particular quarters are analysed below.

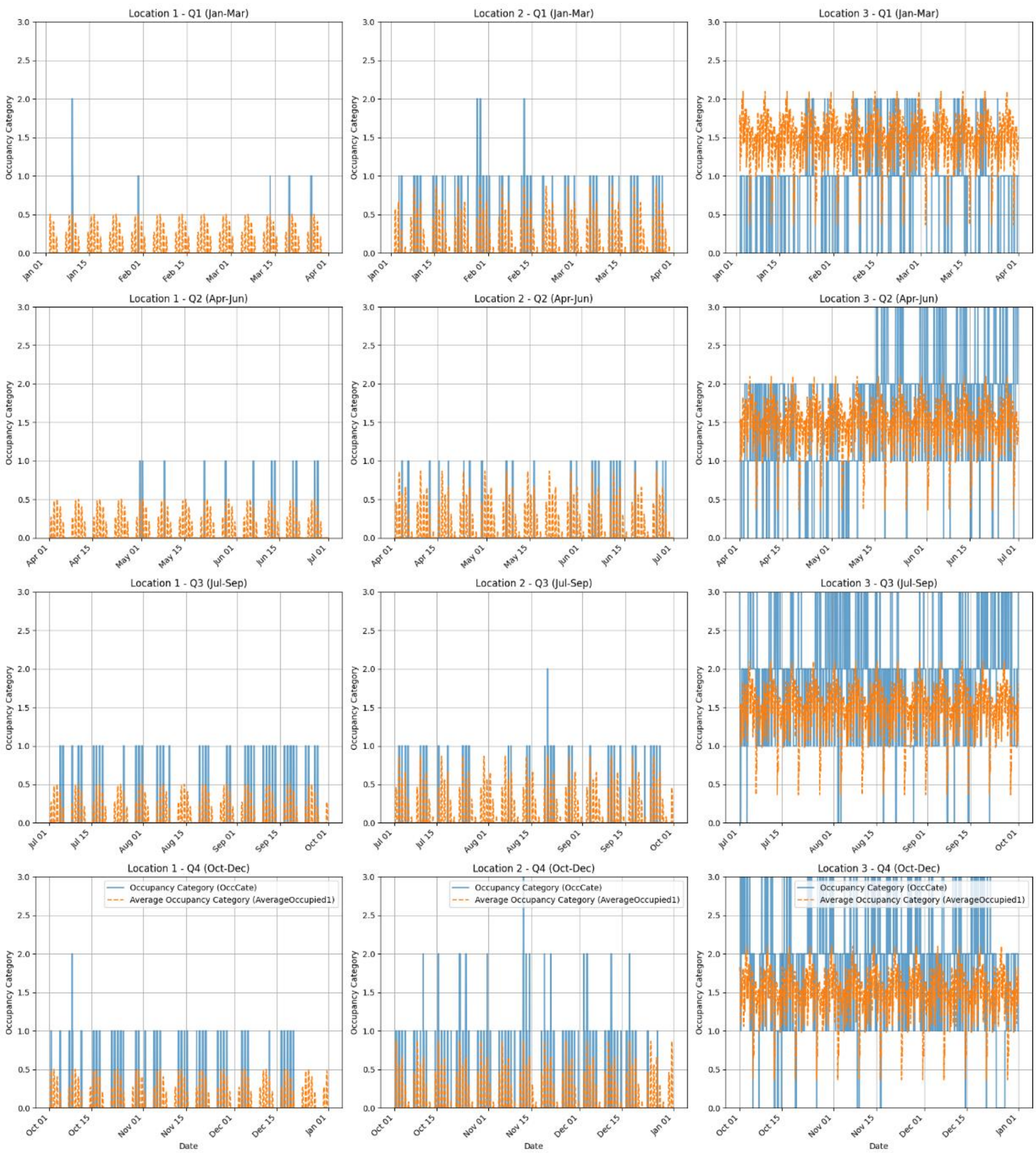


Figure 5-3 Comparison of Actual and Average Occupancy Categories Across Different Quarters for Three Charging Stations

Q1 (Jan-Mar)

The occupancy rate in Stat-1 and Stat-2 are generally low, especially within Stat-1, with most values lingering around the 0 and 1 range. The actual value of the occupancy categories closely matches the average value, suggesting a stable and predictable overall utilization rate. This is positive for this type of study. Furthermore, the study reveals a limited number of peak points that do not significantly deviate from the average values of the occupancy categories. The actual occupancy category at Stat-2 is slightly more variable than the average, indicating some usage fluctuations. Stat-3 exhibits markedly greater occupancy compared to the other two stations. The occupancy pattern exhibits greater dynamism, suggesting a busier location. Furthermore, the first half of the first month of this quarter has a lower occupancy rate than the rest.

Q2 (Apr-Jun)

The pattern in both Stat-1 and Stat-2 is generally not much different from what it was in the first quarter, as the occupancy rate is still low in most periods. There is no significant deviation from Q1. This indicates that consumption habits remain unchanged during the first half of the year. This quarter, the consistent usage in Loc-3 has not changed. Additionally, the station is still in high demand throughout this period. During the second half of this quarter, the occupation rate increased more.

Q3 (Jul-Sep)

Occupancy has experienced a slight increase in comparison to the first and second quarters, especially for Stat-1, but overall, the trend has remained stable. generally, both stations still cater to a small number of users with a low rate of utilization. Stat-3 continues to maintain the high occupancy rate it achieved at the end of the second quarter. This high rate continued until the end of this quarter.

Q4 (Oct-Dec)

Although Stat-2 saw a very slight increase compared to the first three quarters of the year, Stat-1's occupancy rates remained nearly the same as in the third quarter. However, the occupancy rates for both stations are still low, clearly indicating weak demand for these terminals during this year in general. The two stations witnessed a significant decrease in occupancy as the last days of the year approached. Although the last few days of the year saw a relative decrease in occupancy at Stat-3, most of this quarter saw the station maintain its high occupancy rate throughout the year, making it a critical hub for EV users in the area.

Whole-Year Quarterly Data

The following figures display the quarterly data over the course of the whole year.

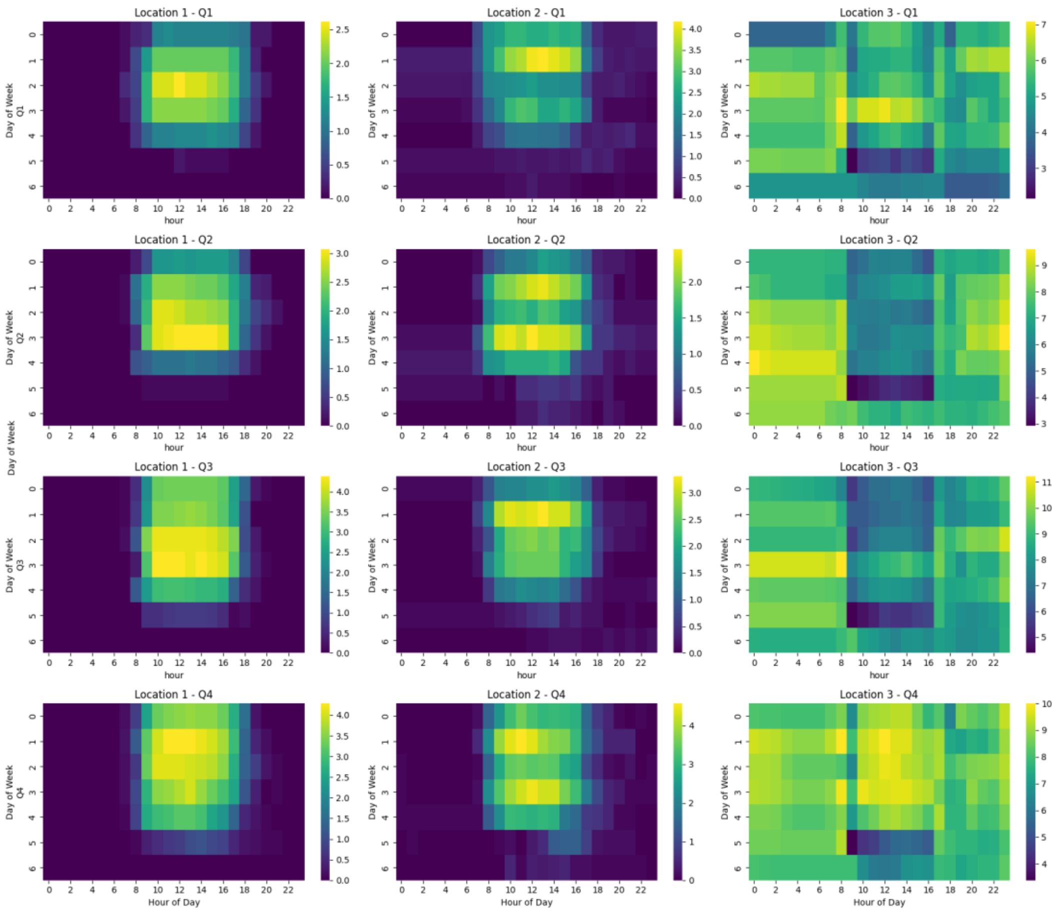


Figure 5-4 Total Occupancy for Each Station for Each Quarter of the Year

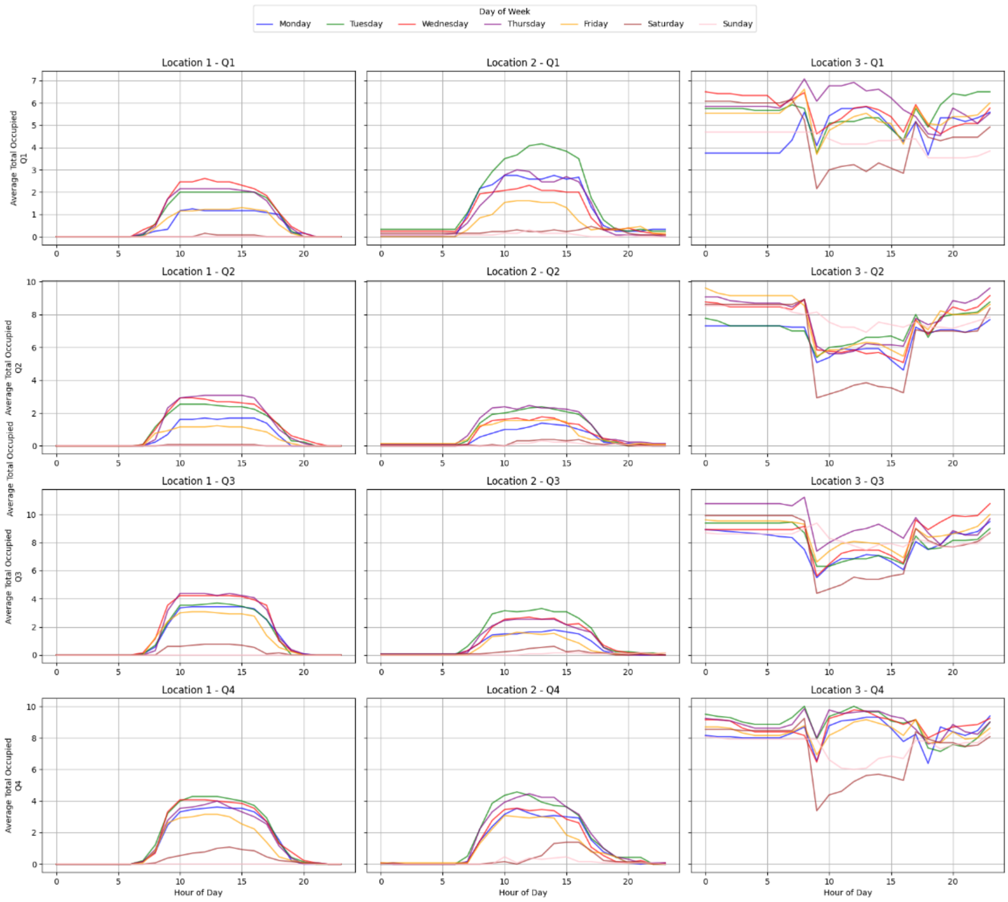


Figure 5-5 Average Total Occupancy of Each Station for Each Quarter of the Year

5.1.1.2.3. Hour of the Day

Hours of the day may have a distinct impact on a charging station's overall occupancy rate. In order to investigate this, the variation in the occupancy rate of the sites in the study sample with the change in hours of the day have been discussed. This may explain whether the hours of the year have an effect on the charging station occupancy rate. The occupancy category graphs in Figure 5-6 show that, during weekdays, two key transition periods in the occupancy rate categories can be observed. The first period is between (6:00 am) and (8:00 am), while the second period is between (2:00 pm) and (5:00 pm). In Stat-1 and Stat-2, the occupancy rate increases and begins to decline in the second period. In contrast, in Stat-3, the occupancy rate begins to decrease in the first period and gradually increases in the second period. However, returning to the increase in the second period in Stat-3 takes longer and fluctuates more. Because of the very low occupancy during the weekends, especially in Stat-1 and Stat-2, this variation is absent, and the general pattern remains flat.

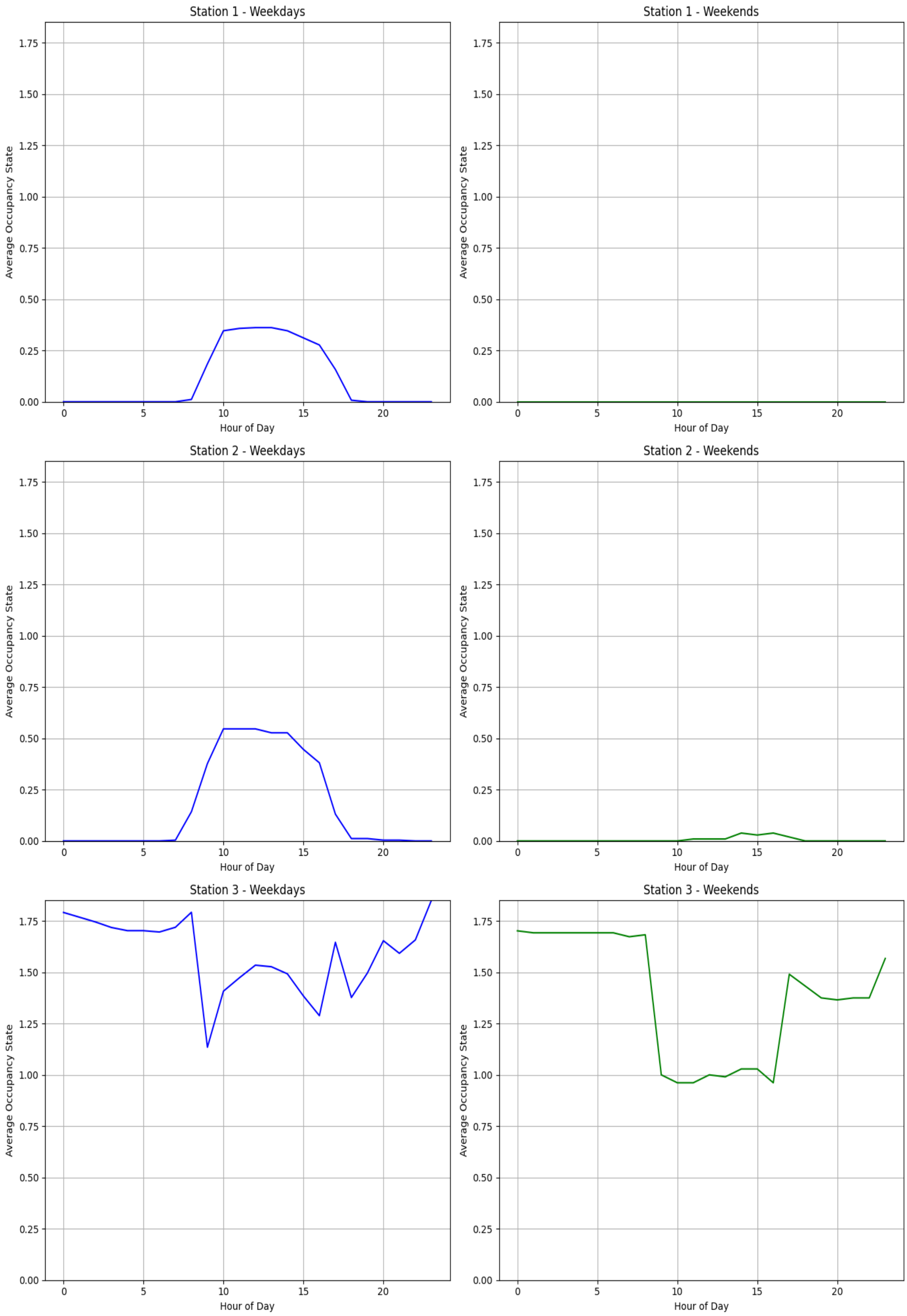


Figure 5-6 Average Occupancy Category During the 24 Hours Weekdays and Weekends

5.1.2. Statistical Analysis on Occupancy

Statistics is a scientific discipline that focusses on gathering, arranging, and analysing data, as well as making conclusions about the entire population based on samples (Ali and Bhaskar, 2016). It is necessary to carefully plan the study, choose a good sample, and select an adequate statistical test. In this research, the researcher used both descriptive statistics and appropriate statistical tests to support some of the data presentation results as well as to evaluate some hypotheses about the study data.

5.1.2.1. Descriptive Statistics

Descriptive statistics analyse data by grouping them into variables to determine the typical values and the range of values for each variable in a dataset. Measures of central tendency, such as the mean, median, and mode, are statistical summaries that represent usual values (Guetterman, 2019). Measures of variability, such as variance, standard deviation (SD), and range, are used to represent the spread of values. Descriptive statistics collectively offer insights about the data's distribution and the frequency of values within the dataset, as demonstrated by a histogram graphic (Guetterman, 2019). These descriptive statistics are foundational in understanding the overall structure of the data before proceeding to more complex analyses.

Occupancy per Station

Given that the forecasting process primarily focusses on the target variables ("TotalOccupied", "OccCate"), this section provides a statistical description of these variables for each electric charging station. Where:

N (Sample Size): The number of observations in the dataset for the respective variable.

$$N = count\ of\ non-mising\ values\ in\ the\ dataset$$

Equation 5-1

Minimum: The minimum value in the respective value of the dataset.

$$Minimum = \min\ (data)$$

Equation 5-2

Maximum: The maximum value in the respective value of the dataset.

$$Maximum = \max{(data)}$$

Equation 5-3

Mean: The sum of all values divided by the number of values.

$$Mean = \frac{\sum_{i=1}^{N} x_i}{N}$$

Equation 5-4

Where x_i is each individual value in the dataset.

Std.Error: The standard error of the mean is the SD divided by the square root of the sample size.

$$Std.Error = \frac{Standard\ Deviation}{\sqrt{N}}$$

Equation 5-5

SD: The SD is the square root of the variance. It is a measure of how spread out the values are in the dataset.

$$Std.Deviation = \sqrt{\frac{\sum_{i=1}^{N}(x_i - Mean)^2}{N}}$$

Equation 5-6

Variance: The variance is the average of the squared differences from the mean.

$$Variance = \frac{\sum_{i=1}^{N}(x_i - Mean)^2}{N}$$

Equation 5-7

The following tables (Table 5-1, Table 5-2, and Table 5-3) display the corresponding statistics of the occupancy variable for each site. These descriptive statistics tables, produced using the SPSS platform, confirm the higher occupancy count for Stat-3 compared with Stat-1 and Stat-2. The occupancy categories also vary, with Stat-2 and Stat-3 having higher maximum categories compared to Stat-1.

Table 5-1 Stat-1 Records Descriptive Statistics of Stat-1 Records

	N	Minimum	Maximum	Mean		SD	Variance
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
TotalOccupied	8730	0	8	.80	.016	1.453	2.112
OccCate	8730	0	2	.08	.003	.276	.076
Valid N (listwise)	8730						

Table 5-2 Descriptive Statistics of Stat-2 Records

	N	Minimum	Maximum	Mean		SD	Variance
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
TotalOccupied	8730	0	9	.75	.014	1.292	1.670
OccCate	8730	0	3	.13	.004	.356	.126
Valid N (listwise)	8730						

Table 5-3 Stat-1 Records Descriptive Statistics of Stat-3 Records

	N	Minimum	Maximum	Mean		SD	Variance
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
TotalOccupied	8730	0	14	7.29	.028	2.633	6.931
OccCate	8730	0	3	1.53	.009	.798	.637
Valid N (listwise)	8730						

5.1.2.2. Statistical Tests: Impact of Seasonal Quarters on Occupancy

At times, descriptive statistics by themselves are insufficient to provide a complete and initial understanding of the researched sample. Thus, the researcher employed inferential statistics, which provide stronger insights than descriptive statistics for making inferences and generalizations that apply to a larger group. Selecting the suitable statistical test for hypothesis testing requires careful consideration of several essential factors. Nevertheless, it is crucial to choose the suitable statistical analysis prior to commencing the study, during the planning phase, and ensure that the selected sample size is optimal. These cannot be determined capriciously after the studies have concluded and data has already been gathered (Nayak and Hazra, 2011).

The purpose of the statistical test in this section is to assess whether the occupancy ("TotalOccupied") at EVCSs is significantly influenced by seasonal quarters (Q1: Jan-Mar, Q2: Apr-Jun, Q3: Jul-Sep, Q4: Oct-Dec).

5.1.2.2.1. Normality Test (Kolmogorov-Smirnov Test)

Firstly, occupancy value has been checked for following the normality distribution by performing Kolmogorov-Smirnov Test. Table 5-4 shows the result of the Kolmogorov-Smirnov test of normality for the "TotalOccupied" variable across three different locations ("LocationID" 1, 2, and 3). The Kolmogorov-Smirnov test is a test used to assess whether a given sample comes from a normally distributed population. The null hypothesis for this test states that the sample data follows a normal distribution.

Table 5-4 Result from Normality Test (Kolmogorov-Smirnov Test)

LocationID	Statistic	p-value
1	0.393	<0.001
2	0.645	< 0.001
3	0.979	<0.001

Analysis for Loc-1 (LocationID = 1): It is shown at Loc-1 that the Kolmogorov-Smirnov statistic is (0.393), with Sig. is less than (0.001). Therefore, the p-value is less than the level of significance of (0.05). The null hypothesis in this case can thus be rejected. The result obtained here implies that the "TotalOccupied" at Loc-1 is significantly deviated from normal distribution. That is, the level of occupancy measured at Loc-1 does not follow a pattern of normal distribution.

Analysis related to Loc-2 (LocationID = 2): For Loc-2, the Kolmogorov-Smirnov statistic is (0.378) with a significance level of less than (0.001). Thus, the p-value is less than (0.05). Therefore, the null hypothesis can be rejected. It can, therefore, be inferred that the occupancy distribution at Loc-2 has a significant deviation from normal distribution. It can be said that the occupancy pattern at Loc-2 may be subjected to some anomalies that may vary across time or prevailing conditions.

Loc-3 (LocationID = 3.00) Analysis: For Loc-3, the Kolmogorov-Smirnov statistic is (0.102), with a significant value of less than (0.001). Although this statistic is lower than those observed for the previous locations, the significance value remains less than (0.05), leading to the rejection of the null hypothesis. This means that the distribution of "TotalOccupied" at Loc-3 also significantly deviates from a normal distribution, indicating non-normal characteristics in occupancy.

Since the p-value for each site is below (0.001), results for the Kolmogorov-Smirnov test conducted at all three sites suggest that "TotalOccupied" is not normally distributed. This is expected due to the anticipation of nonlinear and complex drivers of occupancy across sites, hence requiring nonparametric approaches (or indeed any other kind of approach that does not consider normality while analysing or modelling this kind of data). In all cases, the rejection shows that further analysis has to take into consideration the non-normality of the data, where more advanced methodologies may be required to handle such skew or irregular distributions. Therefore, to test the assumptions related to this data, Kruskal-Wallis H-test is used to assess whether the occupancy "TotalOccupied" at EVCSs is significantly influenced by seasonal quarters.

The results of the ADF test confirmed that the time series data for each station is stationary. Specifically, Station 1 yielded an ADF statistic of -11.42 (p-value < 0.001), Station 2 yielded -12.37 (p-value < 0.001), and Station 3 yielded -7.25 (p-value < 0.001). In all cases, the p-values were significantly below the 0.05 threshold, allowing the null hypothesis to be rejected and affirming that the series does not require differencing or transformation to achieve stationarity.

These findings support the direct application of forecasting models on the raw data without prior differencing, simplifying the preprocessing phase while confirming the data's readiness for time series modelling.

5.1.2.2.2. Kruskal-Wallis H-Test

Since the data is not normally distributed, the Kruskal-Wallis H-test was used to assess whether there are significant differences in occupancy between the quarter groups for each location. The Kruskal-Wallis H-test was chosen due to the non-normal distribution of data, the need to compare multiple groups (quarters), and the suitability of the test for continuous data. It provided a robust and reliable method for testing whether the occupancy "TotalOccupied" varied significantly across different seasonal quarters.

Hypotheses

Null Hypothesis (H_0): There is no significant difference in occupancy ("TotalOccupied") across the seasonal quarters (Q1, Q2, Q3, Q4).

$\boldsymbol{H_0}$ **:** $\text{Median}_{Q1} = \text{Median}_{Q2} = \text{Median}_{Q3} = \text{Median}_{Q4}$

Alternative Hypothesis (H_A): There is a significant difference in occupancy ("TotalOccupied") across at least one pair of seasonal quarters (Q1, Q2, Q3, Q4).

$\boldsymbol{H_A}$**:** At least one median difference between the four quarters The results are shown in Table 5-5, and are discussed below.

Table 5-5 Kruskal-Wallis H-test Results

Location	Kruskal-Wallis H	Degree of freedom	p-value
Loc-**1**	73.910	3	< 0.001
Loc-**2**	77.052	3	< 0.001
Loc-**3**	2256.396	3	< 0.001

As the Kruskal-Wallis H-statistics at Loc-1 is (73.910) with (3) degrees of freedom, the p value is less than (0.001). Therefore, since the p-value is less than the (0.05) common threshold level, the null hypothesis can be rejected. This tends to indicate that there are large differences between the medians of the range of occupancy between groups at Loc-1. The occupancy at this location thus cannot be said to have a similar trend across categories and may be seasonal or temporal.

At Loc-2, the Kruskal-Wallis H statistic is with a recorded value of (77.052), with (3) degrees of freedom and a p-value less than (0.001). Again, the p-value was below the accepted level of significance 0.05; hence the null hypothesis was rejected. The medians varied markedly across the different groups of Loc-2. The result agrees with that of Loc-1 in suggesting that the temporal or seasonal factors are the important causes of variation in occupancy.

On the other hand, Loc-3 considerably increased the Kruskal-Wallis H, and the statistic observed was (2256.396) with (3) degrees of freedom with a p-value less than (0.001). The very high H value, coupled with a p-value below (0.05), reflects very significant differences among the medians of the different groups with respect to Loc-3. This fact is indicative that Loc-3 has a much greater occupancy variation compared with Loc-1 and Loc-2. Results similar to this may reflect increased temporal variations at Loc-3 or even greater variation among the several groups.

The Kruskal-Wallis test results indicate that for all three locations, the median occupancy levels "TotalOccupied" differ significantly across the groups being compared. The p-values for all locations are less than (0.001), indicating that the occupancy levels are not evenly distributed over time, and there are distinct differences in median occupancy between different categories. Loc-3, in particular, shows a much higher H value, implying greater variability and more pronounced differences in occupancy levels compared to Loc-1 and Loc-2. These findings highlight the influence of time-based or seasonal factors on the availability of EVCSs at these locations, making it important to take these variations into account when modelling and planning for charging station usage.

As the Kruskal-Wallis H-test results indicate that there is a statistically significant difference in the distribution of "TotalOccupied" across the different categories of "quarter", pairwise comparisons were conducted to determine where the differences lie between the quarters.

5.1.2.2.3. Pairwise Comparison of Quarters

The pairwise comparisons aim to identify which quarters have significantly different distributions in terms of occupancy, with significance levels adjusted by the Bonferroni correction to account for multiple comparisons. For each location, an analysis of variance is executed by quarters of the year (Q1, Q2, Q3, Q4) depending on "TotalOccupied". Table 5-6 shows the results obtained from this analysis:

Table 5-6 Pairwise Comparison of Quarter Results

	Loc-1			**Loc-2**			**Loc-3**		
	Sig.	Adj.Sig	Sig.?	Sig	Adj.Sig	Sig.?	Sig.	Adj.Sig	Sig.?
Q1 vs. Q2	0.376	1.000	No	< 0.001	0.000	Yes	< 0.001	0.000	Yes
Q1 vs. Q3	**< 0.001**	**0.000**	**Yes**	**< 0.001**	**0.000**	**Yes**	**< 0.001**	**0.000**	**Yes**
Q1 vs. Q4	< 0.001	0.000	Yes	0.953	1.000	No	< 0.001	0.000	Yes
Q2 vs. Q3	< 0.001	0.000	Yes	0.104	0.622	No	< 0.001	0.000	Yes
Q2 vs. Q4	**< 0.001**	**0.000**	**Yes**	**< 0.001**	**0.000**	**Yes**	**< 0.001**	**0.000**	**Yes**
Q3 vs. Q4	0.753	1.000	No	< 0.001	0.000	Yes	0.531	1.000	No

The pairwise comparison of occupancy levels ("TotalOccupied") across quarters for the three locations highlights several significant differences, reflecting temporal variations in occupancy patterns. Notably, the comparisons between Q1 and Q3 as well as Q2 and Q4 are significant for all three locations, indicating consistent differences in occupancy levels during these quarters. However, there are also specific cases where no significant differences were observed.

In Loc-1, no significant difference was found between Q1 and Q2, and between Q3 and Q4, suggesting similar occupancy trends during these periods. In Loc-2, no significant differences were found between Q2 and Q3, or between Q1 and Q4. For Loc-3, significant differences were found for most pairwise comparisons, except for Q3 and Q4, which showed similar occupancy levels.

These results indicate that while certain quarters have consistent differences across all locations, suggesting potential seasonal effects, there are also quarters that exhibit similar occupancy patterns, particularly in Loc-1 and Loc-2. Understanding these temporal variations can help in effective resource allocation and strategic planning to meet occupancy demands throughout the year.

5.1.3. Feature Analysis Results

5.1.3.1. Feature Selection: Embedded Feature Selection Using Lasso

The Lasso embedded feature selection was implemented in Python by importing the *Lasso* library from *sklearn.linear_model*. The lasso function called before model training to check the features and provide the most important features to be used in the model training. Figure 5-7 shows the result of performing the Lasoo method for feature selection.

```
Selected Features: ['Day of Week', 'day', 'month', 'week_of_year', 'hour', 'sin_hour', 'cos_hour', 'AverageOccupied', 'Temp', 'Code', 'WindS']
Not Selected Features: ['Is Weekend', 'sin_day_of_week', 'cos_day_of_week', 'Prec']
Feature Scores:
Day of Week        0.039061
day                0.007453
month              0.149921
week_of_year       0.000048
Is Weekend        -0.000000
hour              -0.001873
sin_hour           0.006643
cos_hour           0.124713
sin_day_of_week   -0.000000
cos_day_of_week   -0.000000
AverageOccupied    1.100213
Temp               0.032038
Prec              -0.000000
Code               0.000489
WindS             -0.003434
```

Figure 5-7 Embedded Feature Selection Results

The features that obtained higher values (such as “AverageOccupied”, “cos_hour”, and “month”) were kept among the features, while on the other hand features with low value (such as “Is_Weekend”, “sin_day_of_week”) were not selected.

5.1.3.2. *Feature Importance* ***(Weather Data)***

This section explores feature analysis as applied for the implemented dataset, which focuses on two major items, as described below.

- **Feature selection:** this involves picking a few significant predictors that widely impact the effectiveness of the predictive models. Hence, selecting the most informative features will enable a reduction of dimensionality or removal of redundant and irrelevant information, as well as facilitate efficiency in the model towards generalizing.

- **Feature importance:** this assesses the contribution of each feature chosen towards the models’ prediction effectiveness. Understanding the importance of features enhances the interpretability of the models and offers critical insights regarding the factors that significantly affect EVCS access.

Collectively, these evaluations guarantee that models are as effective as they are efficient while using the most significant data that may precisely estimate charging station occupancy.

5.1.3.2.1. GCT

GCT was employed in this study to check the historical data of the variables (meteorological variables like temperature, precipitation, weather code, and wind speed) provide predictive power to the response variable "TotalOccupied" which represents the occupancy rate of EVCSs. Such identification of causal relationships among those variables should help in improving the feature selection methodology to predict the occupancy of charging stations. To evaluate the influence of weather features on charging station occupancy, GCTs were conducted between the target variable, "TotalOccupied", and each weather predictor ("Temp", "Prec", "Code", and "WindS"). The tests were performed using lags ranged from 1 to 4, where a lag refers to the number of previous time steps (hours, in this case) of the predictor variable used to test whether it helps forecast the current value of the target variable. The resulting p-values for each lag configuration are summarized in Table 5-7. The GCT assesses whether past values of one time series (the predictor) contain information that helps predict future values of another time series (the target). A low p-value (less than 0.05) indicates that the predictor Granger-causes the target variable at a given lag.

Table 5-7 P-Values from GCTs between Weather Variables and "TotalOccupied"

	Temp	Prec	Code	WindS
Lag1	0.215	0.641	0.840	0.693
Lag2	0.048	0.749	0.530	0.182
Lag3	0.001	0.094	0.376	0.341
Lag4	0.000	0.175	0.430	0.078

5.1.3.2.2. Interpretation of Results

The p-value for temperature **(0.215)** in lag1 is above the (0.05) threshold, suggesting no significant Granger causality at this lag. In lag2, the p-value drops to (**0.048**), slightly below (0.05), which indicates a potential Granger causal relationship. After this in Lag3 and Lag4, the p-values of (**0.001**) and (**0.000**) suggesting strong evidence of Granger causality, reinforcing the significant predictive power of temperature at higher lags. This means the temperature shows a significant Granger causal effect on "TotalOccupied" from lag2 onwards, implying that it is a valuable predictor for the model.

Unlike for the temperature, the p-values for precipitation, code, and wind speed remain above (0.05) across all lags, with the lowest being (**0.094**) for "Perc" at lag3, and (**0.078**) for "WindS" at lag4. This indicate that Precipitation, Code, and Wind Speed do not have a significant Granger causal effect on "TotalOccupied" and may not enhance the predictive model.

The GCT results therefore show that, out of all the weather variables, "Temp" feature has a statistically significant predictive relationship with charging station occupancy at lags greater than one. Hence, this depicts that temperature inclusions do enhance the predictability of the model without making it unnecessarily complex and running the risk of overfitting. In contrast, precipitation, weather code, and wind speed are found not to have significant Granger causality with "TotalOccupied" at any lag. These features might bring a little more predictive power but including them would overcomplicate the model with no added value for model performance.

5.1.3.3. Feature Importance Analysis: LightGBM and SHAP

The feature importance analysis was carried out to understand the various factors affecting the variance in the occupancy forecast at EVCSs, using the LightGBM methodology coupled with SHAP. These have provided the most telling insight into the relevance of several features for improving the predictive power of the model. By identifying key predictors and understanding their impact, the analysis helps to better comprehend the factors affecting charging station occupancy, which is crucial for optimizing the predictive model and informing decision-making processes.

5.1.3.3.1. LightGBM Model Results and Feature Importance Interpretation

The LightGBM model was employed to predict the occupancy levels of EVCSs. The model's performance was evaluated using the MSE and MAE, which were found to be (**0.8776)** and (**0.5844)**, respectively. These metrics indicate a satisfactory level of accuracy in the model's predictions, with the MAE suggesting that, on average, the model's occupancy predictions are within approximately (0.58) units of the actual values. The feature importance scores derived from the LightGBM model provide insights into which variables significantly influence occupancy predictions. The features and their corresponding importance scores are presented in Table 5-8 and Figure 5-8, and the implications of these results are discussed below.

Table 5-8 Features and Corresponding Importance Scores Using LightGBM Model

Feature	Importance Score
day	608
AverageOccupied	470
month	422
Temp_lag4	219
Day of Week	206
Temp	182
AverageOccupied1	170
cos_hour	143
Temp_lag2	94
Temp_lag1	93
sin_ day_of_week	89
Temp_lag3	76
hour	65
cos_day_of_week	50
LocC2	49
LocC1	29
sin_hour	28
LocC3	7
Is Weekend	0
week_of_year	0

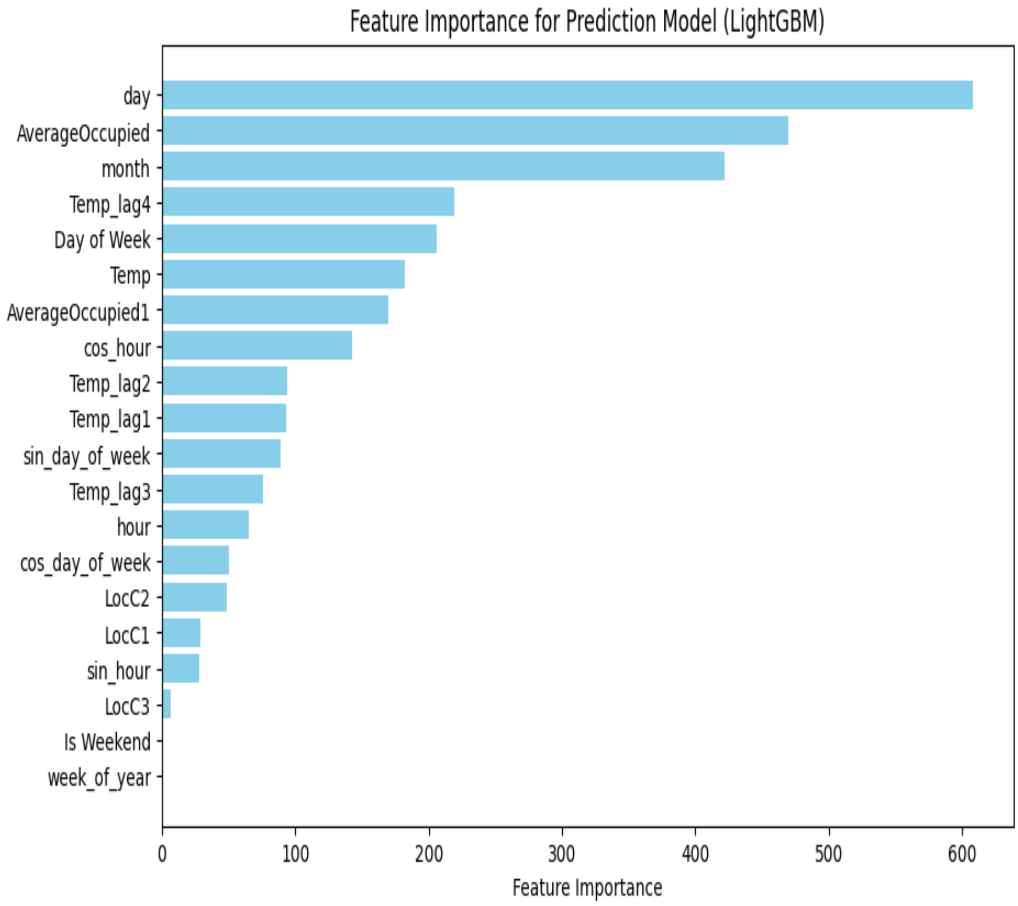


Figure 5-8 Feature Importance for Prediction Model (LightGBM)

Temporal Features: The most relevant features for charging station occupancy are temporal variables. The feature “day” shows a maximal importance score value of (608), hence the occupancy is critical for the exact days of the month. The “month” and “Day of Week” features rank second and third (respectively), with importance scores of (422) and (206), to reflect monthly and weekly occupancy patterns. Other features representing these are “hour”, “cos_hour”, and “sin_hour” contributing to occupancy, reflecting the time of day with daily cycles of use.

Historic Occupancy: Historical occupancy data is crucial for predicting future usage The features “AverageOccupied” and “AverageOccupied1” have importance scores of (470) and (170), respectively, underscoring the relevance of past occupancy levels in forecasting future demand.

Geological Features: Temperature-related features drive the model’s predictions on occupancy significantly. Features like “Temp, “Temp_lag1”, “Temp_lag2”, "Temp_lag3", and "Temp_lag4" all have a generally high importance score combined. An example is the importance score of (219) for the "Temp_lag4". This demonstrates that temperature with lagged values has a higher effect than on occupancy. Temperature variables with lag therefore indicate that past temperature conditions have a longer-lasting impact on present charging behaviour; which confirms the result of GCT.

Geographical Identifiers: Different locations of charging stations bring about varied occupancy rates. The "LocC2" variable importance score is (49), while for "LocC1" and "LocC3", the importance scores are (29) and (7), respectively. This confirms that some particular locations bear a lot of importance in predicting the occupancy level.

Less Important Features: Importance scores of (0) were attained by the attributions "Is Weekend" and "week_of_year", indicating that they do not contribute towards the model's predictability at this setting. Low explanatory power is indicated by the fact that the weekend pattern and the exact week in the yearly cycle do not feature.

5.1.3.3.2. SHAP Analysis

SHAP values were used to quantify and visualize feature importance and to understand the impact of each feature on the model's prediction. The plot in Figure 5-9 illustrates the global feature importance for all samples by plotting the SHAP values. The x-axis indicates the impact of the feature on the model's output, while the colour represents the magnitude of each feature's value (blue for low, red for high).

(a)

(b)

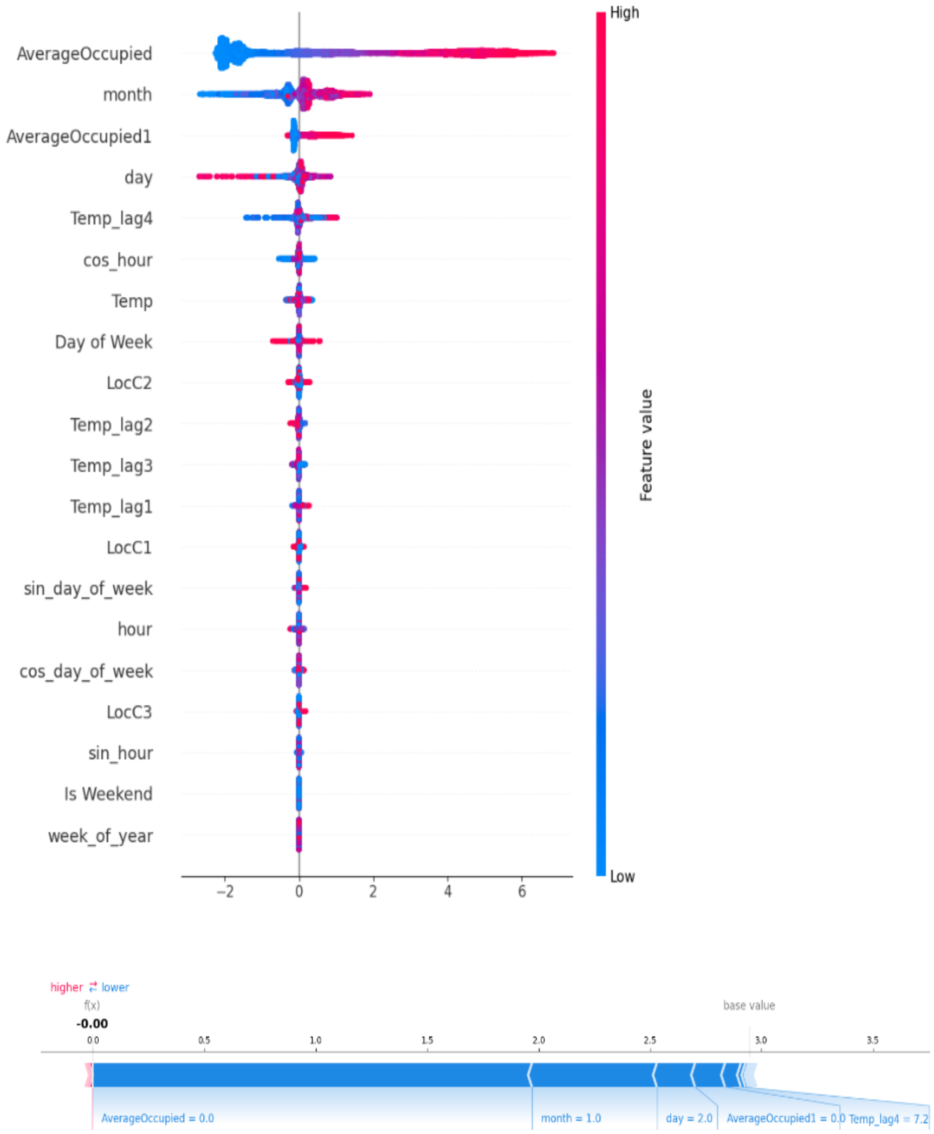


Figure 5-9 (a) Feature Importance Summary, (b) SHAP Force Analysis

AverageOccupied emerged as the most important feature, showing a consistent positive impact on model prediction. High values of AverageOccupied (shown in red) have a strong positive effect on increasing the prediction. **Month** and **Day** also significantly influence the predictions, as indicated by their prominence in the SHAP plot. The month feature, represented by a mix of red and blue, shows that seasonal variations significantly impact predictions, which could be associated with seasonal charging behaviours. T**emperature Lags (Temp_lag1 to Temp_lag4)** also contributed to the predictions, although their impacts were more dispersed and generally lower compared to AverageOccupied.

The presence of both red and blue suggests that the temperature's influence on occupancy changes over different conditions. **Temporal Features (Day of Week, hour, and cos_hour)** also have noticeable effects. For example, Day of Week indicates that certain days are associated with increased or decreased occupancy, potentially highlighting differences in charging behaviour on weekdays versus weekends.

One interesting observation is the difference in distribution for **Day** compared to **AverageOccupied**. **Day** shows an even mix of red and blue SHAP values around zero, indicating that the impact of the day feature on the model output is context-dependent and changes throughout the dataset. In contrast, **AverageOccupied** shows a much stronger relationship between the feature value and the prediction. High AverageOccupied values generally push the model's prediction upwards, showing a consistent influence regardless of the context.

5.1.3.3.3. SHAP Force Plot Analysis

The SHAP force plot, which shows the prediction for a single observation, provides a detailed visualization of how each feature contributes to the final prediction value. In this specific prediction example, **AverageOccupied** had a strong impact, pushing the prediction down. The fact that no red colour is present on the left side indicates that features like AverageOccupied and Temp_lag4 had a reducing effect on the final prediction. **Month** and **Day** also exerted an influence on the prediction, but their effect was relatively more balanced. The blue colour indicates that the features contributed to lowering the predicted value. In this case, the combination of the month, day, and Temp_lag4 resulted in a decrease in the final prediction of occupancy.

Overall, these analyses suggest that **AverageOccupied** is a crucial driver of occupancy predictions, consistently increasing predictions when the average occupancy is high, while other features like **Day**, **Month**, and **Temperature Lags** have more varied impacts depending on the specific scenario. The results of the SHAP analyses align with expectations, highlighting how both temporal and occupancy history features contribute to understanding EVCS occupancy.

5.2. Discussion

As illustrated in **Error! Reference source not found.** and Figure 5-1, variations in the occupancy pattern suggest significant differences in the charging event and its trend across all stations, even within a single area. These differences could be attributed to the factors adumbrated below.

5.2.1. Effect of Spatial Features on the Occupancy State

The occupancy state of EVCSs is influenced by several spatial features, as demonstrated by the variation patterns across the three locations (Temple Green Park and Ride, Elland Road Park and Ride, and Woodhouse Lane Car Park). The spatial characteristics of each location could play a significant role in shaping user behaviour and consequently the occupancy levels at these stations.

5.2.1.1. Geographical Determinants

The geographical characteristics of a station's location are crucial in determining its occupancy trends. For instance, Temple Green Park and Ride, located in an industrial area, shows a stable occupancy pattern with minimal variability, suggesting a steady user base that likely comprises local commuters who prefer this station for its proximity to the M1 and other industrial facilities. This was clearly reflected in the very low mean and SD in the descriptive statistics tables (Table 5-1, Table 5-2, and Table 5-3). The consistent usage at this site indicates that its location is highly convenient for individuals commuting to and from nearby workplaces, which supports a stable demand.

In contrast, Elland Road Park and Ride, situated near both residential and industrial areas, shows lower peaks in occupancy. This location's moderate usage could be influenced by competing alternatives in the vicinity, such as other charging facilities or more convenient options for residents. Its distance from major highways and other strategic locations may also make it less appealing, leading to underutilization during certain periods.

Woodhouse Lane Car Park, located near educational institutions, commercial centres, and a hospital, exhibits a high level of variability in occupancy, with distinct peaks observed towards the end of the year. Also, this is clearly notable from the descriptive statistics table with highest mean and SD. This suggests that the location is subject to intermittent demand, potentially driven by events such as academic schedules, hospital visits, or shopping. The sharp variations in occupancy indicate that the station caters to a fluctuating user base, likely influenced by events and seasonal activities.

5.2.1.2. Station Capacity and Amenities

The capacity and available amenities at each station also play a role in influencing occupancy. Stations with more charging ports, convenient parking, and nearby amenities such as cafes or shopping centres are likely to attract more users. For example, the variability observed at Woodhouse Lane Car Park might be partly due to the presence of nearby amenities that attract transient users, particularly during peak shopping seasons or holidays. On the other hand, Temple Green Park and Ride, which shows consistent occupancy, may provide reliable services that attract regular commuters who value the convenience of its industrial location over additional amenities.

5.2.1.3. Price-Driven Determinants

Another spatial feature that may affect occupancy is the pricing strategy employed at different stations. For example, price reductions during off-peak hours can influence charging behaviours, with users opting to charge their vehicles during times when rates are lower. Such price-driven factors can result in noticeable patterns in day-to-day occupancy levels, particularly in urban areas where price-sensitive..

In this study, while direct pricing data was not available, occupancy patterns observed across the different stations; particularly in Figure 5-1 suggest that Temple Green maintains a relatively stable occupancy level, whereas Woodhouse Lane exhibits greater variability, including sharper peaks and dips. These differences may reflect underlying pricing strategies or incentive schemes, such as differential pricing during peak hours, that influence user decisions. Although this interpretation is inferred from usage trends rather than confirmed by direct pricing data, the pattern aligns with findings from previous studies (e.g.,(Sun et al. 2020)) that highlight the impact of dynamic pricing on EV charging behaviours.

5.2.1.4. Seasonal and Event-Driven Influences

Seasonal events and local activities are also key determinants of charging station occupancy. Locations like Woodhouse Lane, which experience high variability, might be influenced by periodic events such as academic semesters or local gatherings that attract a large number of temporary visitors. The spikes in occupancy during specific months suggest that the station caters to users whose demand is event-driven, such as students during the beginning of a new term or shoppers during the holiday season. This contrasts with the more stable patterns observed at Temple Green, indicating that different spatial features can lead to markedly different occupancy behaviours. The mean values indicate that Stat-3 has significantly higher occupancy and occupancy category values compared to Stat-1 and Stat-2. This suggests that Stat-3 generally experiences a higher frequency of usage, with more vehicles utilizing the charging spaces on average.

The standard error quantifies the level of accuracy in defining the mean estimate. Standard errors of less magnitude suggest more accurate estimations, while higher standard error for Stat-3, particularly for “TotalOccupied”, indicates a higher level of unpredictability in occupancy, as anticipated based on the bigger SD and range. While regarding the value of SD, Stat-3 has the highest SD for both “TotalOccupied” and “OccCate”, indicating more variability in occupancy compared to Stat-1 and Stat-2. This suggests that Stat-3 experiences a wider range of occupancy levels throughout the observation period. As with SD, the variance for Stat-3 is much higher than that of Stat-1 and Stat-2, further highlighting the greater variability in occupancy at Stat-3.

As the variability and variance of the data can significantly affect the DLM’s training process and its accuracy, understanding how variability and variance influence the model can help in deciding the best pre-processing steps and model architecture. Although higher variability and variance can lead to difficulty learning patterns, increase the risk of overfitting and longer training time, this does not necessarily indicate inadequate model performance. If the significant variability arises from significant variations (e.g., seasonal trends, diverse user behaviours), the model can still acquire knowledge of these patterns. However, to effectively manage heightened complexity, more advanced methods may be needed, such as regularization, data augmentation, or a more complicated model architecture.

5.2.2. Temporal Effect on Occupancy Patterns

5.2.2.1. Days of the Week

The occupancy patterns across the different days of the week reveal notable trends. For all stations, Mondays and Fridays tend to show lower occupancy rates compared to the midweek days (Tuesdays, Wednesdays, and Thursdays). This pattern may be explained by the behaviour of EV owners, who may use home chargers during the weekends, thus reducing their need for public charging on Mondays. Similarly, on Fridays, users might prefer to return home and use their home chargers over the weekend. This observation is supported by the low occupancy rates at public charging stations during weekends, particularly at Stat-1 and Stat-2, suggesting a preference for home charging. Stat-1 and Stat-2 appear to primarily serve commuters, as indicated by the sharp rise in occupancy during the morning and a corresponding decline in the evening, which aligns with typical workday travel patterns.

These stations see minimal activity during weekends, reflecting the behaviour of customers who reduce their travel activities outside of workdays. Stat-3, in contrast, serves a more diverse set of users. The occupancy rates during both weekdays and weekends are higher and more variable, suggesting that Stat-3 is located in an area characterized by mixed activities, such as a city centre, shopping district, or entertainment venue. This station's accessibility and appeal to a broader user base, beyond just commuters, is likely responsible for its distinct occupancy patterns. The results for both occupancy level in weekdays and weekends in this research aligns with the result obtained from other works introduced by Soldan et al. (2021) and Lucas et al. (2019).

5.2.2.2. Quarters of the Year

Overall, the occupancy trends across the quarters indicate that each station experiences different levels of seasonal influence. For Loc-1, the test concluded that there is no significant difference in the median occupancy levels between Q1 and Q2, as well as between Q3 and Q4. This could indicate that the occupancy follows different patterns for each half of the year, where the station starts the year with lower occupancy and gets slightly busier in the second half. This difference might be attributed to pricing reasons or changes in site amenities, such as more industrial businesses opening nearby. More likely, this difference is not cyclical and may not be repeated consistently. In Loc-2, the test concluded that there is no significant difference in occupancy between Q2 and Q3, and similarly, between Q4 and Q1. This suggests a clear seasonal pattern where the station tends to get busier during the colder months (Q4 and Q1), possibly due to increased demand in colder weather, while it becomes slightly quieter in the better weather conditions of Q2 and Q3.

For Loc-3, the results indicate that occupancy shows more stability towards the end of the year, with no significant difference observed between Q3 and Q4. One possible reason for this stability could be that occupancy demand reaches equilibrium, possibly influenced by consistent commuter behaviour or a steady local environment, leading to fewer fluctuations in occupancy levels towards the end of the year. There is another notable point from the statistical test common to all locations: *Q1 is always significantly different from Q3, and Q2 is always significantly different from Q4*. This consistent finding suggests that there may be distinct differences in occupancy between the first and third quarters, as well as between the second and fourth quarters, which could be influenced by various factors.

One possible explanation or this finding is that these differences may be related to variations in seasonal activities, local events, or changing user behaviour patterns across different times of the year. For instance, Q1 (at the beginning of the year) might have lower activity levels, which could pick up by Q3 due to mid-year events, increased travel, or changing work schedules. Similarly, Q2 might be less busy compared to Q4, possibly due to weather conditions or year-end activities that increase demand towards the last quarter. These differences highlight the importance of understanding temporal patterns to effectively allocate resources and meet user demand at different times of the year.

The statistically significant results from the Kruskal-Wallis H-test demonstrate that at least one median difference exists among the quarters for all locations. However, the pairwise Mann-Whitney U tests further clarified that some quarters do not exhibit significant differences, providing valuable insights into occupancy dynamics. These findings can help inform targeted resource allocation and planning strategies for each station, considering the seasonal and location-specific variations.

5.2.2.3. Hour of the Day

The analysis of occupancy patterns by hour of the day reveals that peak hours for Stat-1 and Stat-2 occur between 6 AM and 6 PM, which aligns with typical working hours. This suggests that these stations serve areas with a high concentration of workplaces or industrial sites, catering to commuters during their workday. This trnds supports the finding of other works provided by Soldan et al. (2021) and Lucas et al. (2019). On the other hand, Stat-3 shows an opposite trend, with lower occupancy rates during these hours, possibly due to its location in a more residential or commercial area where charging demand is less tied to workday commuting. Instead, Stat-3 may serve users who charge their vehicles during off-peak hours, driven by different activities such as shopping, dining, or leisure. Spatial features, including geographical location, station capacity, nearby amenities, pricing, and seasonal influences, play a critical role in determining the occupancy state of EVCSs. Understanding these factors is essential for optimizing the placement and management of EVCI to meet the varying needs of users effectively.

5.2.3. Feature Importance (Weather Data)

Integrating weather data into predicting EVCS occupancy has demonstrated substantial but selective value. Specifically, while temperature consistently showed a Granger causal relationship with occupancy—especially at lags greater than one—other weather variables, such as precipitation and wind speed, did not. This selective relevance aligns with findings from other predictive studies, such as Feng et al. (2022), which leveraged environmental data for forecasting parking space availability. Like this study, they found that specific variables yielded predictive power, while others added complexity without benefits. This selective inclusion of weather variables thus reduces model overfitting risks, and enhances generalization, as also suggested by Akshay et al. (2024) regarding the forecasting of EV demand and power consumption.

The dual use of LightGBM and SHAP provided interpretability into the feature impact at both global and local levels, confirming “AverageOccupied” and “day” as key predictive factors, similar to Qiao and Lin (2021), who highlighted granular time-based charging behaviour patterns for forecasting occupancy. This alignment with Qiao and Lin (2021) findings reflects the growing consensus that temporal features—like day, month, and hour—are essential for accurate occupancy predictions in EVI. The SHAP analysis further supported this finding by detailing the individual contribution of these temporal features, echoing Sao et al. (2021), who noted the predictive significance of combining temporal data with additional context for more reliable occupancy predictions.

Furthermore, the selective use of weather data in this study contrasts with models that indiscriminately incorporate multiple external factors, as seen in Douaidi et al. (2023) who employed a federated learning approach to protect privacy but encountered issues with heterogeneity and communication overhead. By focusing on temperature alone as a significant predictor, this model’s architecture remains robust yet interpretable, balancing computational efficiency and data relevance. This approach addresses a gap in previous research, such as Ostermann et al.’s (2022) attempt to improve occupancy prediction for public EV charging, which encountered scalability and real-time processing challenges due to more complex models that included broader environmental features.

The findings from the analysis of weather and temporal features directly address **RQ2-1** (“Which specific temporal and environmental features improve the quality and relevance of the training data for predicting EVCS occupancy?”). The study identified that among the environmental features analysed, temperature emerged as a significant predictor, with GCTs demonstrating a statistically relevant relationship between temperature and occupancy levels, particularly at higher lags. This indicates that temperature has a meaningful influence on EV charging behaviour, suggesting it can reliably enhance model predictions. However, other weather features, such as precipitation, wind speed, and weather codes, did not show substantial impacts.

Including only temperature in the model, therefore, not only improves predictive relevance but also prevents unnecessary model complexity and reduces the risk of overfitting. This finding is consistent with similar work by Qiao and Lin (2021), who noted that selective environmental data integration could refine model efficiency. For temporal features, the study highlights day of the week, month, and hour as essential contributors to occupancy prediction accuracy. LightGBM and SHAP analyses confirmed that these temporal indicators are strongly correlated with occupancy patterns, reinforcing their value in capturing seasonality and daily usage trends. The inclusion of these temporal features helps the model accommodate cyclical demand patterns and peak usage periods, supporting findings from Akshay et al. (2024) and Sao et al. (2021), which emphasized the predictive strength of temporal variables in EV-related models.

Overall, this study supports the integration of temperature, day of the week, month, and hour as core features to improve training data quality and enhance the relevance of EVCS occupancy predictions. These features capture both the cyclical nature of EV usage and the specific environmental conditions influencing user behaviour, directly answering the sub-research question by providing a refined approach to feature selection for enhanced prediction outcomes. After understanding the data and analyse the charging patterns from the data, in the following chapter DLMs will be trained to predict the occupancy state of EVCSs fir the different locations.

Chapter 6: Models Development and Comparison

This chapter investigates the use of DLMs to forecast the occupancy of EVCSs. It mainly aims to address research questions 1, 3, and sub-question 2.2. The experiments design contains a models' implementation, training and evaluating where this chapter ended by selecting an appropriate model from several proposed. Then was followed in the later level by refining the chosen model to better align with the characteristics of the dataset and the nature of the task.

6.1. Initial Modelling Development and Evaluation Results

After a preliminary study, a review of previous research, and understanding the data and nature of this task, five DL architectures were selected to perform the predictive models: LSTM, GRU, 1D-CNN, TCN, and an Ensemble model (comprising 1D-CNN and TCN). As it stated in Chapter 4, this stage was split into two phases, to consider both classification and regression processes. In each experiment, models were trained on the same historical dataset with initial hyperparameters and basic architectures. The most suitable model for each experiment was identified by comparing the evaluation metrics resulting from each model's implementation; the selected model was then used for the subsequent stages of this research. This stage is dedicated to conducting the experimental study that addresses the following research questions:

RQ1: How accurately can DLMs predict the availability of EVCSs based on historical data?

RQ2-2: How does the use of aggregated data from multiple locations versus location-specific data impact the generalization and accuracy of DLMs in predicting EVCS occupancy?

RQ3: How does indicating the prediction of the EVCS occupancy as a classification versus regression task affect the performance and accuracy of DLMs?

6.1.1. Phase 1: Classification Models for the Occupancy State Prediction Results

The objective of this phase was to identify the optimal DL architecture to predict the occupancy category of a charging station. In other words, the important concern in this phase is to evaluate the efficiency of such models in predicting the occupancy state of the charging stations. This classification methodology enables insights into how various models perform in handling categorical outcomes, and how these are useful in decision-making systems that require the generation of categorical predictions. The primary goal is therefore to determine how well the proposed models can predict the availability of charging stations. This multi-classification model comparison allows to assess the performance of various models and compare its ability in such tasks.

6.1.1.1. Input-Data Preparation

The architecture of the models used in this phase was configured to leverage time-series data and capture patterns in occupancy trends based on the location, time of day, and other relevant features. These models were chosen due to their known strengths in processing sequential data and their ability to capture both short- and long-term dependencies in the dataset. The dataset, as described in Chapter 4, consists of various features, including time-related data, location indicators, occupancy categories, and weather data. To prepare the input data for the classification approach in this part, several preparatory steps were applied to the main data.

First, the target variable "OccCate" was encoded to different labels using "*LabelEncoder*", then the encoded labels were converted into categorical format using "*to_categorical*", which is suitable for multi-class classification problems.

The dataset was then split into three subsets: 60% for training, 20% for validation, and 20% for testing. These split ratios were selected based on empirical evaluation: multiple data splitting schemes were trialled (including 70/15/15 and 80/10/10), and their effect on model performance was assessed using key evaluation metrics such as validation loss, classification accuracy, and signs of overfitting (e.g., performance drop on the validation set). The 60/20/20 configuration was ultimately chosen as it provided the best trade-off between sufficient training data and robust validation/testing evaluation, while also minimizing overfitting and providing stable convergence during model training.

Features were normalized using "*StandardScaler*", ensuring that each input feature tended toward centrality and variance of the data, which facilitated faster convergence during training and improved model performance.

6.1.1.2. Description of Models

The five selected models used in this experiment are briefly described below, and are summarized in Table 6-1.

1D-CNN: This model comprises successive one-dimensional convolutional layers designed to extract spatial features from the input data, followed by max-pooling layers that reduce the dimensionality by summarizing feature information and mitigating overfitting. A fully connected layer was then added before the output layer, which predicted the category of occupancy using a *softmax* activation function.

TCN: This model uses use causal convolutions to handle temporal dependencies within the dataset. The architecture consists of combined layers of convolutional filters using dilation techniques to capture long-range dependencies effectively, making it more suitable for sequence modelling tasks.

Ensemble Model: This hybrid model combined the strengths of both 1D-CNN and TCN model architectures. Outputs from 1D-CNN and TCN were concatenated in this model to allow it to cover both spatial and temporal features at the same time. A fully connected layer, followed by *softmax* output, then predicted the class labels.

LSTM: The LSTM architecture exploited its essential memory for effectively capturing temporal relationships, which are of crucial importance when dealing with sequential data. In this regard, this framework was set up with one LSTM layer followed by a dense layer to convert the hidden representations into occupancy classification.

GRU: Like LSTM, other networks use the GRU architecture to capture long-term dependencies in sequences using its gating mechanism. Generally, GRUs tend to be more computationally efficient than LSTMs, thus becoming a practical option to handle large sequential data.

The batch size of 32 and the Adam optimizer were chosen based on their general effectiveness in deep learning applications, providing a balance between convergence speed and stability. The categorical cross-entropy loss function was used for classification tasks due to its suitability for multi-class problems. Early stopping was implemented to prevent overfitting, ensuring that training halts when validation performance no longer improves. The architectures and initial hyperparameters were determined based on exploratory experiments conducted during preliminary stages. These experiments provided insights into how the dataset characteristics such as the time-series nature and class distribution interacted with different model types and parameter settings. A dense layer with 64 units was included in each architecture to process the features extracted by preceding layers. This size ensures sufficient capacity for feature representation without risking overfitting, particularly given the dataset size and complexity.

Table 6-1 Summary of Models' Architectures

Model	Architecture Summary	Common Settings
1D-CNN	Input layer 1D Convolutional Layer (32 filters, kernel size = 3, ReLU activation) MaxPooling Layer (pool size = 2) Flatten Layer Dense Layer (64 units, ReLU activation) Output Layer (softmax activation)	**Epochs**: 80 **Batch Size**: 32 **Optimizer**: Adam **Loss Function**: Categorical Cross-Entropy **Callbacks**: Model checkpoints and early stopping to save the best model during training.
TCN	Input layer Temporal Convolutional Layer (32 filters, kernel size = 3) Dense Layer (64 units, ReLU activation) Output Layer (softmax activation)	
Ensemble (1D-CNN + TCN)	Input layer 1D-CNN (32 filters, kernel size = 3, MaxPooling, Flatten) TCN (32 filters, kernel size = 3) Concatenate the outputs of 1D-CNN and TCN Dense Layer (softmax activation)	
LSTM	Input layer LSTM Layer (32 units) Dense Layer (64 units, ReLU activation) Output Layer (softmax activation)	
GRU	Input layer GRU Layer (32 units) Dense Layer (64 units, ReLU activation) Output Layer (softmax activation)	

This experiment was divided into two scenarios (Combined input-data / Location-specific input-data) to address the sub-question (**RQ2-2**):

First, in Scen-1, the models were trained on a combined dataset includes records from three different locations. This approach allowed the models to learn from a comprehensive set of data and generalize across different locations. Second, in Scen-2, the dataset was split by location into three subsets, and the models were trained independently on each subset. This allowed for a more location-specific evaluation, helping to determine whether models perform better when focused on data from a single location versus the full dataset. The evaluation metrics (accuracy, precision, recall, F1-score) and confusion matrix with area under the curve (AUC) analysis were monitored and reported then compared to evaluate the model's performance.

6.1.1.3. Scenario 1: Models Trained on Combined Dataset (Scen-1)

The models were developed via the *Adam* optimizer, which adaptively modifies the learning rate throughout training to expedite the adaption process. Given that the challenge involves multi-class classification, the models optimized the categorical *cross-entropy* loss function. The training was configured for (80) epochs, with an early termination criterion activated after (20) epochs of no enhancement. Early stopping was utilized to avert overfitting, ensuring the model ceased training as its performance on the validation data stabilized, hence conserving time and decreasing the likelihood of overfitting. Validation accuracy and loss were observed to prevent overfitting. A *ModelCheckpoint* callback was employed for each model, preserving the optimal version according to validation accuracy. This guaranteed that the final model employed for testing reflected the optimal performance from the training scenario. The training method allowed real-time monitoring of training and validation accuracy, permitting continuous evaluation of model performance over epochs.

The left side of Figure 6-1 represents the training accuracy for the five models among the (40) cycle of the training epochs, while the right side of Figure 6-1 shows the calculated loss for the models within the (40) epochs. The TCN model achieved the highest training accuracy with approximately (95%) at the end of the training. This indicates that the TCN was the most successful model at learning from the combined dataset in this stage. Following the TCN, the Ensemble model obtained the second-best performance in this stage. The progress of the Ensemble model shows a steady growth in the accuracy where it eventually stabilized just below (90%). Not too far behind the Ensemble model, 1D-CNN reached a training accuracy around (85%). Both of the studied RNN varieties (LSTM and GRU) achieved the lowest training accuracy scores, with final values just over (80%). These outcomes indicate that these models appear to struggle more with learning patterns in the combined dataset compared to the others.

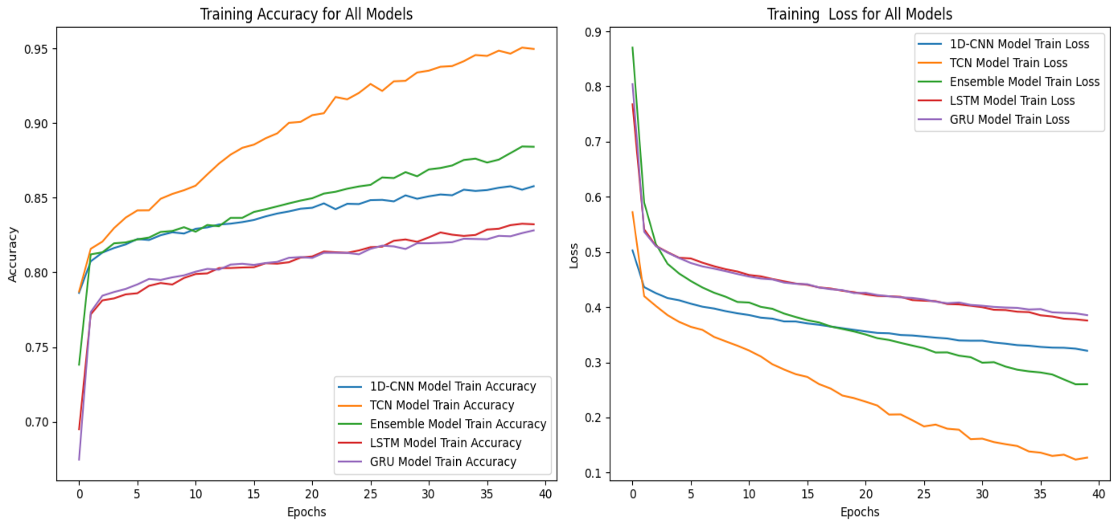


Figure 6-1 Accuracy and Loss per Epoch for All Models

On the right side of Figure 6-1, the training loss graph demonstrates the decline in loss for each model across epochs, providing insights into how effectively each model minimizes error. The TCN model not only achieved the lowest final loss, but also showed a smooth and consistent reduction in loss over epochs. This smooth and consistent reduction in loss indicates effective learning and convergence. The Ensemble model also demonstrated a steady decrease in loss, though not as sharply as the TCN model, showing that it was able to minimize errors fairly well. The 1D-CNN model initially had a rapid decrease in loss before stabilizing around (0.3). Both the LSTM and GRU models showed higher final losses, converging more slowly and with less pronounced improvement over epochs.

It can be seen that the TCN model stands out as the best performer, showing both high accuracy and low loss by the end of training; thus, it is a strong candidate in attempting to predict occupancy with a combined dataset. The Ensemble model also performed well, showing that archiving TCN with other architectures leads to robust results. In contrast, the LSTM and GRU models resulted in lower accuracy and higher loss, which might mean that these models are less effective for such a configuration of a dataset. This result underscores how important the choice of a model is, and it shows the potential of TCN for occupancy data with temporal patterns.

6.1.1.3.1. Models' Performance on Unseen Data

In evaluating machine learning models, it is necessary to validate their performance against unseen data to be sure of its generalization capabilities. The fact that a model has a high degree of accuracy on its training data does not really show whether the model is capable of performing well on new and unseen data. In addition, models must be verified on test data not used during their training to ensure that they only memorize the patterns seen during training but can predict accurately in a new environment.

Most of the models, as can be seen in Figure 6-2, show fairly acceptable accuracy when tested on previously unseen experimental samples. This further confirms the ability of the selected models to learn from historical data and provide reasonably acceptable future predictions. Once again, the TCN model outperforms the other models by having the highest test accuracy and lowest loss value. The rest of the models are not far behind, with slightly lower accuracy, however, significantly higher loss values, especially for the Ensemble model and, to a less significant extent, the LSTM and GRU models.

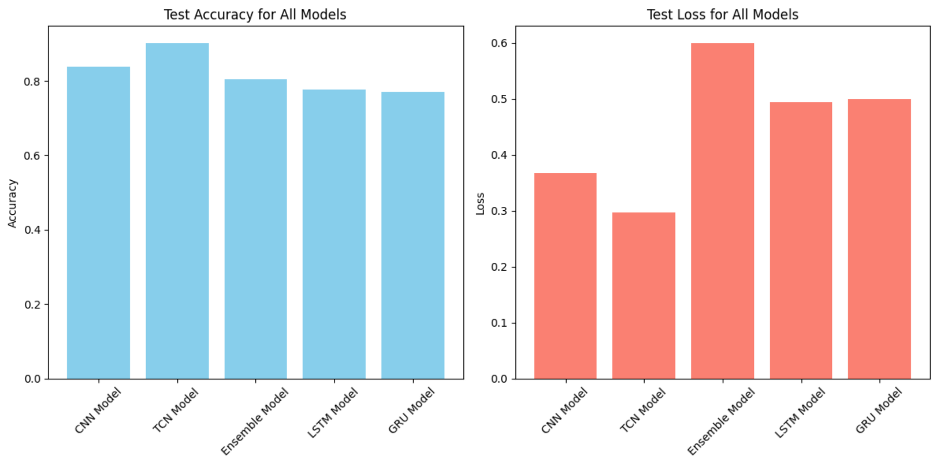


Figure 6-2 Models' Performance on Test Data

6.1.1.3.2. Scen-1 Classification Report

This classification report demonstrates the recorded classification scores and metrics. The report assists to evaluate the models' performance as well as the model selection for this scenario.

6.1.1.3.2.1 Evaluation Matrices Table

The use of classification evaluation metrics such as *precision*, *recall*, *F1-score,* and *support* are crucial in understanding and comparing the performance of machine learning models, especially in imbalanced datasets like occupancy prediction. Specifically, for example, the distribution of the different categories of the occupancy state in station 1 was about (58% for class 0, 20% for class 1, 13% for class 2, and 9% for class 3).

Precision indicates the proportion of true positive predictions among all predicted positives, giving insight into a model's accuracy in classifying a particular class without overestimating it. *Recall* (or *sensitivity*) measures the ability of the model to identify all true positives, which is important when missing instances of a particular class has a significant cost. The *F1-score* combines both precision and recall into a single metric, providing a harmonic mean that balances the trade-off between false positives and false negatives. Finally, *support* reflects the number of instances in each class, helping to contextualize the metrics, as high performance in a majority class (with high support) can mask poor performance in minority classes (with low support) (Sokolova and Lapalme, 2009). Together, these metrics provide a comprehensive view of model performance, particularly in multi-class problems, where different classes may present distinct challenges for prediction.

Table 6-2 shows the evaluation metrics comparison which highlights the performance of five different models in this scenario across four classes (0 to 3); where the four classes are (empty, almost empty, moderate full, full). The models are evaluated using precision, recall, F1-score, and support, providing insight into how well each model classifies different occupancy states.

Table 6-2 Evaluation Metrics Comparison

Model	**precision**				**recall**				**F1**				**support**			
	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
1D-CNN	0.947	0.707	0.618	0.621	0.956	0.693	0.686	0.381	0.951	0.700	0.650	0.472	3283	1094	659	202
TCN	0.967	0.816	0.790	0.773	0.977	0.825	0.771	0.658	0.972	0.821	0.780	0.711	3283	1094	659	202
Ensemble	0.880	0.682	0.531	0.000	0.975	0.401	0.772	0.000	0.925	0.505	0.630	0.000	3283	1094	659	202
GRU	0.906	0.603	0.487	0.000	0.955	0.431	0.734	0.000	0.930	0.502	0.586	0.000	3283	1094	659	202
LSTM	0.913	0.510	0.462	0.000	0.954	0.576	0.401	0.000	0.933	0.541	0.429	0.000	3283	1094	659	202

The 1D-CNN model performs exceptionally well in class 0, with a precision of (0.947), recall of (0.956), and an F1 score of (0.951), indicating that it can accurately classify the empty or near-empty state (class 0). However, for class 1 and class 2, the performance declines, with the F1 score dropping to (0.700) and (0.650), respectively, reflecting difficulty in handling intermediate occupancy levels. The model significantly underperforms in class 3 (full occupancy), with an F1 score of (0.472), due to low recall (0.381), meaning the model often misses instances of this class.

The TCN model achieves the best overall performance across all classes. In class 0, it has near-perfect metrics, with a precision of (0.967), recall of (0.977), and F1 score of (0.972). It also performs exceptionally well in class 1 and class 2, with F1 scores of (0.821) and (0.780), indicating it is highly effective at predicting intermediate occupancy states. Notably, the TCN model achieves the highest performance in class 3 (F1 = 0.711), handling the most complex cases much better than the other models.

The Ensemble model shows mixed performance. It performs well for class 0 with an F1 score of (0.925), and reasonable results for class 2, where F1 reached (0.630), but it struggles significantly in class 1 and completely fails to predict class 3, as indicated by an F1 score of (0.000) for class 3. This suggests that while combining 1D-CNN and TCN enhances performance in some classes, the ensemble lacks the ability to generalize well to more complex and underrepresented classes like class 3.

GRU model performs well in class 0 with an F1 score of (0.930), but it starts to show weaknesses in other classes. In class 1, the F1 score drops to (0.502), reflecting difficulty in handling moderate occupancy levels. Similarly, the F1 score for class 2 is (0.586), and the model fails to make any correct predictions for class 3, resulting in an F1 score of (0.000). This poor performance in the more complex classes suggests that the GRU model struggles underrepresented data.

The LSTM model shows a similar trend to the GRU model, performing well in class 0 with an F1 score of (0.933) but experiencing significant performance drops in other classes. For class 1, the F1 score is (0.541), and for class 2, it is (0.429), indicating difficulty in classifying more complex occupancy states. Like GRU, LSTM completely fails to predict class 3, with an F1 score of (0.000). This suggests that the used structure of LSTM model struggles with complex transitions and minority classes in the dataset.

6.1.1.3.2.2 AUC and ROC Analysis

The AUC and receiver operating characteristic (ROC) curves provide an overview of how well each model distinguishes between different occupancy states. The higher the AUC, the better the model's ability to correctly classify instances of that class. The five models evaluated with their respective AUC values for class 0, class 1, class 2, and class 3.

As seen in the AUC calculation in Table 6-3, among all the occupancy states, the TCN model performed the best, with the highest AUC values. It was particularly effective for the complex states of a fully occupied level, where its AUC stands at (0.99), hence proving its strong temporal dependency and subtle transitions capture ability in occupancy data. The 1D-CNN model also fares very well; it mostly follows after the TCN in most cases, especially for empty and near-empty states' identification, with an almost-perfect AUC of (0.99) for both.

Table 6-3 AUC Calculation for All Models

	1D-CNN	**TCN**	**Ensemble**	**LSTM**	**GRU**
Class 0	0.99	0.99	0.98	0.97	0.97
Class 1	0.94	0.97	0.90	0.87	0.86
Class 2	0.95	0.98	0.92	0.92	0.91
Class 3	0.97	0.99	0.93	0.92	0.91

Generally, the Ensemble model functions effectively, but usually performs worse than the TCN and 1D-CNN models, majorly on higher occupancy states, with AUCs around (0.90-0.93). While LSTM and GRU work in some cases, they truly fall behind the other models, especially for intermediate occupancy states, where their AUCs drop right down to (0.86-0.87). Lastly, TCN is the best model that grasped the occupancy pattern, while 1D-CNN stands a little behind. The most problematic ones are LSTM and GRU for fine or mid-level occupancy states.

6.1.1.3.2.3 ROC Curve Interpretation

The ROC curves for each model are represented in Figure 6-3, which displays the balance between true positive rate-sensitivity and the false positive rate across varying classification thresholds. The closer the curve is to the upper-left corner, the better the model's performance in classifying that class. The TCN ROC curves are always the closest together, towards the top left corner for all classes-a good indicator of its superior performance in all classes of occupancy. Similarly, ROC curves are also strong for 1D-CNN, though for class 0 and class 3, while for class 1 and class 2, TCN outperforms it. The Ensemble model, although it might combine the powers of 1-D CNN and TCN, does not perform better compared to the latter; the evidence is that the ROC curve for class 3 is further from the top left of the TCN. LSTM and GRU exhibit the weakest ROC curves, especially in class 1 and class 2, where their curves deviate more from the ideal, confirming their struggles with intermediate occupancy states.

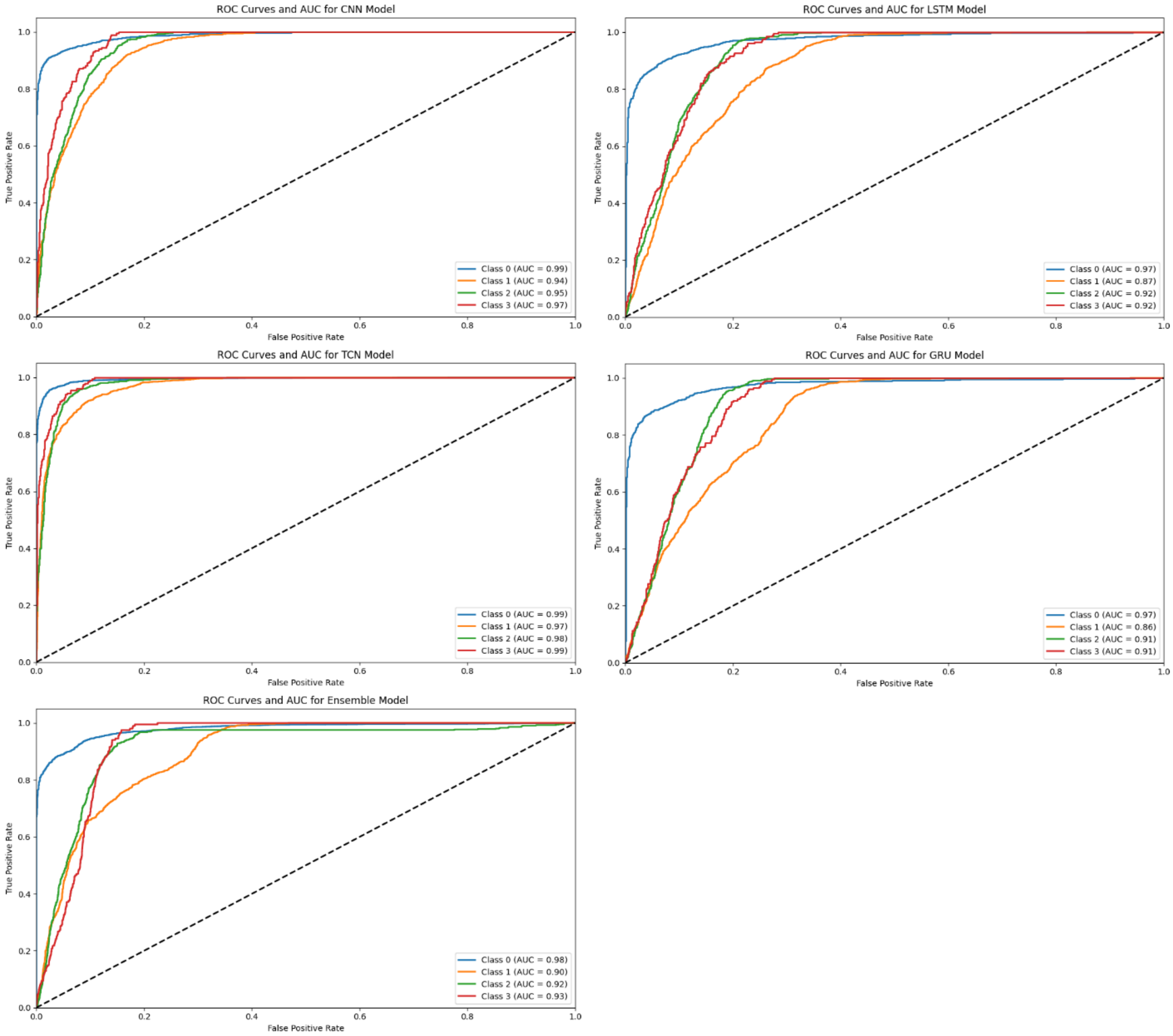


Figure 6-3 ROC Curves and AUC for All Models

Although the Precision and Recall values for some models on Location 3 are close to or equal to zero, the corresponding AUC values remain high. This may appear inconsistent at first glance, but it is statistically valid. The AUC and ROC Curve evaluates the model's ability to distinguish between classes across different thresholds. It is threshold-independent and reflects the model's ranking capability. In contrast, Precision and Recall are threshold-dependent metrics. When the predicted probabilities for the positive class are consistently low and fall below the classification threshold, the model may correctly rank positive samples higher (reflected in a high AUC) but fail to classify any correctly as positive at the chosen threshold; resulting in P = 0 and R = 0.

This behaviour is particularly common when the dataset is highly imbalanced, as is the case with Location 3, where the minority classes are underrepresented. Thus, while the AUC reflects the model's underlying ability to separate classes, the low precision and recall reveal its difficulty making accurate classifications under fixed thresholds for rare events.

6.1.1.3.2.4 Confusion Matrices

Confusion matrices help to represent a model's performance clearly and in detail, showing the frequency of each true class label that was correctly or incorrectly predicted. For this task of occupancy classification, a confusion matrix provides quantitative meaning about the number of correct and incorrect predictions for each of the classes, 0-3, and then highlights aspects where the model does well or worse. This is especially useful in imbalanced datasets like this one, where a clear variation in the distribution for the classes is apparent. For instance, class 3 was clearly underrepresented. The results from this matrix were visualized in plots to show how each model distinguishes between similar occupancy states, such as moderate class 2 and high class 3. These classes can be particularly challenging to differentiate using numerical metrics like accuracy. Thus, confusion matrices provide more profound insights into the challenges of classification, and we can identify which models act best for certain classes, therefore creating a much more detailed picture of model performances.

The confusion matrices for the five models displayed in Figure 6-4 provide further validation of the classification report, AUC scores, and ROC curves, clearly illustrating how well each model predicts the four occupancy classes (0 to 3). The matrices show the exact number of correctly and incorrectly classified instances for each class, confirming the models' relative strengths and weaknesses in predicting specific occupancy levels.

The 1D-CNN model exhibits high precision, recall, and F1 scores, with all values being higher than (0.94), demonstrating high performance in the classification of the empty state. In contrast, the model makes many errors while classifying the fully occupied state, correctly classifying only (77) out of (202) instances as this class and yielding low recall and F1 scores for this class. In fact, the model often confuses class

3 with class 2, reflecting its inability to distinguish between moderate and full occupancy levels.

The overall best performance comes from the TCN model, which optimally classifies empty occupancy states (class 0) and complex full occupancy states (class 3), where it correctly predicts (133) out of (202) instances with an F1 score of (0.711). This may be attributable to TCN having strong temporal modelling, but it most importantly handles the distinguishing between states of moderate to high more effectively, while the other models notably struggle with this.

The Ensemble model gives fair performance for empty states, but exhibits huge difficulties with complex states of occupancy; class 3 hardly reaches a correct classification, as nearly all instances are classified as class 2. This results in an F1 score of (0.000) for class 3, showing that the ensemble approach fails to improve the performance for complex classes compared to TCN alone. Class 1, class 2, and class 3 are the major weaknesses of this model. It misclassifies several cases belonging to these classes. As a matter of fact, it fails to predict any instances in class 3 completely, thus further fortifying the poor performance of the model for complicated transitions between occupancy levels.

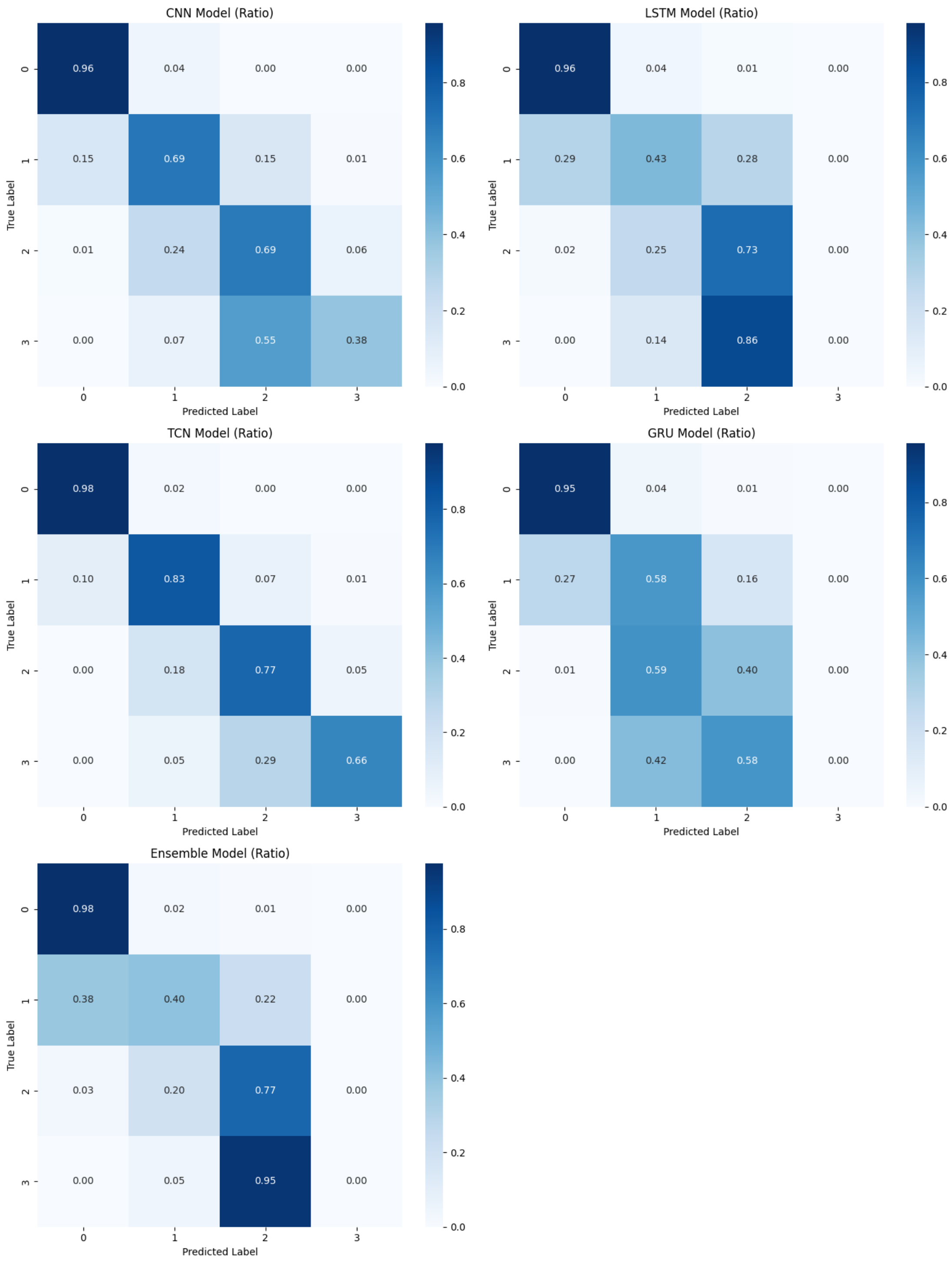


Figure 6-4 Confusion Matrix for All Models

While the high occupancy states pose a big challenge for the GRU model, it classifies most instances of class 3 to be class 2. Similar to LSTM, GRU fails to handle subtle differences in higher occupancy states well, as reflected by the F1 score being exactly (0.000) for class 3, with generally lower AUC.

6.1.1.4. Scenario 2: Specific Location Model Training (Scen-2)

In this scenario, the full input data was split by location into three sub-input data. Accordingly, each model from the five models was trained on the three sub-input data separately. This experiment yielded three trained versions of each model, each targeting a specific location. The models' performance for each location was then evaluated. As a result, three distinct models were produced, with each model tailored to a different location. The metrics obtained from the evaluation of each specific-location model were compared with the results obtained in Scen-1 to assess the model's generalization abilities.

The models used the *Adam* optimizer, which adaptively adjusts the learning rate during training to accelerate the adaptation process. Considering the challenge in multi-class categorization, the models optimized the categorical *cross-entropy* loss function. The training was set for (40) epochs, with an early termination criterion implemented after (20) epochs without improvement. Early stopping was employed to prevent overfitting, ensuring the model halted training as its performance on the validation data plateaued, hence saving time and reducing the risk of overfitting. Validation accuracy and loss were monitored to avert overfitting. *The ModelCheckpoint* callback was utilized for each model, retaining the optimum version based on validation accuracy. This ensured that the final model utilized for testing exhibited the optimal performance derived from the training scenario. The training strategy facilitated real-time observation of training and validation accuracy, enabling ongoing assessment of model performance across epochs.

6.1.1.4.1. Models' Training Performance

Figure 6-5 shows the results for training the models (1D-CNN, TCN, Ensemble, LSTM, and GRU) separately with the corresponding data from each location (Loc-1, Loc-2, and Loc-3), to evaluate how each model adapts to the special characteristics of the locations.

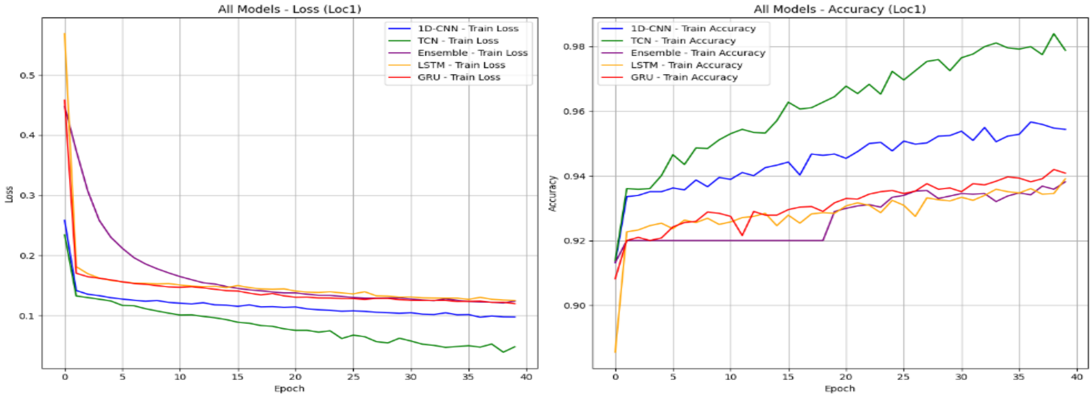
Loc-1: The TCN model has the best performance overall at Loc-1 because it has the lowest training loss and highest accuracy, nearing (98%) at the end of training. This shows that the TCN model really has caught the temporal patterns in data under Loc-1. The 1D-CNN model also sees good performance, reaching about (94%) accuracy and keeping the loss low, which indicates it is fitting well to this location. The Ensemble model follows, with competitive accuracy at a slightly higher loss than the TCN and 1D-CNN. Both LSTM and GRU have larger losses and lower accuracies, meaning they struggle more with the specific data attributes of Loc-1.

Loc-2: At Loc-2, the TCN again performs with the highest accuracy and lowest loss, similar to its performance in Loc-1, meaning it is good at capturing patterns specific to this place. Its accuracy is about to touch (95%) with a minimal loss at the end of the training process. The 1D-CNN model's accuracy is also high but slightly lower than TCN's; the loss will not be high either. The Ensemble model gives an increase in accuracy, which closely follows TCN and 1D-CNN. LSTM and GRU models follow similar trends observed in Loc-1 with lower accuracy and a higher loss, thereby showing that both models underperform consistently when handling the data for Loc-2.

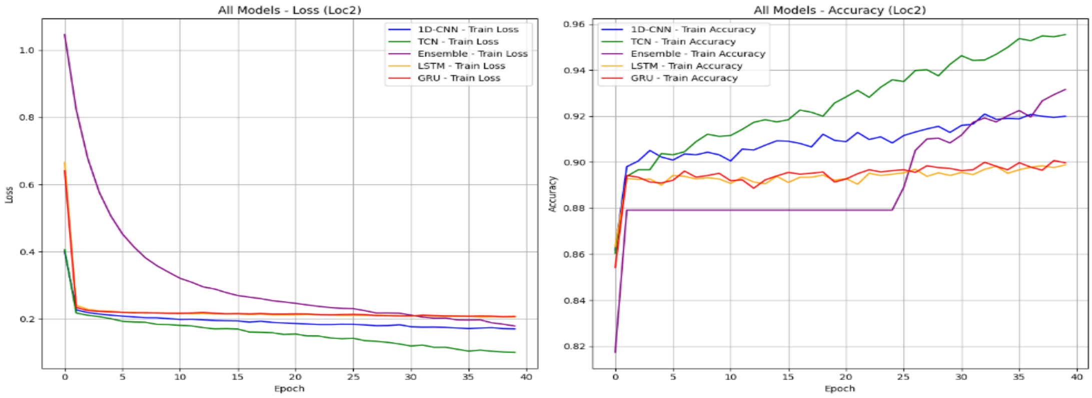
Loc-3: The data from Loc-3 is more challenging, and it can be noticed that all models have a slightly worse performance compared to the previous locations. Still, the TCN model performs the best, with an accuracy of approximately (85%) and the lowest training loss; this shows the adaptability of this model, even on harder data. The 1D-CNN model has accuracy over (80%) but with a slightly higher loss. The Ensemble model gives quite reasonable performance, although it does not match TCN or 1D-CNN. On the other hand, the LSTM and GRU models, are performing quite badly, with both models failing to exceed (70%) accuracy and showing relatively high training loss, which indicates a potential difficulty in capturing the temporal features in Loc-3 data.

Across all three locations, the TCN model consistently performs best, with high accuracy and low loss, which means it is strong in handling unique characteristics in the data at each location. The 1D-CNN model follows closely with good adaptability and competitive performances. The Ensemble model has moderate success but, more often than not, lags behind the TCN and 1D-CNN models. The LSTM and GRU models have a lower accuracy with a higher loss at all locations, indicating that they might not be well suited to capture the specific temporal patterns of this task. These results underline the versatility and potential of TCN as the most effective model for location-specific occupancy prediction.

(a) Loc-1



(b) Loc-2



(c) Loc-3

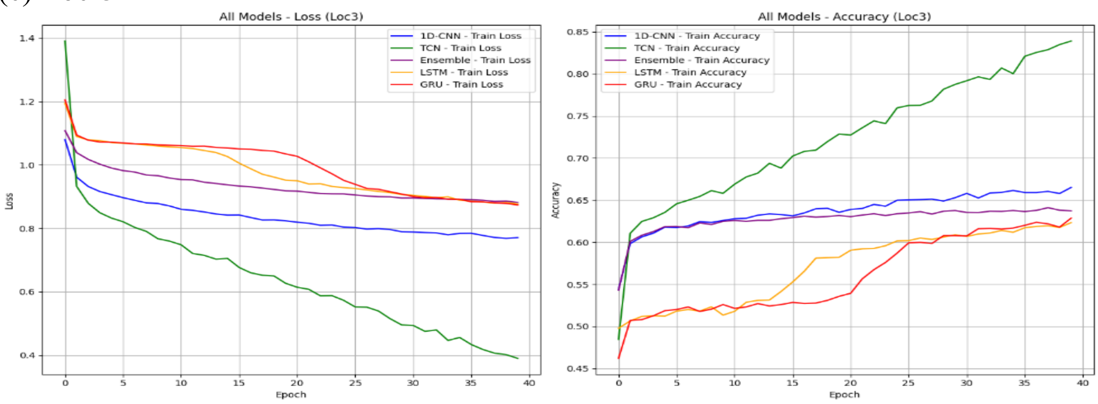


Figure 6-5 Models' Loss and Accuracy During the Training Epochs, (a) Loc-1, (b) Loc-2, and (c) Loc-3

6.1.1.4.2. Models' Performance on Unseen Data

Table 6-4 shows the accuracy and loss data for the models when trained and tested on the same location-data. When comparing Loc-1 and Loc-2, performance stayed strong throughout most models. Loc-1 generally showed higher accuracy and lower test loss compared to Loc-2, although the differences were minor. Surprisingly, the Ensemble model, which combines the 1D-CNN and TCN models, did not perform as well as the individual models in these locations. This may suggest possible an unsuccessful grouping of the two model types in this task.

Table 6-4 Test Accuracy and Loss Comparison

	Test Accuracy			**Test Loss**		
Model	L1	L2	L3	L1	L2	L3
1D-CNN	0.95	0.92	0.61	0.117	0.169	0.877
TCN	0.97	0.93	0.79	0.109	0.174	0.899
Ensemble	0.92	0.90	0.62	0.403	0.212	0.892
LSTM	0.94	0.90	0.55	0.145	0.205	1.006
GRU	0.94	0.90	0.54	0.149	0.207	1.051

6.1.1.4.3. Scen-2 Classification Report

This classification report demonstrates the recorded classification scores and metrics. The report assists to evaluate the models' performance as well as the model selection for this scenario. The report contains the evaluation metrics, AUC, and confusion matrices.

6.1.1.4.3.1 Evaluation Matrices Table

The metrics related to classification in Table 6-5, namely precision, recall, and F1 score, all vary across the three different locations: Loc-1, Loc-2, and Loc-3. Each model performs well and poorly with respect to a certain category of occupancy and the locations in question.

Table 6-5 Classification Matrices for All Models

	Model	precision					recall					F1					support			
		0	*1*	*2*	*3*	*W-P*	*0*	*1*	*2*	*3*	*W-R*	*0*	*1*	*2*	*3*	*W-F1*	*0*	*1*	*2*	*3*
Loc-1	*1D-CNN*	0.974	0.673	0	0	0.948	0.969	0.729	0	0	0.948	0.971	0.7	0	0	0.947	1599	144	3	0
	TCN	0.975	0.852	0.5	0	0.964	0.989	0.722	0.333	0	0.966	0.982	0.782	0.4	0	0.965	1599	144	3	0
	Ensemble	0.916	0	0	0	0.839	1	0	0	0	0.916	0.956	0	0	0	0.876	1599	144	3	0
	GRU	0.956	0.661	0	0	0.930	0.977	0.514	0	0	0.937	0.966	0.578	0	0	0.932	1599	144	3	0
	LSTM	0.95	0.685	0	0	0.927	0.982	0.438	0	0	0.935	0.966	0.534	0	0	0.929	1599	144	3	0
Loc-2	*1D-CNN*	0.955	0.657	0.25	0	0.919	0.966	0.608	0.125	0	0.923	0.961	0.632	0.167	0	0.921	1548	189	8	1
	TCN	0.946	0.735	0.5	0	0.921	0.981	0.529	0.375	0	0.929	0.963	0.615	0.429	0	0.922	1548	189	8	1
	Ensemble	0.943	0.556	0	0	0.896	0.955	0.524	0	0	0.903	0.949	0.54	0	0	0.900	1548	189	8	1
	GRU	0.939	0.551	0	0	0.892	0.957	0.487	0	0	0.901	0.948	0.517	0	0	0.896	1548	189	8	1
	LSTM	0.942	0.541	0	0	0.894	0.952	0.524	0	0	0.901	0.947	0.532	0	0	0.897	1548	189	8	1
Loc-3	*1D-CNN*	0.613	0.696	0.561	0.488	0.616	0.253	0.621	0.783	0.233	0.611	0.358	0.656	0.654	0.316	0.595	150	749	667	180
	TCN	0.767	0.817	0.786	0.724	0.791	0.66	0.822	0.804	0.728	0.792	0.71	0.82	0.795	0.726	0.791	150	749	667	180
	Ensemble	0	0.691	0.578	0.51	0.570	0	0.65	0.814	0.289	0.620	0	0.67	0.676	0.369	0.584	150	749	667	180
	GRU	0.692	0.573	0.507	0.6	0.561	0.12	0.605	0.703	0.017	0.540	0.205	0.589	0.589	0.032	0.499	150	749	667	180
	LSTM	0.561	0.59	0.523	0.625	0.566	0.153	0.63	0.703	0.028	0.555	0.241	0.609	0.6	0.053	0.517	150	749	667	180

Loc-1 Performance: In Loc-1, the models are effective in predicting category 0, which is the majority class, since most of the models' precisions, recalls, and F1 scores are close to (1.0). Among them, the TCN model is outstanding for category 0, having the highest F1 score of (0.982), hence reflecting the ability of correct recognition of this category. The classes 1 and 2, F1 scores significantly drop. Precisely, the performance of 1D-CNN gives an F1 score of (0.700) for class 1 and fails in the detection of classes 2. The Ensemble model, though it turned out perfectly for class 0 with (1.0) recall, failed in the prediction of classes 1, 2; this may indicate the limit of distinguishing other classes because of the domination of one class. This performance implies that even though models can predict high-frequency categories in Loc-1 quite reliably, the minority classes suffer greatly, thereby showing the need for further model tuning or the use of some balancing techniques to deal with imbalanced data.

Loc-2 Performance: In Loc-2, the results are also on the high precision, recall, and F1 scores in class 0 for all models tested. The TCN model outperformed them again, and the F1 score was (0.963) in class 0, which means this model also has a good predictive capability for the dominating class. For its effectiveness of prediction for class 1 and more, it is unstable. Like Loc-1, the performance in predicting the minor categories (category 2 and 3) remains very weak, with zero recall and F1 score metrics from most of the models. The inadequacies noticed in categories 2 and 3 would suggest that models may not generalize well for the less frequent occupancy states in Loc-2. This is probably due to class imbalance or a lack of variability within the training data for these categories.

Loc-3 Performance: For most models, Loc-3's performance was rather fair among the different classes, while still sustaining low performance compared to Loc-1 and Loc-2. The TCN model has the highest values for most categories: (0.710) for category 0, (0.820) for category 1, (0.795) for category 2, and (0.726) for category 3, hence solidifying its strength in distinguishing between all four levels of occupancy. This may be indicative that TCN serves better for complex data distributions where the categories are balanced. For Loc-3, the 1D-CNN model performed very poorly by achieving an F1 score of just (0.358) for category 0, while for the rest of the categories, the score continues to fluctuate. That clearly shows its inability to adapt to added complexity in data from this location. The Ensemble model performs inconsistently in Loc-3, even though it does well for category 2, with its recall at (0.814), but for categories 0 and 3, this recall is low. Similarly, though GRU and LSTM show decent F1 scores over categories, they are considerably inferior to those from TCN. For example, the predicted scores for classes 0, 1, 2, and 3 are (0.205), (0.589), (0.589), and (0.032), respectively, indicating that the GRU model is finding the effective modelling of the data distribution for Loc-3 difficult.

Overall, though all models performed very well in the majority class (category 0) of both Loc-1 and Loc-2, they performed poorly on the minority classes, which reduced the precision, recall, and F1 score for categories 1, 2, and 3. From the above, TCN has shown consistent better generalization performance on the three locations, and especially for Loc-3, it had a relatively balanced performance on the different categories, which may indicate better suitability for a complex and diverse dataset. These results highlight the importance of selecting models based on the inherent features of location-specific data and further pinpoint the requirement for additional approaches such as balancing data and tuning the model to improve performance for underrepresented classes.

To test models' generalization, in another experiment, each trained model for each specific location was tested on combined test data, containing (10%) unseen data from each location. The combined test data allows to test model's generalization. The average accuracy and loss for each location-model then was calculated. The classification report resulted in this experiment can be seen in Table 6-6.

Table 6-6 Classification Report for All Models Among the Three Locations

	Model	**precision**					**recall**					**F1**					**support**			
		0	*1*	*2*	*3*	*W-P*	*0*	*1*	*2*	*3*	*W-R*	*0*	*1*	*2*	*3*	*W-F1*	*0*	*1*	*2*	*3*
Loc-1-Models	*1D-CNN*	0.646	0.575	0	0	0.527	0.964	0.168	0	0	0.641	0.774	0.26	0	0	0.541	1647	549	333	90
	TCN	0.653	0.636	0.667	0	0.629	0.971	0.191	0.006	0	0.651	0.781	0.294	0.012	0	0.554	1647	549	333	90
	Ensemble	0.629	0	0	0	0.396	1	0	0	0	0.629	0.772	0	0	0	0.485	1647	549	333	90
	GRU	0.643	0.497	0	0	0.509	0.954	0.158	0	0	0.633	0.768	0.24	0	0	0.533	1647	549	333	90
	LSTM	0.63	0.462	0	0	0.493	0.997	0.011	0	0	0.629	0.772	0.021	0	0	0.490	1647	549	333	90
Loc-2-Models	*1D-CNN*	0.646	0.502	0	0	0.511	0.94	0.202	0	0	0.633	0.766	0.288	0	0	0.542	1647	549	333	90
	TCN	0.659	0.579	0.333	0.333	0.590	0.96	0.226	0.003	0.011	0.652	0.782	0.325	0.006	0.022	0.561	1647	549	333	90
	Ensemble	0.629	0	0	0	0.396	1	0	0	0	0.629	0.772	0	0	0	0.485	1647	549	333	90
	GRU	0.646	0.491	0	0	0.509	0.937	0.204	0	0	0.632	0.765	0.288	0	0	0.541	1647	549	333	90
	LSTM	0.644	0.477	0	0	0.505	0.937	0.193	0	0	0.630	0.764	0.275	0	0	0.538	1647	549	333	90
Loc-3-Models	*1D-CNN*	0.969	0.29	0.195	0.13	0.699	0.058	0.568	0.751	0.233	0.259	0.109	0.384	0.31	0.167	0.194	1647	549	333	90
	TCN	0.941	0.324	0.234	0.18	0.696	0.107	0.61	0.715	0.756	0.312	0.192	0.423	0.352	0.291	0.264	1647	549	333	90
	Ensemble	0	0.216	0.202	0	0.071	0	0.581	0.691	0	0.210	0	0.314	0.313	0	0.106	1647	549	333	90
	GRU	0	0.146	0.211	0	0.057	0	0.41	0.685	0	0.173	0	0.216	0.322	0	0.086	1647	549	333	90
	LSTM	0	0.205	0.181	0	0.066	0	0.488	0.712	0	0.193	0	0.288	0.289	0	0.097	1647	549	333	90

From the classification report above, it can be clearly seen that across models and all locations the precision, recall and f1 scores are significantly higher for class 0 compared to other classes, especially class 2 and class 3. Also, across all models and locations, the precision, recall, and F1 scores for class 3 are extremely low. Some models, like TCN and 1D-CNN, show relatively consistent performance across different locations for class 0. However, their performance still drops significantly for other classes. Ensemble, GRU and LSTM models perform poorly across classes, particularly for class 3, where their F1 scores are close to zero.

6.1.1.4.3.2 AUC and ROC Analysis

The AUC graphs shown in

Figure 6-6 ROC Curves for All Models in All Locations

for the models across three locations provide valuable insights into model performance for each class and location.

For **Loc-1**, the TCN model achieved the highest precision, recall, and F1 scores for classes 0 and 1, with an impressive AUC score nearing (0.99) for both classes. The 1D-CNN model closely followed, performing well with AUCs around (0.97). Both models demonstrated strong predictive capabilities, while the Ensemble model showed weaker performance in class 1, reflected in its lower recall and precision values. The GRU and LSTM models displayed competitive results, though slightly lagging TCN and 1D-CNN, particularly in their AUCs and confusion matrix results.

In **Loc-2**, the TCN and 1D-CNN models continued to perform strongly, especially in class 0, achieving AUC scores around (0.96) and (0.95), respectively. The Ensemble model exhibited some inconsistencies, particularly in class 1, where its recall was lower. The LSTM and GRU models, while decent in class 0, struggled with lower performance in other classes, as seen in both the confusion matrices and classification reports. Notably, the TCN model maintained an edge over the others, showing robust generalization.

Loc-3 presented the most challenging scenario, with overall lower accuracy across all models compared to the other two locations. The TCN model, once again, performed the best, particularly in class 0, where it achieved an AUC of (0.95). However, all models struggled significantly with class 1, with much lower AUCs (below 0.8) and recall values across the board, as reflected in the confusion matrices. The 1D-CNN and Ensemble models showed declining performance in this location, particularly in class 2, where AUC scores dropped to (0.76), indicating greater difficulty in making accurate predictions for this class.

As observed in Table 6-6 and Figure 6-6, certain models for Location 3 report zero Precision and Recall while maintaining relatively high AUC values. This is a reflection of the different ways these metrics assess performance. AUC reflects the model's discrimination power over all thresholds, while Precision and Recall are influenced by the specific decision threshold and can yield zero when positive instances are missed, particularly under imbalanced class distributions. This highlights the importance of interpreting AUC in conjunction with threshold-sensitive metrics.

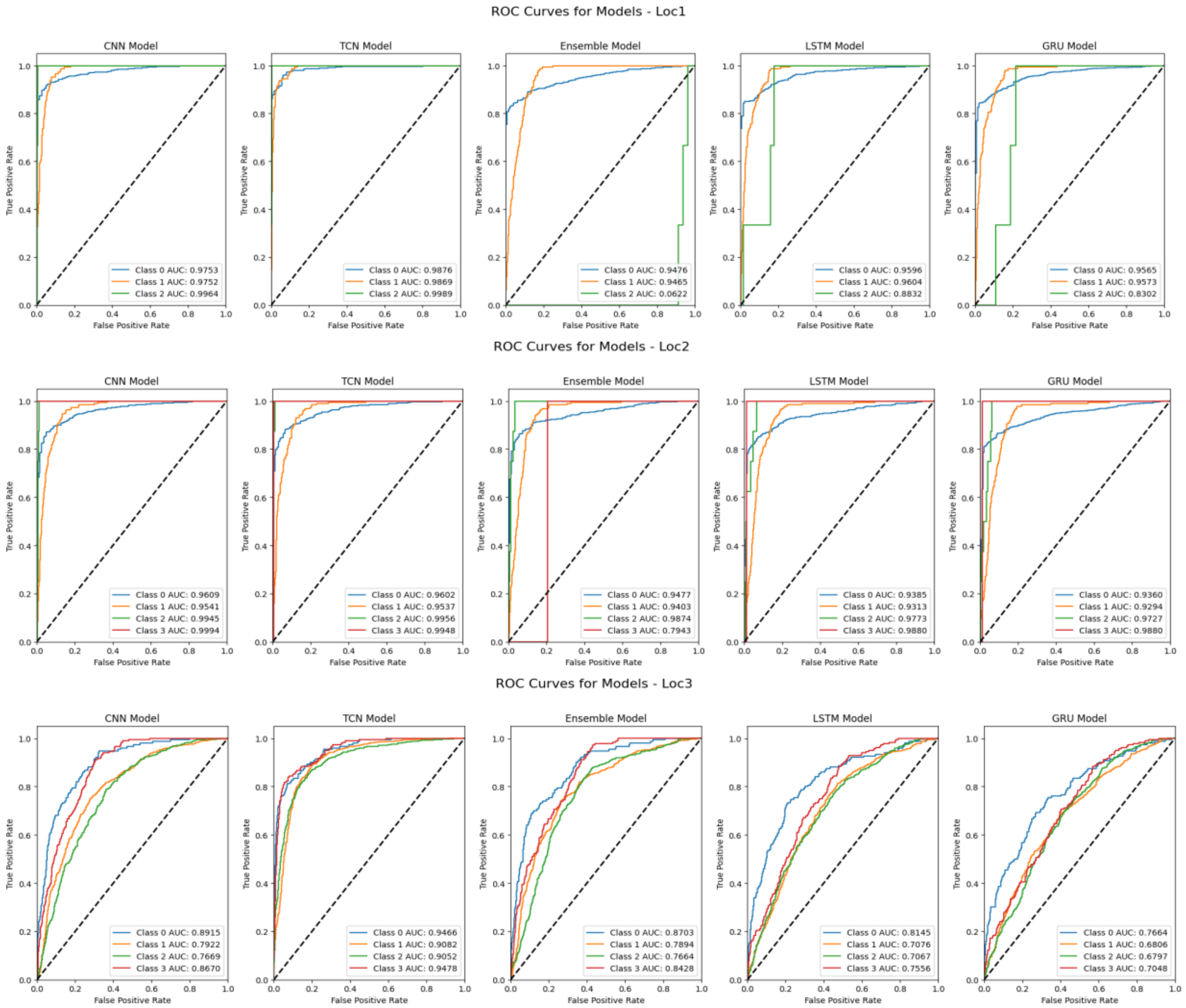


Figure 6-6 ROC Curves for All Models in All Locations

6.1.1.4.3.3 Confusion Matrices

The confusion matrices shown in Figure 6-7 for each model at Loc-1, Loc-2, and Loc-3 provide a broad insight into the classification performance and highlight the differences in accuracy and misclassification patterns. All models perform variably in forecasting the occupancy categories ranging from 0 to 3 for the locations under consideration, there are substantial variations noticeable in Loc-3. It should be noted that since there is no representation of class 3 in the training data for site 1, the results only display classes 0, 1, and 2.

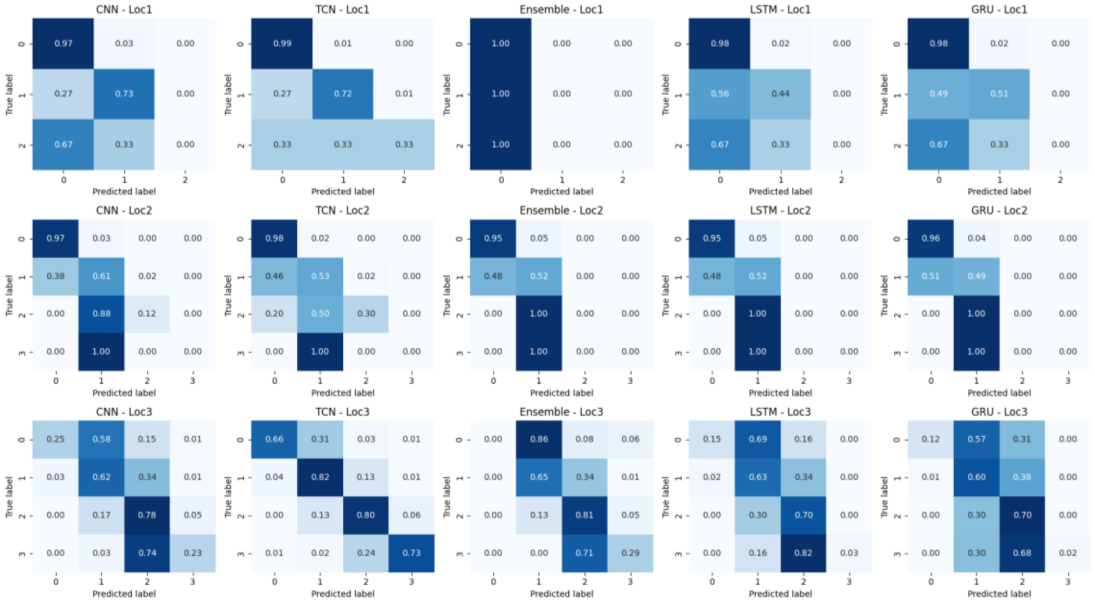


Figure 6-7 The Confusion Matrix for All Models in All Locations

Loc-1 Performance: Loc-1 presents the strong performance of all the models, especially for class 0, as the highest number of instances is rightly predicted for it. For instance, the 1D-CNN model has predicted (1549) true positives of class 0, with very negligible misclassifications in the other classes. At the same time, the TCN and Ensemble models performed also well for class 0, correctly classifying (1582) and (1599), respectively; hence, these models are good in finding this majority class. However, minor misclassifications occur in categories 1 and 2 across models. The LSTM model, for instance, correctly predicts 81 instances in category 1, but it does have a few misclassifications, where some instances are classified under category 0. The GRU model also has the same kind of misclassifications, with about 70 correct in category 1, and it shows slight confusion against category 0. The models are very accurate at Loc-1, mainly separating category 0 from other categories, although some problem remains with distinguishing categories 1 and 2 accurately.

Loc-2 Performance: In Loc-2, the models are very consistent and show high levels of accuracy in their prediction of category 0. For example, the TCN model correctly predicts (1518) instances in category 0, while the Ensemble model is comparably high with its (1478) true positives in the same category. This trend follows from what has been established in Loc-1 and therefore indicates that category 0 is usually the one that remains the most predictable across different locations. However, there is a slightly higher misclassification rate across categories 1 and 2 compared to Loc-1. More specifically, the 1D-CNN model misclassifies some instances in category 1, correctly classifying (71), while a few instances were misclassified as either category 0 or category 3. A similar trend can be observed from the LSTM and GRU models, with minor misclassifications between categories 1 and 2, thus highlighting that discerning among the middle categories remains a difficult task. Despite these minor difficulties, the general performance of classification in Loc-2 remains remarkably high, especially regarding the dominating class, class 0.

Loc-3 Performance: Loc-3 was harder for all of the models, as reflected in the confusion matrices. All the models studied here show striking reductions in their correct classification rates for category 0 compared to Loc-1 and Loc-2. For example, the 1D-CNN model is only able to correctly classify (38) instances in category 0, while it makes profound misclassifications in categories 1, 2, and 3. It has also been noticed that all the models run this pattern, where TCN also shows a similar distribution, correctly classifying (99) instances in category 0 and showing marked confusion in other categories. Aggregations of misclassifications occur in the lower diagonal cells of the matrices, showing the ambiguity that is especially apparent in distinguishing neighbouring occupancy levels within this context. For example, category 1 in the Ensemble model records (487) correct classifications but gives way to considerable overflows into categories 0 and 2. On a similar note, the GRU model performs worse in Loc-3, high misclassifications among the categories, especially category 2, where there are (285) instances predicted to belong to category 1.

It follows from the results that all models behave well for Loc-1 and Loc-2, having a high recognition rate for the majority class, category 0. Most of the differences in performance were between the TCN and Ensemble models; this reflects a kind of robustness in capturing the prevailing occupancy pattern. In contrast, Loc-3 presents certain challenges, with all models finding it hard to achieve the same levels of accuracy, as depicted by the increase in misclassifications among the categories. Issues seen with Loc-3 may relate to specific data patterns or more complex occupancy dynamics at this site. This again points to the fact that further model calibration or location-specific features might be required for overall good classification performance.

In summary, all models achieved robust performance for both Loc-1 and Loc-2, but their performance is severely degraded at Loc-3, which has a highly imbalanced set due mainly to the minority and intermediate occupation classes. These outcomes suggest that local occupation features are far more important, and that additional contextual information may be required to boost the accuracy of classification in complex locations.

6.1.2. Phase 2: Predictive Modelling as a Regression Task

This phase focusses on the regression task, where the goal is to predict the total occupancy as a continuous value rather than classifying it into predefined categories. To estimate the availability more precisely, all five models (1D-CNN, TCN, LSTM, GRU, and the Ensemble model (TCN + LSTM)) were trained on the exact number of occupied charging stations.

6.1.2.1. Models Description

The Ensemble model combining TCN and LSTM was chosen to leverage the strengths of both architectures. This synergy enables TCN to model long-range temporal dependencies and find subtle patterns in sequential data, while LSTM can deal with short-term dependencies and make sure that recurrent layers do not have problems with vanishing gradients. By combining them, the Ensemble model is able to see both long-term and short-term changes in occupancy patterns. This could improve its ability to predict what will happen in the regression task.

First, the five models were trained on the full dataset, which contains records from three different locations, similar to the previous classification task. As a last step, the impact of location-specific training was also examined, and the models were trained on splits containing data from only one location. This produced three variants for each of the five architectures. The best case (whether a single model, trained on the full dataset, outperforms models that are trained by location regarding predictive accuracy) is determined by comparing all the different evaluation results.

Table 6-7 Summary of Model Architectures

Model	Architecture Summary	Common Training Settings
1D-CNN	Input layer 1D Convolutional Layer (64 filters, kernel size = 3, ReLU activation) MaxPooling Layer (pool size = 2) Flatten Layer Dense Layer (100 units, ReLU activation) Output Layer (1 unit for regression output)	Optimizer: Adam. loss function: MSE Evaluation metric: MAE
TCN	Input layer 1D Convolutional Layer (64 filters, kernel size = 3, causal padding, ReLU activation) 1D Convolutional Layer (64 filters, kernel size = 3, causal padding, ReLU activation) MaxPooling Layer (pool size = 2) Flatten Layer Dense Layer (100 units, ReLU activation) Output Layer (1 unit for regression output)	
Ensemble (TCN + LSTM)	TCN Block: 1D Convolutional Layer (64 filters, kernel size = 3, causal padding, ReLU activation) 1D Convolutional Layer (64 filters, kernel size = 3, causal padding, ReLU activation) Flatten Layer LSTM Block: LSTM Layer (64 units, ReLU activation) Concatenation of TCN and LSTM outputs Dense Layer (100 units, ReLU activation) Output Layer (1 unit for regression output)	
LSTM	Input layer LSTM Layer (64 units, ReLU activation) Dense Layer (100 units, ReLU activation) Output Layer (1 unit for regression output)	
GRU	Input layer GRU Layer (64 units, ReLU activation) Dense Layer (100 units, ReLU activation) Output Layer (1 unit for regression output)	

6.1.2.2. Scenario 1: Models Trained on a Combined Dataset (Scen-1)

6.1.2.2.1. Preparation

Since the target value in this scenario is the Total Occupation value "Continuous Variable", the input dataset was changed, with the removal of the categorical feature "*OccCate*" and the target value being moved to point to "*TotalOccupied*". The training epochs were set to (30), while validation loss was monitored to avoid overfitting. A

ModelCheckpoint callback was utilized for each model, which retained the best version, predicated on validation accuracy. This ensured that the final model utilized for testing represented the best performance from within the training scenario. In addition, the training procedure allows for training and validation loss metrics, therefore enabling real-time assessment of model efficiency over time across epochs.

The input data was split in a cross-location manner as shown in Figure 6-8, where each sub-dataset contains data from all three locations. More precisely, the training data consisted of (70%) of the data from Loc-1, combined with (70%) from Loc-2 and (70%) from Loc-3. Analogously, (15%) of data from each location formed the validation set, while the test set was completed with the remaining (15%) of data from each location. This made for a final split of (70%) for training, (15%) for validation, and (15%) for testing, while ensuring that all data subsets were representative of the three locations.

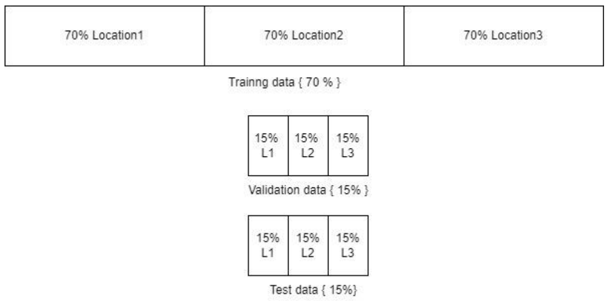


Figure 6-8 Cross-Location Data Split

A cross-location split enhanced the model's generalization, and a similar cross-location test set was tested by training and validating the model on data from all three locations. The model encountered different patterns and characteristics during its training, which are likely to increase to its robustness. This also ensured that the model was tested on data that it had never seen during the training and validation process, thus giving a more realistic estimate of the performance in generalizing to new and unseen data.

Procedure1: pseudocode Split Dataset by Location (Cross-Location Split)

```
procedure SPLIT_BY_LOCATION(dataframe, train_ratio, val_ratio, test_ratio)
  Initialize train_data, val_data, test_data as empty dataframes
  for each unique location in dataframe['location_id'] do
     loc_data ← Filter dataframe where 'location_id' equals current location
     loc_len ← length of loc_data
     train_end ← Integer value of loc_len * train_ratio
     val_end ← train_end + Integer value of loc_len * val_ratio
     loc_train ← First 'train_end' rows of loc_data
     loc_val ← Rows from 'train_end' to 'val_end' in loc_data
     loc_test ← Rows from 'val_end' to the end in loc_data
     Append loc_train to train_data
     Append loc_val to val_data
     Append loc_test to test_data
  end for
  return train_data, val_data, test_data
end procedure
```

In a standard data split, where data is split sequentially, the model in this case would be trained on the majority of Loc-1 and part of Loc-2, and the rest of the locations would be utilized for validation and testing. That would lead to weak generalization and reduced performance by the model, since it would overfit to patterns in some locations without capturing the diversity of all the three locations. By splitting the data across all locations for training, validation, and testing, the model is better positioned for good generalization and performance variations across locations. The model's performance was evaluated on the test set in terms of MSE, MAE, and R^2 values.

6.1.2.2.2. Training and Evaluation Results

The performance of all five models was visually compared to identify the model learning pattern during the running epochs. Figure 6-9 shows the progress of the five models during the training epochs, as described below.

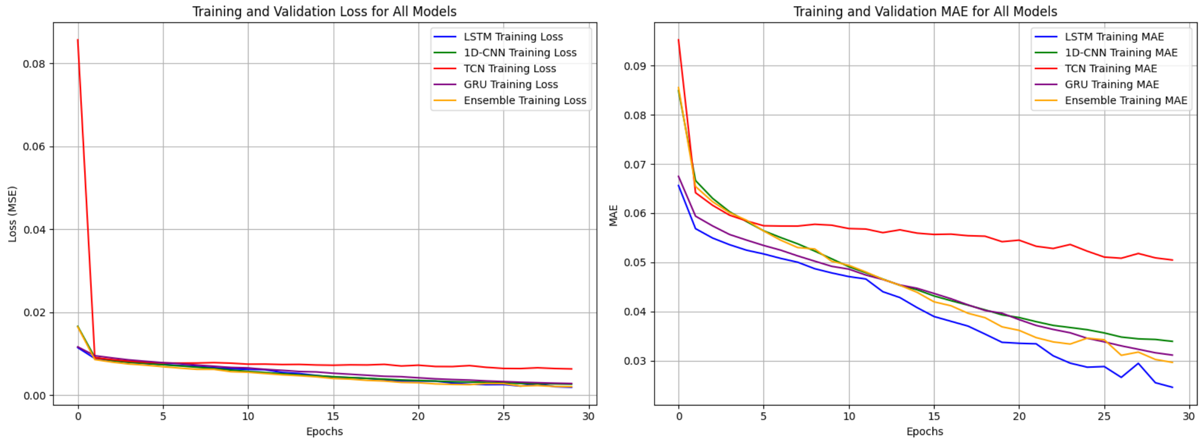


Figure 6-9 Progression of Training Loss (Left) and Training MAE (Right) For Each Model Over the Epochs

Models' Training Performance: Concerning MSE and MAE, Figure 6-9 shows that LSTM, Ensemble, and GRU models were slightly better than the others, with the most efficient and stable learning curve. Their performances exhibited less MSE and MAE on the training dataset compared to 1D-CNN and (especially) TCN. Among these, the LSTM and Ensemble model consistently progressed well in both metrics, and were smoothest at the end; thus, they offered excellent fitting for the training data. Conversely, the TCN model initiates at a very high initial loss with MAE; after a few epochs this does decrease considerably, but remains high compared to the other models during training. That indicates the fact that TCN is poorly fitted to the data, and has a limitation in extracting the pattern while that is not the case with GRU, LSTM, or Ensemble. The 1D-CNN also has relatively higher MAE value at the initial training epochs, after which it is stabilized for some epochs during the training, yielding moderate performance. However, it was clearly outperformed by GRU, LSTM, and the Ensemble models in light of their better convergence and lower final training losses.

Models' Performance on Test Data: Evaluating the models' performance on test data provides a clearer view on models' performance especially with respect to the generalization capability of each model. The results displayed in Table 6-8 show that the Ensemble model achieved the best MSE value with (4.389) on average, hence capturing the underlying pattern in the data and staying accurate upon testing on new data. The second best MSE (4.721) was recorded for the LSTM model, followed by GRU, 1D-CNN, and then TCN, with (4.923), (4.927), and (4.976) (respectively).

In terms of the calculated MAE value for each model, the LSTM shows a balancing in combining both lower MSE and MAE where its MAE value was the lowest and the best with (1.228). Achieving a lower absolute error on a combined dataset highlights the potential to develop a more generalized model. Similar to the LSTM model, the GRU model also achieved the second lowest MAE value with (1.379). The Ensemble model, unlike its MSE value, recorded a moderate absolute error with (1.417) which may reflect some model overfitting the data. Both TCN and 1D-CNN again recorded the highest absolute error with (1.46) and (1.553) respectively. This may confirm that model struggling to train the pattern of the data.

Table 6-8 Models' Performance on Test Data

Model	MSE	MAE
1D-CNN	4.927	1.553
TCN	4.976	1.468
LSTM	4.721	1.228
GRU	4.923	1.379
Ensemble	4.389	1.417

Evaluating models with R2: The R^2 metric makes a difference in the evaluation of model performance by indicating how well the model explains the variance in the target variable. The inclusion of R^2 in this approach along with both MSE and MAE provides for an accessible, interpretable measure of fit that will ensure the models chosen for EV demand and occupancy prediction are capable of capturing the underlying patterns in data in a way that guides practical decision-making. Figure 6-10 shows a comparison of the R^2 values obtained for each model. While the Ensemble model achieved the highest R^2 value with (0.72), the LSTM model attained similar performance with (0.70), and both of these models again outperformed the other three models, which achieved similar R^2 values (0.68).

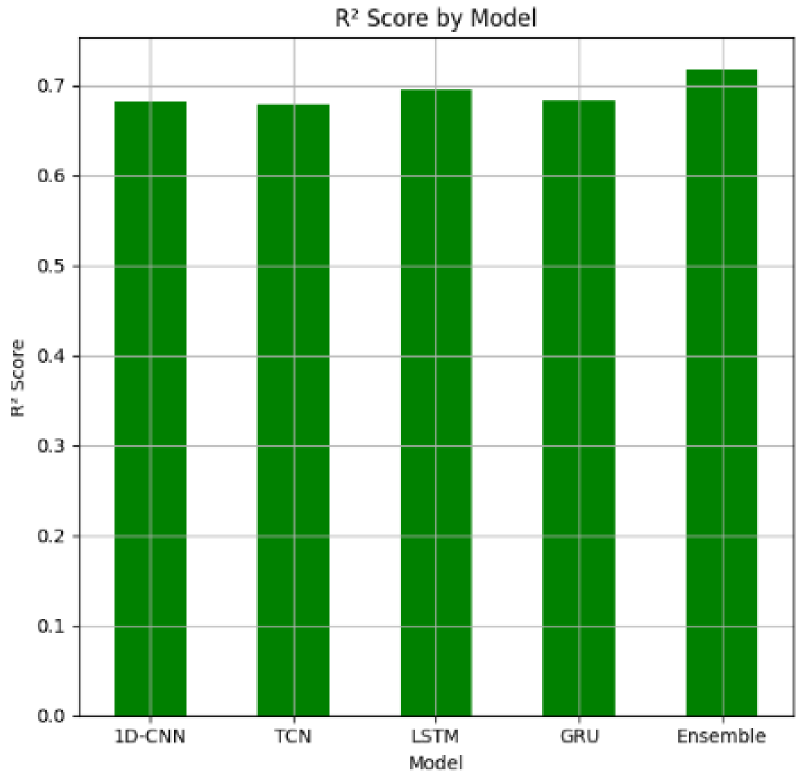


Figure 6-10 Comparison of R^2 Values for Studied Models

6.1.2.3. Scenario 2: Learning on Location-Specific Data (Scen-2)

In this scenario, each model of the five models was trained on three location-specific datasets; each location-dataset contains a historical input data from one specific location. Then the trained models were tested on test data cross locations. The two metrices (MSE and MAE) were shown to evaluate the model performance on unseen data and therefore, evaluate models' ability to generalization. The models' names and the initial settings of the hyperparameters are the same as for Scen-1 (see Table 6-7). The average values of the resultant metrices were calculated to show the overall performance cross all locations.

6.1.2.3.1. Overall Training Performance

Data concerning model progression over the studied training epochs is shown in Figure 6-11, where the first three rows represent the loss curves, and the last three rows represent the MAE curves for every model and location. Each subplot depicts the training and validation performance over (40) epochs.

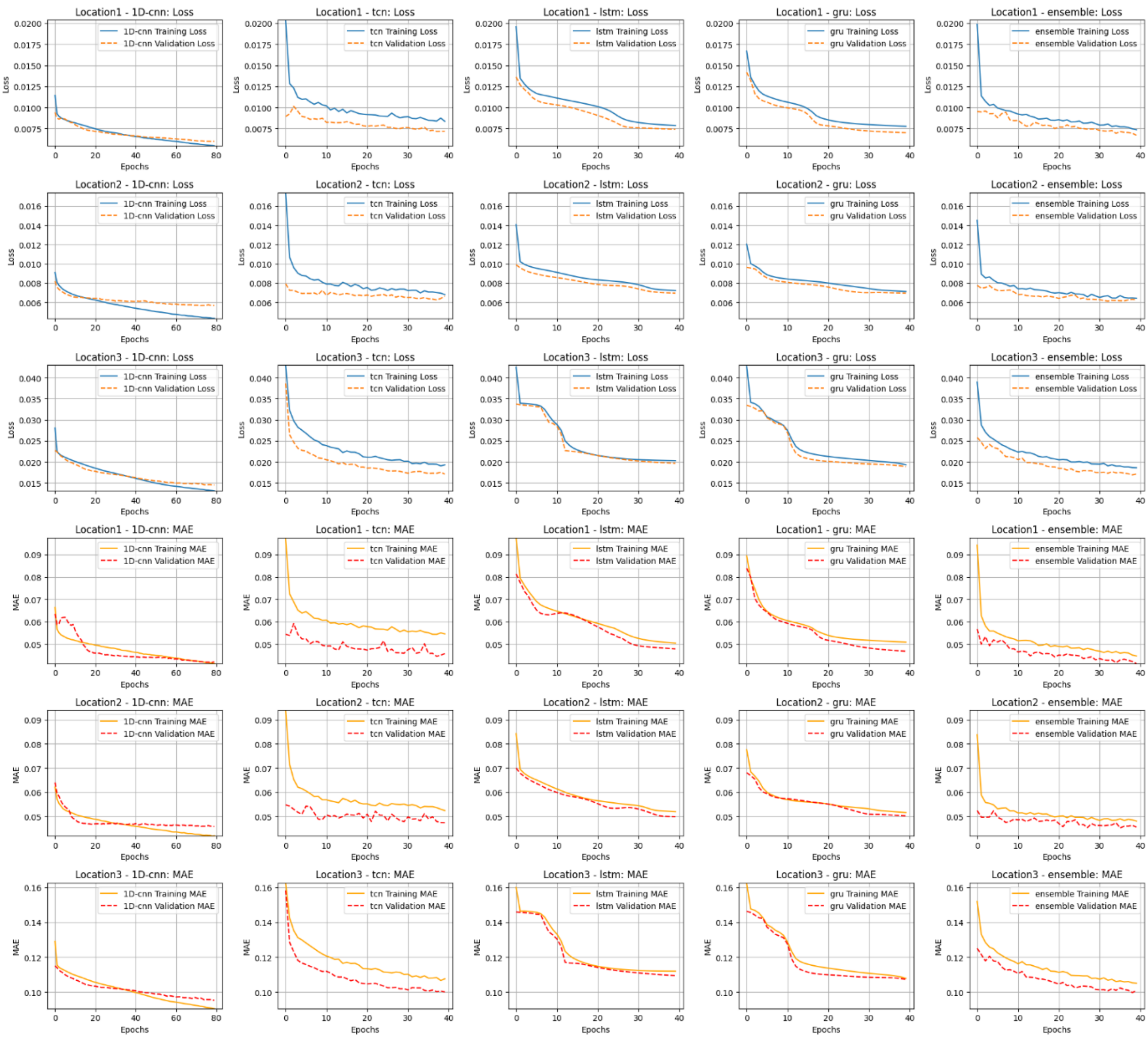


Figure 6-11 Loss, V-Loss, MAE, V-MAE for Each Model by Location

Loc-1

1D-CNN: The training loss decreases from approximately (0.011 to 0.005) over (60) epochs, while validation loss follows a similar trend. The MAE drops steadily to around (0.023), indicating consistent learning performance on training and validation sets.

TCN: Training and validation loss curves converge rapidly within the first (10) epochs and remain stable with minimal divergence (<0.002). Validation MAE closely tracks training MAE across all epochs, maintaining a consistent error range around (0.020 – 0.022).

LSTM: The validation loss plateaus around epoch (27) at approximately (0.0075). The MAE curves are also following a similar pattern.

GRU: The performance is similar to the LSTM trend, as both training and validation losses start to converge earlier. Validation MAE fluctuates slightly but stays within (0.003) of training MAE throughout, suggesting acceptable generalization with mild variance.**Ensemble:** Both loss curves show close convergence, with final training and validation losses around (0.007) and (0.006), respectively. Validation MAE stays within about (0.002) of training MAE, supporting strong generalization performance.

Loc-2

1D-CNN: It can be seen that the training loss is consistently decreasing until it reaches around (0.002), while the validation loss saturates rather early at about (0.006); this could imply overfitting. The same happens to the MAE; the validation MAE is always higher compared to the training MAE.

TCN:TCN performs well in terms of the early convergence of both training and validation loss curves after about (15) epochs. In the MAE plot there is minimal divergence between training and validation with about (0.003), which implies good generalization.

LSTM: A persistent divergence of around (0.001) is observed between training and validation loss after epoch (15). Training MAE remains about (0.003) higher than validation MAE, suggesting mild overfitting.

GRU: GRU performs comparatively better compared to LSTM, with both the training and validation losses stabilized around the same point. In this case, the validation MAE closely follows that of the training MAE; this speaks volumes for its good generalization.

Ensemble: Training and validation loss curves converge to approximately (0.015). Validation MAE remains stable and closely follows the training MAE about (0.019), indicating reliable model performance.

Loc-3

1D-CNN: Training loss decreases to around (0.012), while validation loss remains significantly higher about (0.015), with an MAE gap exceeding (0.004) by epoch (80). This indicates poor generalization.

TCN: TCN trains well with slight divergence between the training and validation loss curves where it reached about (0.003). MAE difference between training and validation remains consistently under (0.002), reflecting stable and generalizable learning

LSTM: The LSTM model has a bit of divergence in training and validation losses. This is where validation loss gets to a higher value than training loss and then stabilizes there. In the MAE plot, the network performs moderately overfitting, as the validation MAE is higher than the training one.

GRU: GRU generalizes well, and training and validation losses converge well at (<0.001) after (40) epochs. The validation MAE is mostly close to the training MAE, indicating good performance on unseen data.

Ensemble: This model shows the best alignment, with both training and validation losses converging to about (0.005) by epoch (20). Validation MAE closely follows training MAE (within 0.001), indicating robust generalization and minimal overfitting.

Among all the locations, TCN and Ensemble models generalize very well, the divergence between training and validation loss and MAE curves is minor. The GRU model is relatively good but with slightly more fluctuation of MAE in validation. The LSTM model has a moderate overfitting problem, especially Loc-3, where the validation loss and MAE are higher than the values in training. Among these three models, the 1D-CNN model has the worst performance. In particular, it suffers from significant overfitting for Loc-3 with generally poor generalization.

6.1.2.3.2. Models' Performance on Test Data

By examining the performance between the models, each trained and tested exclusively at different geographical locations, one could notice a very interesting trend in the accuracy between the three locations, as shown in Figure 6-12 and Table 6-9.

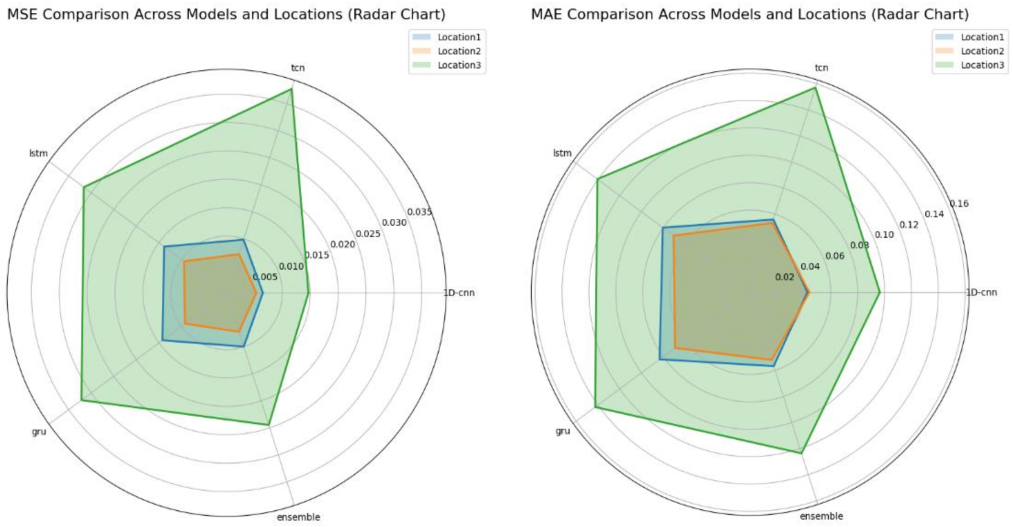


Figure 6-12 MSE and MAE Comparison Across Models and Locations

Table 6-9 Models' Performance on Local Test Data

Location	Model	MSE	MAE
Loc-1	1D-CNN	0.007	0.043
	TCN	0.010	0.056
	LSTM	0.014	0.080
	GRU	0.014	0.083
	Ensemble	0.010	0.057
Loc-2	1D-CNN	0.005	0.044
	TCN	0.007	0.053
	LSTM	0.009	0.070
	GRU	0.009	0.070
	Ensemble	0.007	0.052
Loc-3	1D-CNN	0.015	0.097
	TCN	0.038	0.157
	LSTM	0.032	0.141
	GRU	0.032	0.143
	Ensemble	0.024	0.124

For Loc-1, the best results were those from the 1D-CNN model, which attained the lowest value of MSE (0.007), while MAE was (0.043). The 1D-CNN model thus seems capable of learning the peculiar temporal pattern from Loc-1, and hence allows a good predictability. The performance of the TCN and Ensemble models was relatively good, giving an MSE of (0.010) and MAE of (0.056) and (0.057), respectively, while the LSTM and GRU gave a slightly higher error rate of (0.014) for the MSE and (0.080) and (0.083) for the MAE of both, showing the probable unsuitability of these models to the nature of data for Loc-1.

A similar trend was observed for Loc-2. Again, the 1D-CNN model performed best with an MSE of (0.005) and MAE of (0.044). followed by TCN and Ensemble, with MSE of (0.007) and MAE values of (0.053) and (0.052), respectively, hence showing good strength in the case of Loc-2. However, again LSTM and GRU models produced higher MSEs and MAEs of (0.009) and (0.070), respectively, where both the models recorded an MSE and MAE showing the difficulty of the recurrent architecture in learning the location-specific pattern over Loc-2.

As expected, Loc-3 proved to be more challenging for all models, as evidenced by higher error rates across the board. The results from the 1D-CNN showed relatively better performance compared to other models, though it had a higher MSE of (0.015) with an MAE of (0.097), which is considerably higher compared to the values derived from Loc-1 and Loc-2. For Ensemble, it had a value of (0.024) for MSE and (0.124) for MAE, performing comparatively better than TCN, LSTM, and GRU. Among these, TCN has the highest error contribution with an MSE of (0.038) and MAE of (0.157). A reasonable conclusion would be that Loc-3 has more complex or sophisticated patterns, and to achieve comparable results , more location-based features and/or further model tuning need to be incorporated.

6.1.2.3.3. Cross-Location Models' Performance

Table 6-10 presents the results of each model for different locations (given that they are trained in one location and tested in another). The models trained in Loc-1 and tested in Loc-2 showed medium success. Among them, GRU and LSTM resulted in slightly low MAE values of (0.102) and (0.103), respectively. Nevertheless, all of them presented quite significant drops in their performance when executed on Loc-3, and all architectures gave an MAE of above (0.4). The best MAE of (0.420) was given for the GRU model, but these high error rates also suggest further that models trained on Loc-1 and Loc-2 do not generalize well to Loc-3.

Table 6-10 Cross-Location Model Evaluation

Training Location	**Model**	**Testing Location**	**MSE**	**MAE**
Loc-1	1D-CNN	Loc-2	0.021	0.113
	TCN	Loc-2	0.019	0.113
	LSTM	Loc-2	0.019	0.103
	GRU	Loc-2	0.020	0.102
	Ensemble	Loc-2	0.019	0.106
	1D-CNN	Loc-3	0.220	0.432
	TCN	Loc-3	0.211	0.424
	LSTM	Loc-3	0.226	0.440
	GRU	Loc-3	0.209	0.420
	Ensemble	Loc-3	0.225	0.439
Loc-2	1D-CNN	Loc-1	0.035	0.137
	TCN	Loc-1	0.034	0.130
	LSTM	Loc-1	0.035	0.129
	GRU	Loc-1	0.035	0.126
	Ensemble	Loc-1	0.034	0.129
	1D-CNN	Loc-3	0.196	0.404
	TCN	Loc-3	0.227	0.441
	LSTM	Loc-3	0.227	0.441
	GRU	Loc-3	0.214	0.427
	Ensemble	Loc-3	0.225	0.439
Loc-3	1D-CNN	Loc-1	0.179	0.398
	TCN	Loc-1	0.199	0.421
	LSTM	Loc-1	0.185	0.406
	GRU	Loc-1	0.176	0.396
	Ensemble	Loc-1	0.138	0.351
	1D-CNN	Loc-2	0.176	0.397
	TCN	Loc-2	0.199	0.427
	LSTM	Loc-2	0.184	0.409
	GRU	Loc-2	0.175	0.397
	Ensemble	Loc-2	0.135	0.347

The generalization was moderate for models trained in Loc-2 then tested in Loc-1. Whereas for the GRU model, the MAE was the lowest, amounting to (0.126), and for Ensemble and LSTM (0.129). At the same time, when these models were tested on Loc-3, all of them have shown a rather high error rate with MAEs around (0.4). The 1D-CNN model slightly outperformed others with an MAE equal to (0.404), but these results show a rather limited generalization capability from Loc-2 to Loc-3. Lastly, the models that were trained on Loc-3 and tested on Loc-1 and Loc-2 exhibited relatively poor performance on both test locations. Testing on Loc-1 yielded the best performance for the Ensemble model, with an MAE of (0.351); similarly, the Ensemble model relatively fared better during the conduct of the test on Loc-2, recording an MAE of (0.347).

These results on the generalization performance of the model to other locations are dictated by data patterns specific to Loc-3 alone, showing some unique attributes of Loc-3 that most likely have not been represented during training on Loc-1 or Loc-2.

6.1.2.3.4. Loss and MAE Convergence Analysis

The loss and MAE convergence curves in Figure 6-11 further present perspectives into the effectiveness of the training and validation at different locations. The 1D-CNN has a fast convergence of loss and MAE across all locations within the first few epochs; this generally means that their learning processes are effectively done, with a low likelihood of overfitting. This is further confirmed by the model’s metrics, which showed fast convergence and a very good performance in Loc-1 and Loc-2.

The Ensemble model also demonstrates good balance between convergence and error rate, especially for Loc-1 and Loc-2, though they were noisier on both validation loss and MAE, especially for Loc-3. That would mean that the models can do well in capturing the temporal patterns; still, additional tuning might be required to generalize well for more complicated data patterns, as in Loc-3. On the other hand, LSTM and GRU converge slower than other models and are characterized by higher losses and MAEs on validation, especially for Loc-3.

Overall, the 1D-CNN and Ensemble models showed very good curve converge and robustness for most scenarios, including achieving the lowest values for MSE and

MAE in individual location runs. In contrast, the performance of the LSTM and GRU models were somewhat comparable but were less effective in the case of Loc-3. Cross-location results in Table 6-10 clearly show the limit of model generalizability, especially in cases where the model learned from Loc-1 or Loc-2 was transferred to Loc-3. This explains the fact that Loc-3 has something different in terms of the pattern of data; hence, this location requires further personalized techniques in order to be generalized well, such as transfer learning or adding extra contextual features.

6.2. Discussion

6.2.1. Comparative Model Performance: Location-Specific vs. Combined Data

The results obtained in both scenarios in Stage 1 clearly indicated great deviations between the model performances resulting from either location-specific training or combined dataset training. For instance, location-specific training generally resulted in higher accuracy associated with lower loss, which may mean an improved learning of the localized pattern and the unique occupancy trend especially in Loc-1 and 2. This result aligns with the work of Lucas *et* al. (2019), who noted that local dynamics are an important part of any EVCS occupancy forecast. However, this was not enough to say that location specific improve the model's performance better than the combined data in Scen-1. In contrast, as stated in the comparative results in this research, the models which were trained on the combined dataset demonstrated better generalization performance across different locations and were a little worse at capturing local specifics.

This result can be clearly seen when comparing visually the results in Figure 6-3 and Figure 6-4, where the confusion matrix for all models in the Scen-1 shows a clear view about how models, especially TCN, were able to be generalize well over different locations to predict different classes. Among the models compared, within both stages, while most of the models performed reasonably well, training on combined dataset in Scen-1 enhance more the models' generalization. On the other hand, training on a location specific dataset, although it helps models to reduce training error by capture more local pattern, models appear to be less generalized to other locations. As the loss rate varies when training models on combined dataset and location-specific dataset, thus model accuracy and the ability to generalizations impacted by aggregating different dataset.

6.2.2. Impact of Task Framing: Classification vs. Regression Performance

Furthermore, when this study investigated framing the prediction task as either classification or regression, the best performance was achieved by the models for classification, which were tasked with predicting the categorical occupancy state of a space in this research categorized into four categories (Empty, Moderate Low Occupancy, Moderate High Occupancy, or Fully Occupied) relative to models for regression. This affirms the findings of Hecht *et* al. (2021), who demonstrated that binary predictions perform best concerning structured EV charging tasks. It can be noticed that the strong classification performance of both TCN and then the Ensemble model was achieved for usually highly variable intermediate occupancy states, which are most important to optimize user satisfaction but also resource efficient EVCI.

On the other hand, these regression tasks posed additional challenges, since they are directed toward the prediction of total occupancy values, which are naturally more variable. Variation across space and time, or seldom high-demand events, brought in complexities that the regression models cannot handle with full accuracy. Furthermore, the training features that were available for this research may not provide the aqueduct direction for the model to learn the underlying user charging behaviour. This is consistent with Chen *et al*. (2022), who used classification predictions as input features for regression models, thereby enabling the latter in a hybrid fashion to take into consideration the categorical context while making continuous predictions. Such an integration, as suggested by Nespoli *et al*. (2023), enhance accuracy, particularly in the context of fluctuating conditions. Thus, predicting occupancy as classes may be more suitable than predicting the exact total occupancy, given the high variation and weak patterns that make precise prediction more challenging. As a result, the classification model appears to be a more appropriate choice for this research.

6.2.3. Addressing Temporal Continuity and Cross-Location Split Techniques

When training models on combined datasets, ensuring reasonable coverage throughout locations is vital for accurate forecasting. Because of the temporal constraint in the data, this work followed a cross-location split technique in which data of each location were sequentially structured. This was useful split technique that would avoid model overfitting, where having the split prevent data from a certain location to exist in a training set while not represented in the testing set, therefore reducing training without bias and provide a fair split between locations. This follows the temporal continuity that does not allow for random shuffling and hence conveys affirms the concept explained by Morton *et al*. (2017), concerning the importance of preserving time-series order in the cross-location splits.

This approach mitigates the bias that can arise from sequentially structured data, whereby models might train to learn patterns at one or two locations, and fail to generalize for the third one. In the above example, due to the sequence structure of input, the model was not able to learn the occupancy trend specific to Loc-3, which implies that the model needs equal exposure to all types of locations while training. The challenge in data representation as such, taking into consideration the fact that a model actually requires consistent exposure to location-specific patterns to be expected to perform normally over large geographical areas, is that those diverse regions develop under different occupancy trends.

6.2.4. Average Occupancy Feature Improves Prediction Accuracy

The integration of average occupancy as a feature was important in improving model performance, especially in refining predictions across varying occupancy levels. Local average occupancy gives the general context in which a model can find the trends related to normal usage, thereby reducing errors toward the high and low occupancy conditions. This feature turned out to be useful for capturing occupancy trends that are cyclic, something also pointed out by Nespoli *et al*. (2023), who, by including average occupancy, reached better forecast accuracy, especially for high-demand periods. In this work, average occupancy metrics allowed the models to learn the deviations from normal patterns, hence improving their responsiveness to time-varying demands at varied time scales. Due to this normal occupancy benchmark, the metric served to enhance model stability over various regions and times, developing an adaptable system to forecast changes in occupancy based on historical trends.

6.2.5. Model Selection Criteria: Evaluation and Interpretability

Models' selections were informed by several metrics: *accuracy*, *precision*, *recall*, *F1-score, classification report* and *confusion matrix* for classification tasks, and *MSE and MAE* for regression.

In both Scen-1 and Scen-2 of Phase 1, the TCN model ranked highest in almost all metrics, further strengthening that the TCN model is able to handle challenging temporal patterns, as already demonstrated by Dong *et al*. (2021). Moreover, TCN's strength lies in the handling of temporal data across heterogeneous contexts, it is expected to be strong for combined datasets, as indicated by previous studies such as those by. It follows closely with the Ensemble model (CNN + TCN) which did reasonably well in understanding occupancy variation with subtlety for diverse data. These are industry standard evaluation criteria adopted in the present work on EVCI, and, as indicated by Viswanathan *et al*. (2018), precision and recall are both very important for an accurate prediction of the availability of EVCSs.

In Phase 2, 1D-CNN and the Ensemble model (LSTM + TCN) generally show a higher performance level over the other models. While LSTM and GRU shows also comparative result specially in the Scen-1. Both MSE and MAE were used to evaluate models' performance in this stage. The MSE calculates the average of the squared differences between the predicted and real values. Because it squares the differences, the resulting errors coming from larger differences get much heavier penalties. Therefore, MSE is useful when the larger errors should be as small as possible since large errors are magnified. However, this could sometimes be sensitive enough for outliers, and skew the evaluation for those datasets that contain just a few very extreme values. On the other hand, MAE calculates the unconditional average of the absolute errors. This interpretation provides the direct accuracy of the model and thus estimates the typical magnitude of an error without indicating larger deviations. Because of this, MAE is robust when there are outliers since it deals with all errors linearly, hence giving a better view into how the model will perform across the range of values. Therefore, while the comparison in this study relies more on MAE, it also considers a balanced comparison between both metrics.

Besides, techniques for explainability, such as observing feature importance with average occupancy metrics, enhance the model's transparency to understand which variables lead to the variation of prediction values. Results from such interpretability will help decision-makers to plan EVI, probably enabling them to emphasize only those highly impactful features of occupancy. The features used here are interpretable, as guided by Wang *et al*. (2021); these will contribute to increasing user trust and operational applicability so that predictions are more actionable for infrastructure management.

6.2.6. Implications for Future EVI and Scalability of Models

The conclusions from the presented research pose practical implications for scalability and implementation within management using predictive models for EVI. In other words, this study stipulates and proves that location-specific training generally enhances prediction accuracy, while combined datasets offer better generalization; consequently, we advocate for a hybrid approach, whereby models are first trained on combined data, and are then fine-tuned with location-specific data as more become available. This hybrid training may bring in better model adaptability, a strategy supported by Yang *et al*. (2024) to cope with variations in demand, both geographically and temporally. This could be extremely useful for real-world deployment in conjunction with features like averaging occupancy, given that charging demand is very different across regions using cross-location split techniques. Furthermore, such an approach can enable more sophisticated attention mechanisms that are self-aligned dynamically in time with shifts in traffic, focal demands, and other contextual demands in the future. Real-time information streams, as suggested by Chen et al. (2022), can enhance model responsiveness toward proactive management of EV-EVCI in matching user demands to emergent and fluctuating occupancy patterns.

6.3. Chapter summary

This chapter reviewed predictive models in terms of occupancy at EVCSs, considering the influence of *location-specific information* versus *aggregated* data, the nature of the different predictions of *classification* or *regression* methodologies, different works to keep temporal continuity, the relevance of average occupancy attributes, and model selection criteria. This study reveals deviations in model performance when trained on location-specific data compared to combined datasets. Overall, most models that were trained on location-specific data had higher accuracy with lower loss, which hinted at better learning of localized pattern and unique occupancy trends at respective locations.

However, the location-specific training did not conclusively outperform that of combined data. The best performance of generalization across various locations but slightly less capable of depicting the local specifics correctly was observed for the models trained on integrated datasets. Visual inspection of confusion matrices, especially for the TCN, showed that training on combined data strengthened the generalization capabilities. While most models performed reasonably well, training on combined datasets in Scen-1 improved model generalization. On the other hand, location-specific training captured the local pattern, reducing the training error but losing the ability to generalize well. This again depicts how aggregation of various datasets influences both accuracy and generalization ability.

Model selection was guided by several metrics, according to the task, such as accuracy, precision, recall, the F1-score, classification reports, confusion matrices for the classification tasks, and MSE and MAE for regression tasks. Among most of the metrics in both the scenarios of Stage 1, the ranking was generally one for the TCN model, hence proving its ability to handle much of the complicated temporal patterns. The Ensemble model (combining 1D-CNN and TCN) also performed well in capturing subtle occupancy variations in diverse data. During regression tasks in Stage 2, the performances of both 1D-CNN and the Ensemble model combining LSTM with TCN were relatively better. The LSTM and GRU models achieved similar performances, with better performance captured in Scen-1.

Chapter 7: BiGTCN Classification Model

In this chapter the BiGTCN represents an advanced deep learning architecture that has been designed with the objective of improvement in the modelling of temporal occupancy patterns of an EV Charging Station. First, this chapter gives a brief outline on the model; it outlines its conceptual framework based on the diagrammatic representation and highlights some of the basic definitions. It then goes on to outline how the data preparation was done for the feeding of the model, discusses the unique architectural pieces of the BiGTCN, and how it was trained. Subsequently, after describing the experimental environment it concludes with the results of experiments conducted to demonstrate the effectiveness of the model.

7.1. Overview of the BiGTCN Model

The proposed BiGTCN architecture represents an enhancement to the TCN by integrating a Bidirectional Gated Recurrent Unit (BiGRU) layer. This hybrid structure is designed to address the specific challenges associated with predicting EV charging station occupancy, where capturing both short-term and long-term temporal dependencies is crucial for accurate predictions. While TCNs are known for their efficiency in modelling temporal patterns using dilated convolutions, their unidirectional nature may limit their ability to fully capture dependencies from past and future time steps. To mitigate this limitation, the BiGTCN incorporates a BiGRU layer, which learns bidirectional temporal dependencies, further enhancing the model's capability to process sequential data effectively.

Traditional deep learning architectures such as LSTM, GRU, and standalone TCNs have demonstrated success in time-series forecasting tasks, but each has limitations when applied to complex, highly dynamic datasets like EVCS occupancy data. Specifically:

- Unidirectional Limitations of TCNs: While TCNs excel at efficiently modelling temporal patterns using convolutional filters, they process data in a single direction (past-to-present), potentially missing contextual information from future time steps.

- Sequential Learning Gaps: Sequential models such as LSTMs and GRUs, although bidirectional in some cases, often struggle with capturing long-term dependencies without substantial computational overhead.

The BiGTCN is specifically designed to overcome these limitations by combining the strengths of TCNs (efficient temporal convolution) with the bidirectional learning capability of BiGRUs. The BiGTCN architecture is inspired by the Parallel Spatio-Temporal Attention-TCN (PSTA-TCN) structure proposed by Fan et al. (2023). However, unlike PSTA-TCN, which focuses on integrating spatiotemporal attention mechanisms, BiGTCN emphasizes bidirectional temporal learning to address the specific challenges of EVCS occupancy prediction. To the best of the researcher knowledge, no existing studies have applied this structure to such prediction tasks, making BiGTCN a unique contribution to this domain.

7.1.1. Experimental Development and Rationale Behind the BiGTCN Architecture

Following the comparative evaluation presented in Chapter 6, where five deep learning models (1D-CNN, TCN, LSTM, GRU, and a TCN-LSTM ensemble) were assessed, the research proceeded into a second stage aimed at refining and enhancing the predictive performance of the best-performing models. This phase focused on designing hybrid architectures that could leverage the complementary strengths of different network types to better capture temporal patterns in EV charging station occupancy.

7.1.1.1. Rationale for Exploring Hybrid Architectures

The decision to develop a hybrid model was motivated by two primary observations: Comparative results in Stage 1 (Chapter 6) revealed that TCN and GRU outperformed the other baseline models in both classification and regression tasks. TCN showed strength in extracting temporal features using dilated causal convolutions, while GRU demonstrated effective learning of sequential dependencies with fewer parameters and faster convergence compared to LSTM.

Insights from the literature also supported this direction. Several studies (e.g., (Fan et al. 2023)) emphasized the advantages of TCN in modelling long-range dependencies, while recurrent architectures like GRU have been praised for their ability to manage dynamic sequences. The hybridization of CNNs with RNNs has shown benefits in domains like speech recognition, traffic prediction, and energy forecasting.

These findings prompted the researcher to begin designing combinations of these architectures to create a more expressive and flexible hybrid model for the occupancy forecasting task.

7.1.1.2. First Attempt: TCN + GRU Hybrid

The initial hybrid model combined a TCN layer with a GRU layer. The idea was to allow the TCN to first extract and compress temporal features from the input sequence, which were then passed to the GRU layer to capture sequential dependencies over time.

- Configuration:
    - TCN layer, 64 filters, kernel size 3, dilations [1, 2, 4, 8, 16], dropout 0.2
    - GRU: 1 layer, 100 units.
    - Optimizer: Adam.
    - Epochs: 50.
    - Batch size: 32.

Although this combination improved the model’s ability to detect transitions between occupancy categories, it had a key limitation: the GRU processed input only in the forward direction. As a result, it was unable to learn from future context within the sequence, creating a potential bottleneck in temporal understanding particularly in capturing symmetrical or context-dependent transitions.

7.1.1.3. Refinement: Introducing Bidirectional GRU

To address this limitation, the architecture was refined by replacing the unidirectional GRU with a BiGRU layer. This version, named BiGTCN, retained the same TCN component but enabled the recurrent unit to process sequences in both forward and backward directions, thereby capturing context from past and future time steps.

- Revised Configuration (BiGTCN):
    - TCN: Same as above
    - BiGRU: 1 layer, 64 units each (forward and backward), dropout 0.2
    - Loss: Categorical Cross-Entropy
    - Optimizer: Adam
    - Epochs: 60
    - Batch size: 32

7.1.2. Diagrammatic Representation

The BiGTCN model is designed as a dual-branch structure that combines the strengths of TCN and BiGRU. These branches work in parallel to extract both temporal and spatial dependencies, enabling the model to effectively learn dynamic relationships in EV charging station occupancy data. The architecture of the proposed BiGTCN model is depicted in Figure 7-1 and described below.

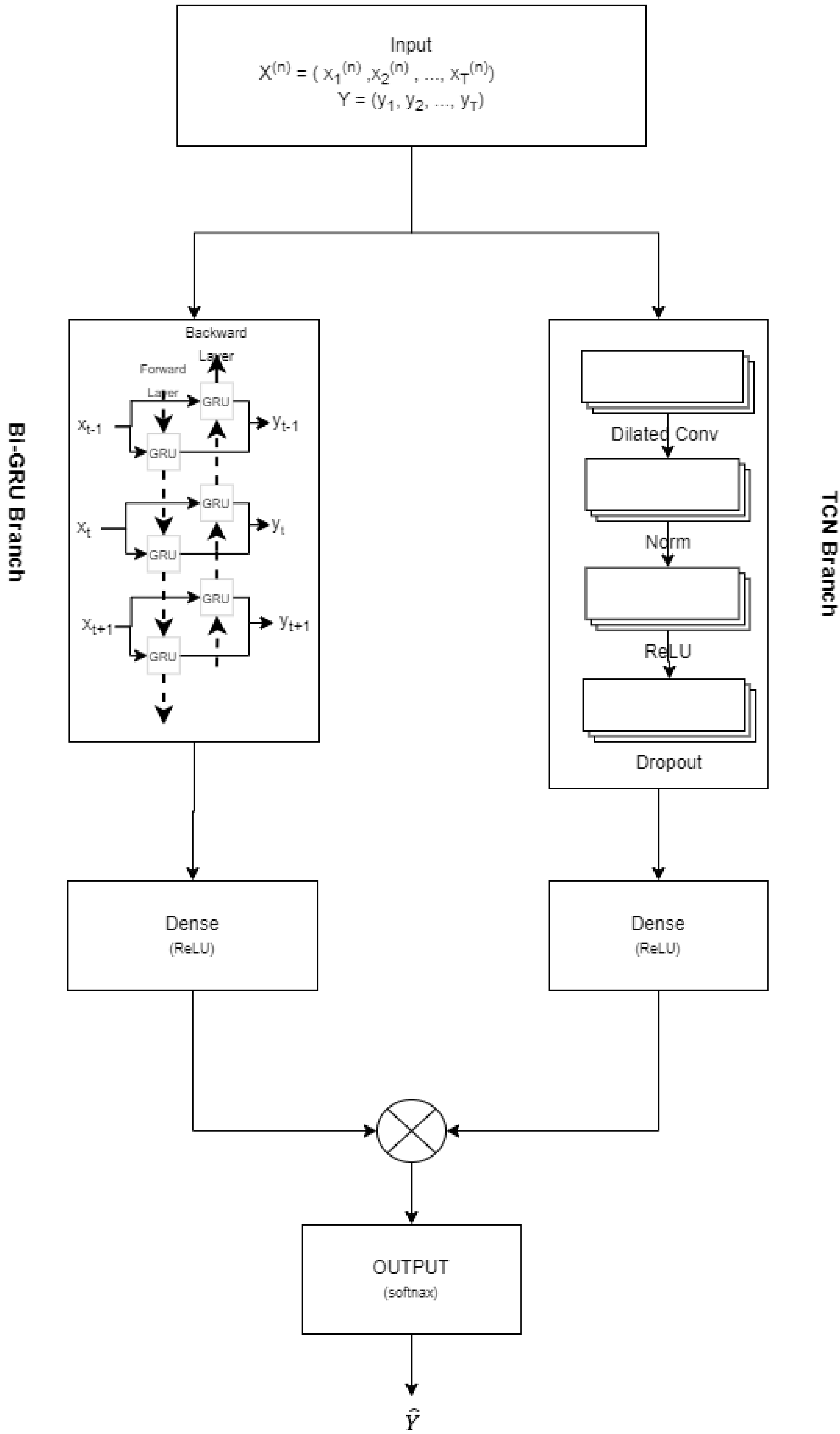


Figure 7-1 Proposed BiGTCN Structure

Input Layer: The input to the model consists of preprocessed and normalized time-series data, including temporal features (e.g., time of day, day of the week) and contextual information (e.g., weather conditions, occupancy history). This input layer has a shape corresponding to the time steps and feature dimensions of the data.

TCN Branch: The TCN branch captures temporal dependencies in the input data using a dilated convolutional layer with 64 filters and a kernel size of 3. Unlike traditional RNNs, the TCN efficiently processes long sequences through its parallel convolutional structure while preserving the causal nature of time-series data. The output of the TCN branch is passed through a Dense Layer with 32 units and a ReLU activation function to further refine the learned temporal features.

BiGRUBranch: The BiGRU branch processes the same input data bidirectionally using a Bidirectional GRU layer with 64 units. This layer captures both forward and backward dependencies, improving the model's ability to learn spatial and temporal relationships. The output of the BiGRU layer is passed through a Dense Layer with 32 units and a ReLU activation function, ensuring that spatial features are effectively extracted and represented.

Feature Fusion: The outputs from both branches (TCN and BiGRU) are concatenated using a Concatenate Layer. This combined feature vector integrates both temporal and bidirectional spatial information, providing a comprehensive representation of the input data.

Output Layer: A fully connected Dense Layer with a *softmax* activation function produces the final output for classification tasks, such as predicting the occupancy category of EV charging stations.

7.1.3. Model Definition

7.1.3.1. Input Data

The input to the model consists of multivariate time series data:

$$\mathrm{X} = \left[X^{(1)}, X^{(2)}, \dots, X^{(n)}\right], where\ X^{(i)} \in \mathbb{R}^{T}$$

Equation (7-1)

Where *n* is the number of features (or series), and *T* is the sequence length or window size.

7.1.3.2. TCN Block

The input data follow a parallel links to the TCN block and the bi-directional GRU block.

TCN Layer: The TCN processes the input X to capture temporal dependencies. A dilated causal convolution is applied:

$$H_{TCN} = ReLU(W_{TCN}\,X + b_{TCN})$$

Equation (7-2)

Where W_{TCN} and b_{TCN} are weights and biases of the convolution layer.

Dense Layer: A fully connected layer reduces the output dimensionality:

$$Z_{TCN} \;=\; ReLU\left(W_{Dense,TCN}\,H_{TCN} \;+\; b_{Dense,TCN}\right)$$

Equation (7-3)

Where $Z_{TCN} \;\in\; \mathbb{R}^{32}$

7.1.3.3. BiGRU Block

The BiGRU block processes the input X to extract the spatial dependencies:

Forward GRU:

$$h_t^f \;= GRU_f(X_t, h_{t-1}^f)$$

Equation (7-4)

Where h_t^f is the forward hidden state at time t.

Backward GRU:

$$h_t^b \;= GRU_b(X_t, h_{t+1}^b)$$

Equation (7-5)

Where h_t^b is the backward hidden state at time t.

Bidirectional Output: At this layer both forward and backward states combine to provide the layer output:

$$h_t = \left[h_t^f, h_t^b\right]$$

Equation (7-6)

Dense Layer: This a fully connected layer to reduce the output dimensionality:

$$Z_{BiGRU} = ReLU\left(W_{Dense,BiGRU}\, h_T + b_{Dense,BiGRU}\right)$$

Equation (7-7)

Where $Z_{BiGRU} \in \mathbb{R}^{32}$, and h_T refers to the hidden state at the last time step *T*.

7.1.3.4. Combined Features

The outputs from both branches are concatenated to provide a concatenated result from both branches:

$$Z_{Combined} = Concatenate\,(Z_{TCN}, Z_{BiGRU})$$

Equation (7-8)

Where $Z_{Combined} \in \mathbb{R}^{64}$.

7.1.3.5. Output Layer

At the end, the combined features pass through a Softmax layer to predict the class probabilities:

$$\hat{Y} = Softmax\,(W_{out}\, Z_{combined} + b_{out})$$

Equation (7-9)

Where $\hat{Y} \in \mathbb{R}^{num\_classes}$ represents the predicted probabilities for each class. BiGTCN Experiments and Results

This section presents the experimental setting that has been used for testing the proposed BiGTCN model for EVCS occupation status forecasting. The dataset and the approach to be adopted are the same as in Chapter 6, which confers the uniformity and comparability of this analysis. Along with the BiGTCN, baseline models were used, including the TCN and 1D-CNN, both of which demonstrably attained good performance (as presented in the previous chapter). Also, to enhance the comparison, researcher additionally added a lightweight version of the proposed model called GTCN, which combines TCN and GRU layers.

7.1.4. Data Preparation

The following dataset includes data compiled from three locations (in the same manner as in Chapter 6). It includes temporal variables, occupancy classification, and weather conditions. Major steps in data pre-processing include the following:

Target Variable Encoding: The column in occupancy category representation is called "OccCate", and it was label-encoded into numerical form using *LabelEncoder* from *Scikit-learn*; this was further one-hot encoded using the function *to_categorical* using *Keras* library afterward for multi-class classification.

Data Splitting: The original dataset was divided into training, validation, and test sets. The testing set is a small portion of the data reserved for testing the final performance, specifically (15%); the remaining (85%) of the dataset comprised training (60%) and validation (25%) subsets. This ensures the division in every subset is representative of the overall distribution of data.

Scaling Features: The scaling of input features was applied using the *StandardScaler* in *Scikit-learn*, for the normalization of data. This scaler was fitted on the training data and applied on both the validation and test set to avoid leakage.

7.1.5. Model Architecture

Three models were used in this experiment, for comparison with the proposed BiGTCN model: 1D-CNN, TCN, and GTCN. These models were structured to investigate different methods of modelling the temporal and sequential patterns that exist within data; similarly, the BiGTCN model will leverage both past and future information through the respective bidirectional element. The shows the summary of the models' architecture:

Table 7-1 Comparison of Layers in Tested Models

Layer/ Component	**1D-CNN**	**TCN**	**GTCN**	**BiTCN**
Input Layer	Input (shape=(number_of_features, 1))			
Hidden Layers	Conv1D(filters=32, kernel_size=3, activation='ReLU') MaxPooling1D(pool_size=2)	TCN (nb_filters=32, kernel_size=3, return_sequences=False)		
Dense Layers	Dense (units=64, activation='relu')	Dense (units=64, activation='relu')	Dense (units=32, activation='relu') (after GRU branch)	Dense (units=32, activation='relu') (after GRU branch)
Concatenation Layers			Concatenate outputs of TCN and GRU branches	Concatenate outputs of TCN and BiGRU branches
Output Layers	Dense (units=num_classes, activation='softmax')			

7.1.6. Training Procedure

All models experienced training and evaluation in a uniform set of conditions to guarantee an equitable comparison. The Adam optimizer was selected since it is efficient while supporting adaptive learning rates. The Loss Function Categorical "cross-entropy" was the most appropriate for this multi-class problem. Models' performance and comparison were assessed using classification metrics (e.g., accuracy, F1-score, and confusion matrix). Each model was trained for (25) epochs, balancing between sufficient learning time and computational efficiency. A batch size of (20) was used to keep updates of gradients constant.

The *ModelCheckpoint callback* from Keras library was used to save the weights from models that have the maximum validation accuracy recorded during training. This helped avoid the problems of overfitting and keep the best performance models in making predictions with new unseen data. Models went through training utilizing the standardized training dataset (X_train, y_train), while the performance on the validation set (X_val, y_val) was systematically observed. The training history, encompassing metrics related to loss and accuracy for both the training and validation datasets, was documented for subsequent analysis.

7.1.7. Experimental Environment

The experiments were conducted using the following tools and libraries:

Programming Language: Python

Data Handling: Pandas for data manipulation and pre-processing.

Numerical Computations: NumPy for efficient numerical operations.

Machine Learning Utilities: Scikit-learn for data splitting, encoding, and scaling.

DL Framework: TensorFlow and Keras for building and training NNs.

TCN Implementation: The "*TCN*" package provided the TCN layer implementation compatible with Keras.

Visualization: Matplotlib for creating plots of the training history.

Care was taken that all the experimental settings were prepared with due diligence, including to standardize the training parameters and enable model evaluations to focus on architectural differences. The objective of this research is to show how much the combination of bi-directional GRUs with TCNs capture complex temporal dependencies relevant for efficient occupancy prediction using the evaluation of performance by the BiGTCN model against recognized baseline models and a reduced hybrid model, GTCN. The results from these experiments will be discussed closely in the sections that follow.

7.1.8. Experimental Results

7.1.8.1. Models' Training Accuracy and Loss

Figure 7-2 shows the accuracy and loss of each model, monitored during the training phase across the (25) epochs. It can be noticed that GTCN and BiGTCN achieve faster convergence, getting higher final training accuracy than the two simpler models, namely a 1D-CNN, and a TCN.

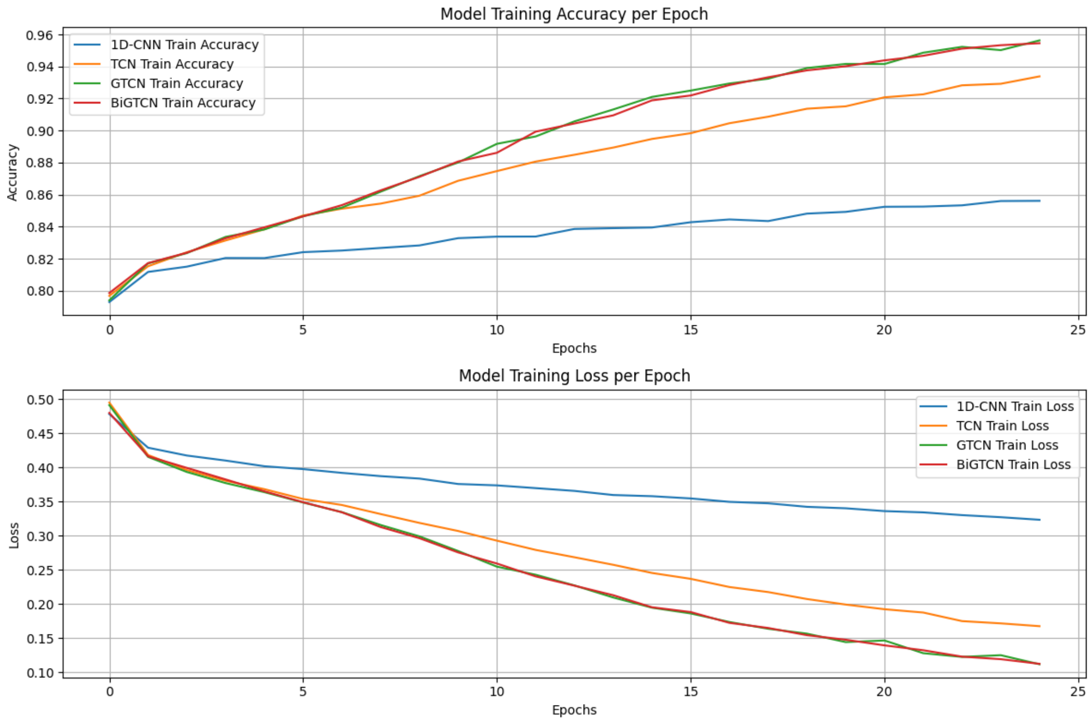


Figure 7-2 Models Training Accuracy and Loss per Epochs

The 1D-CNN presented slower improvement, and never actually beat the threshold of accuracy at (0.86). It also kept the maximum training loss compared to other models, showing a weaker ability to learn from complex features in this data. On the other hand, the TCN, by capturing time-series features with the concept of temporal convolution, outperformed the 1D-CNN with a training accuracy of (0.93).

The GTCN model, integrating the advantages of TCN with GRUs, demonstrated a significant enhancement in training accuracy, achieving (0.96). This improvement is due to its capacity to utilise both temporal and sequential dependencies in the data. The suggested BiGTCN attained a comparable training accuracy of (0.96), leveraging its BiGRU component to capture contextual dependencies in both forward and backward directions. Both GTCN and BiGTCN exhibited a steady reduction in loss across the epochs, ultimately diminishing to approximately (0.12) for each model. Considering the nearly same training performance of GTCN and BiGTCN, further evaluation utilising supplementary metrics is essential to ascertain which model exhibits superior overall performance.

7.1.8.2. Models' Testing Accuracy and Loss

Comparison of test accuracy and loss highlighted the enhanced performance by the proposed architecture, BiGTCN. As can be seen from Figure 7-3 the 1D-CNN produced the minimum test accuracy of (0.83) with the maximum loss value of (0.38) and thus was found incapable when dealing with complex temporal fluctuations in smart EVCS occupancy data.

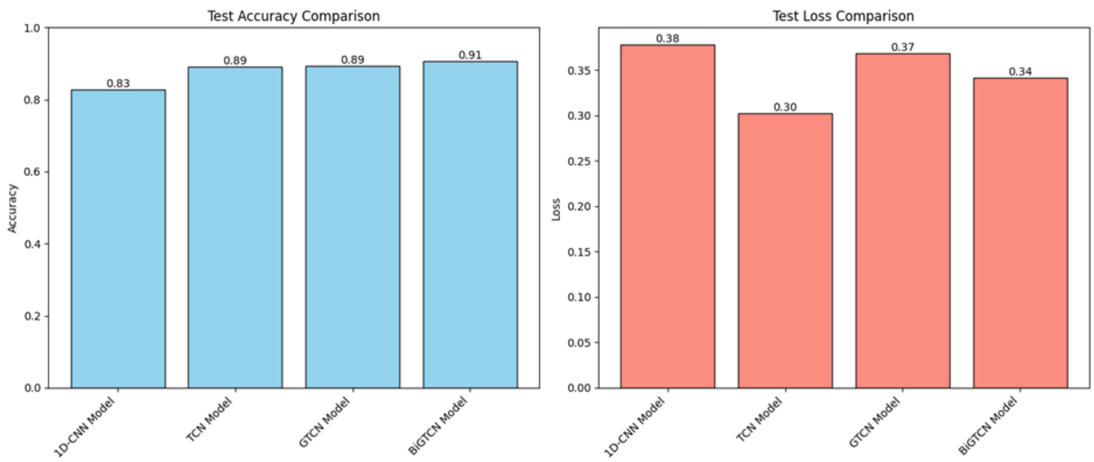


Figure 7-3 Test Accuracy and Loss Comparison

The TCN performed much better, reaching a test accuracy of (0.89) while reducing the test loss to (0.30) which is the lowest loss obtained between the four models' loss. TCN shows a competitive result with the proposed structure during the test on unseen data.

The GTCN achieved a test accuracy of (0.89), equivalent to the TCN, but demonstrated a bigger test loss of (0.37). This suggests that although GTCN achieved comparable accuracy, it was less effective than TCN in reducing loss on the test data. Finally, the BiGTCN model, which outperforms all models, reached (0.91) in accuracy and (0.34) in loss. This proves that integrating BiGRU with TCN is highly efficient in modelling and prediction of EVCS occupancy.

7.1.8.3. Confusion Matrices

The confusion matrices, depicted in Figure 7-4 offer detailed insights into the prediction performance for each occupancy class. The 1D-CNN struggled particularly with classes 1 and 3, showing significant misclassifications. For example, only (55%) of class 1 samples were correctly identified, with (23%) misclassified as class 2.

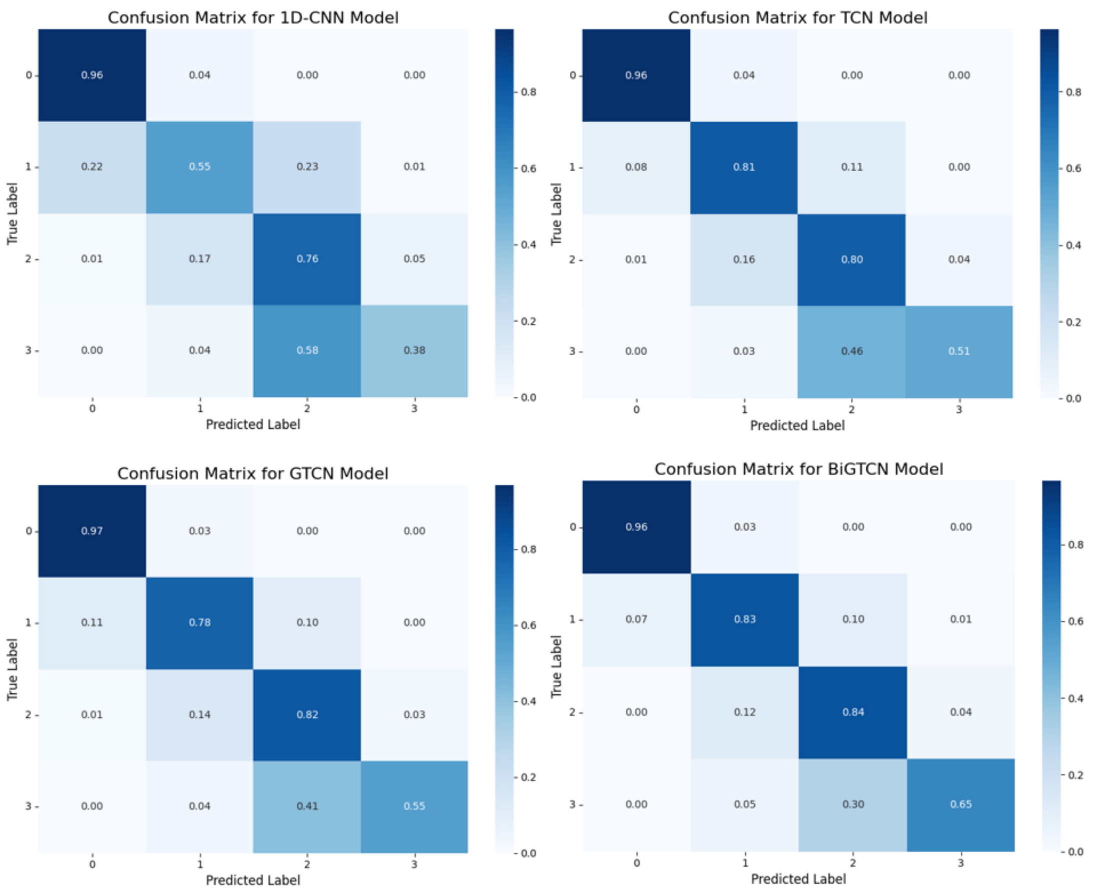


Figure 7-4 Confusion Matrices for All Models

The TCN showed improvements across all classes but still faced challenges with class 3 predictions, misclassifying (46%) of class 3 samples as class 2. The GTCN further improved classification accuracy for all classes by combining TCN with GRU, correctly classifying (78%) of class 1 samples and reducing misclassifications. The BiGTCN achieved the best performance, with notable improvements in distinguishing between classes 2 and 3. Specifically, for class 3, it achieved (65%) correct predictions, significantly reducing the confusion observed in other models. These results demonstrate the BiGTCN's enhanced capability in accurately classifying challenging occupancy states.

7.1.8.4. ROC Curves and AUC Scores

ROC and AUC scores for all models, shown in Table 7-2 and Figure 7-5 further corroborate the superior performance of the BiGTCN. The 1D-CNN demonstrated strong performance for class 0 (AUC = 0.99) but fell short for other classes, especially class 1 (AUC = 0.93).

Table 7-2 The ROC Curves and Area Under the Curve Scores

Class	1D-CNN	TCN	GTCN	BiGTCN
0	0.99	0.99	0.99	0.99
1	0.93	0.96	0.96	0.97
2	0.95	0.97	0.97	0.97
3	0.96	0.98	0.98	0.98

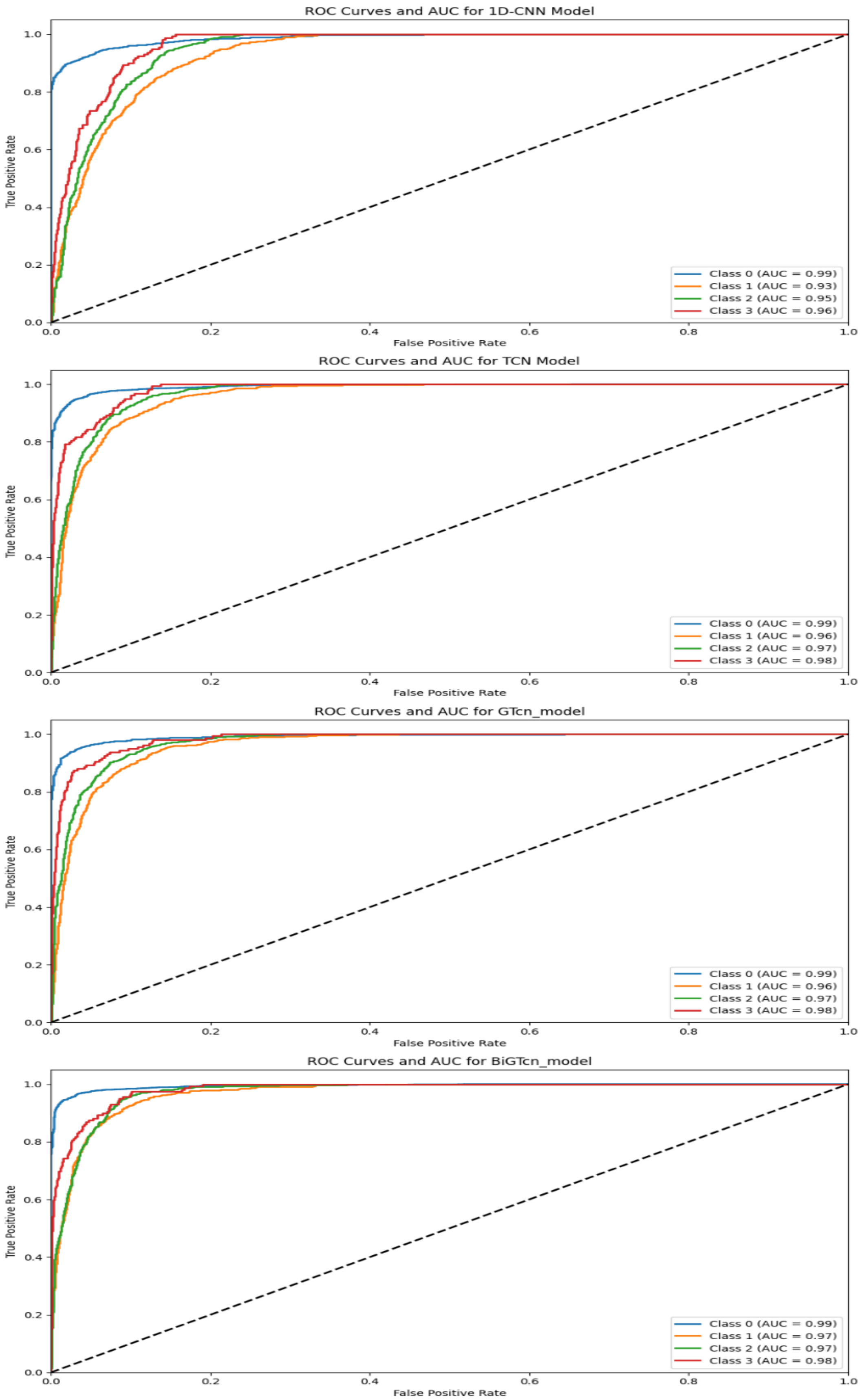


Figure 7-5 The ROC Curves and Area Under the Curve Scores; See Appendix C

The TCN improved the AUC scores across all classes, achieving (0.97) for class 2 and (0.98) for class 3. The GTCN slightly enhanced the AUC scores compared to TCN, with an AUC of (0.97) for class 2 and (0.98) for class 3. The BiGTCN achieved the highest AUC scores for all classes, with (0.99) for class 0, (0.97) for class 1, and (0.98) for both classes 2 and 3. These results underscore its superior ability to differentiate between occupancy states with high confidence.

7.1.8.5. Classification Report

The classification reports are given in Table 7-3. Precisely, the overall performance of each model is summarized. With class 1, the 1D-CNN has a minimum precision of (0.72) and recall of (0.55), while class 3 has a minimum of (0.65) for precision and (0.38) for recall, which means it failed to effectively classify these classes.

Table 7-3 Classification Report (Precision, Recall, F1-Score)

Model	Metrics	0	1	2	3	Macro avg	Weighted avg	Accuracy
1D-CNN	Precision	0.93	0.72	0.56	0.65	0.72	0.83	
	Recall	0.96	0.55	0.76	0.38	0.66	0.83	
	F1-score	0.94	0.62	0.65	0.48	0.67	0.82	0.83
	support	2457	831	482	159	3929	3929	3929
TCN	Precision	0.97	0.80	0.70	0.82	0.82	0.89	
	Recall	0.96	0.81	0.80	0.51	0.77	0.89	
	F1-score	0.97	0.80	0.74	0.63	0.78	0.89	0.89
	support	2457	831	482	159	3929	3929	3929
GTCN	Precision	0.96	0.82	0.72	0.83	0.83	0.89	
	Recall	0.97	0.78	0.82	0.55	0.78	0.89	
	F1-score	0.96	0.80	0.76	0.66	0.80	0.89	0.89
	support	2457	831	482	159	3929	3929	3929
BiGTCN	Precision	0.98	0.82	0.75	0.81	0.84	0.91	
	Recall	0.96	0.83	0.84	0.65	0.82	0.91	
	F1-score	0.97	0.82	0.79	0.72	0.83	0.91	0.91
	support	2457	831	482	159	3929	3929	3929

The classification reports are given in Table 7-3, summarizing the performance of each model across all occupancy classes. Notably, the dataset is imbalanced, class 0 contains 2457 samples, while class 3 has only 159. In this context, reporting only accuracy may be misleading, as a model could achieve high accuracy by predicting majority classes more frequently. To address this, both macro and weighted averages of the evaluation metrics (precision, recall, and F1-score) are reported:

- **Macro average** calculates the metric independently for each class and then takes the average, treating all classes equally. This is especially useful for understanding model performance on underrepresented classes.

- **Weighted average** computes the average by taking the class imbalance into account, weighting each metric by the number of instances in each class.

This distinction ensures a fair and informative evaluation, particularly for challenging classes like class 3. For example, BiGTCN achieved the highest macro F1-score (0.83) and a precision of (0.81) with a recall of (0.65) for class 3, demonstrating improved ability to classify rare yet important states.

It can be seen that the TCN elevated the precision and recall for all classes over (0.80). Relatively, the GTCN further enhanced the precision and recall for classes 1 and 3 to (0.82) in precision, with (0.78) recall for class 1. In general, the highest performances concerning precision, recall, and F1-scores have been obtained by the BiGTCN. Class 3 reached (0.81) in precision and recall of (0.65) for classification of the most difficult classes with a relatively high score.

Overview, among all the runs, the proposed model of BiGTCN has always outperformed the baseline models, 1D-CNN and TCN, and the hybrid model GTCN, for every evaluation metric. The BiGRU component can capture contextual dependencies much better, yielding higher accuracy, better AUC scores, and improved precision and recall, especially for the challenging classes of occupancy. These studies clearly prove that BiGTCN is the best among the structures tested to predict the occupation of an EVCS.

7.1.8.6. Model Performance on Test Data Visualisation

The *actual* versus *predicted* classification plot shown in

Figure 7-6 was developed from the test data set. It shows relative efficiencies in predicting correct classes of EVCS occupancies, with true class labels plotted in blue, whereas orange dashed lines correspond to the predicted class labels over 300 time steps in graphs that follow.

The first graph illustrates deviations of the actual versus predicted values, especially in transitions of occupancy states, for the 1D-CNN model. It struggled to align with each actual class change, most especially for classes 1 and 3; this resulted in misclassifications being very frequent. This indicates that while 1D-CNN captures the temporal pattern to a certain extent, its limitations in modelling complex sequential dependencies significantly restrict its predictive performance.

Notably, the TCN model represented in the second graph outperforms the 1D-CNN, because actual and predicted class labels show much less discrepancy and with more scattered cases of misclassifications. The model has some issues with quick class transitions sometimes, especially around class 3 predictions. This ascertains how great TCN is in temporal pattern recognition, while at the same time it shows weaknesses regarding distinguishing subtle variations occurring in class transitions.

The third graph represents the GTCN model, which already shows distinctly better predictive performance compared to the TCN model. As much as the temporal pattern recognition of TCN is put into use, GTCN bounds the transitions between classes more precisely, due to the sequential modelling efficiencies of GRU. There is a significant decrease in misclassifications, especially for class 1 and class 3, and most of the model predictions have become much more stable.

Among these four graphs, the best correspondence between real classes and their predicted classes is provided by the BiGTCN model. Since a BiGRU leverages contextual information from both past and future, the BiGTCN reduces its prediction error. The transitions among the occupancy states are mapped in a much smoother way, and the model was showing an increase in precision and stability to deal with the challenging categories such as class 3. Visual consistency in the predictions made by the BiGTCN further underlines its superior generalization and better prediction over test data.

On the whole, these visualizations demonstrate the significant advantage of the proposed BiGTCN model over the baseline, 1D-CNN and TCN, and the hybrid model GTCN. Although the TCN and GTCN showed considerable improvements from the much simpler 1D-CNN, the proposed BiGTCN, with a bidirectional architecture, proved better and more reliable in terms of alignment with the actual class labels, hence being the best structure to carry out EVCS occupancy predictions. This conclusion can be investigated by comparing the prediction obtained from each model with the actual value, which will be obtained in the next part.

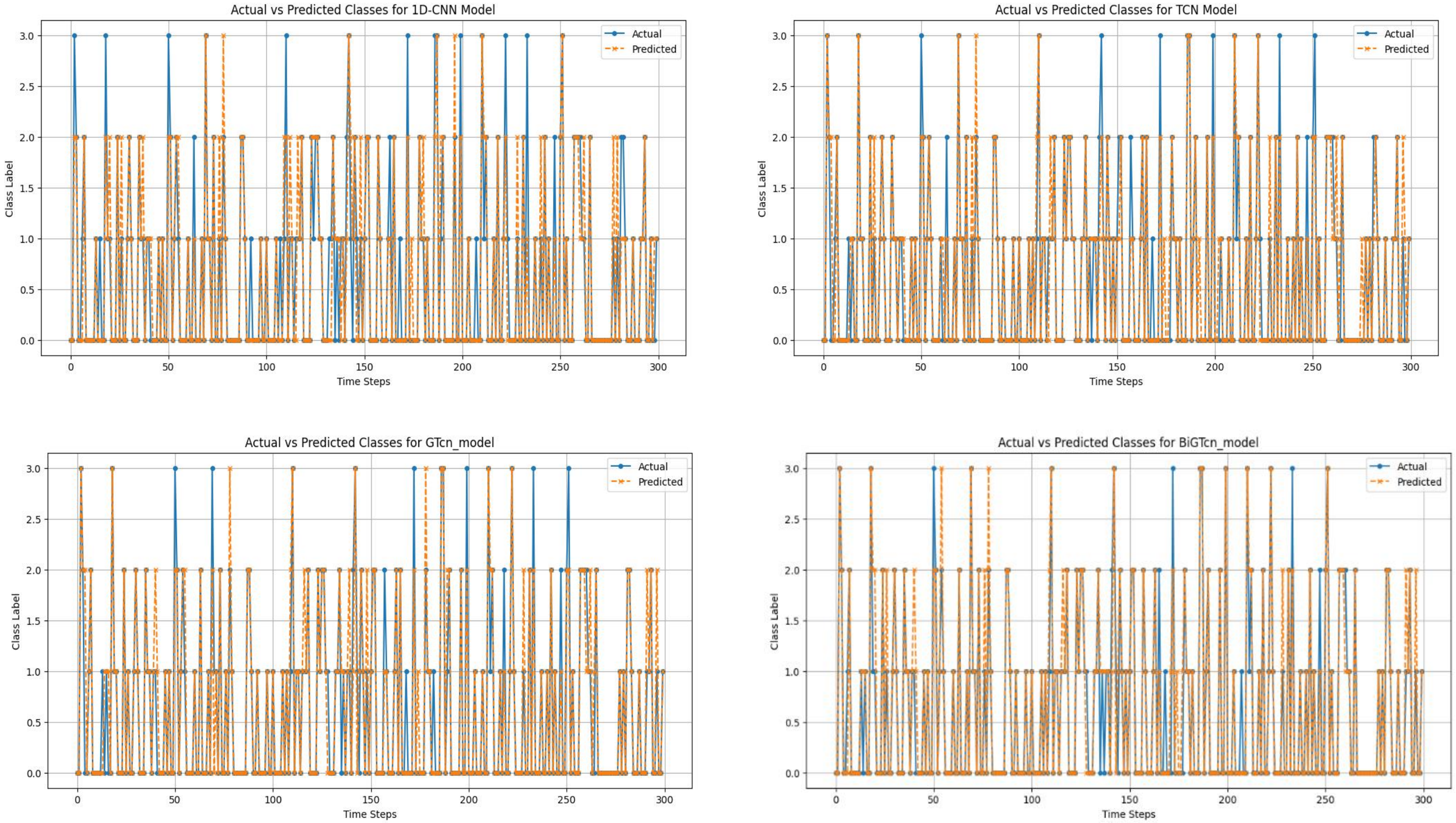


Figure 7-6 Models Performance on Test Data

7.1.8.7. Paired T-Test to Compare Model Predictions with Actual Values

To determine the quantitative and statistical accuracy of the models' predictions, a paired sample t-test was performed for the mean actual values of each model in comparison to predicted values: 1D-CNN, TCN, GTCN, and BiGTCN. This enables the analysis of whether there is a statistically significant difference between the actual occupancy state and the predictions according to the various models. Besides, paired samples' correlations were realized to find out the magnitude of the relationship between the observed and the predicted values, while effect sizes (Cohen's d) were calculated to measure the magnitude of the differences.

The results from the paired sample t-testing shown in the tables below show that there were somewhat large differences between the actual and predicted for several of the models tested. In the 1D-CNN model, the p-value was (0.595), showing no significant difference, thus indicating no significant difference from real values. However, the high SD of 0.420 with a comparatively lower paired correlation of r = 0.875 reflected inconsistency in the predictive performance for the model. The results are also in agreement with performance reflected in the real vs. predicted graph above, where 1D-CNN frequently misclassified and performed poor in the transition between the occupancy states.

Table 7-4 Paired Sample Test Between Actual and Predicted Values.

		Paired Differences							Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Pair 1	Actual - Predicted1DCNN	-.004	.420	.007	-.017	.010	-.531	3928	.298	.595
Pair 2	Actual - PredictedTCN	-.018	.344	.005	-.029	-.008	-3.340	3928	<.001	<.001
Pair 3	Actual - PredictedGTCN	.016	.351	.006	.005	.027	2.815	3928	.002	.005
Pair 4	Actual - PredictedBiGTCN	-.009	.331	.005	-.020	.001	-1.783	3928	.037	.075

Table 7-5 Paired Samples Effect Sizes.

			Standardizer[a]	Point Estimate	95% Confidence Interval	
					Lower	Upper
Pair 1	Actual - Predicted1DCNN	Cohen's d	.420	-.008	-.040	.023
		Hedges' correction	.420	-.008	-.040	.023
Pair 2	Actual - PredictedTCN	Cohen's d	.344	-.053	-.085	-.022
		Hedges' correction	.344	-.053	-.085	-.022
Pair 3	Actual - PredictedGTCN	Cohen's d	.351	.045	.014	.076
		Hedges' correction	.351	.045	.014	.076
Pair 4	Actual - PredictedBiGTCN	Cohen's d	.331	-.028	-.060	.003
		Hedges' correction	.331	-.028	-.060	.003

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

The t-test also revealed a significant difference in the observed and forecasted values within the TCN model with ($p < 0.001$). However, as opposed to the results for the 1D-CNN, the paired correlation coefficient was relatively higher with ($r = 0.918$), suggesting that actual and forecasted values correspond even better to each other. The SD of the paired differences was lower, measured at (0.344), which is higher consistency compared to what was obtained for 1D-CNN. Such results confirm the impression obtained visually from the real versus the predicted graph, where the TCN seemed better aligned with actual class transitions, although it had distinct deviations in parts.

Table 7-6 Paired Samples Correlations

		N	Correlation	Significance	
				One-Sided p	Two-Sided p
Pair 1	Actual & Predicted1DCNN	3929	.875	<.001	<.001
Pair 2	Actual & PredictedTCN	3929	.918	<.001	<.001
Pair 3	Actual & PredictedGTCN	3929	.914	<.001	<.001
Pair 4	Actual & PredictedBiGTCN	3929	.926	<.001	<.001

In the GTCN model, there was a significant positive difference between the observed and estimated values, with ($p = 0.005$) and a paired correlation coefficient of (0.914). The SD was (0.351), somewhat larger than for the TCN, but the effect size Cohen's ($d=0.351$) suggests that overall, the GTCN made better predictions. The results of the t-test confirm the observations made based on plots, as while GTCN did navigate transitions more successfully and with fewer mistakes as compared to TCN, the results show the opposite.

The BiGTCN model, as the proposed structure, had the strongest paired correlation ($r = 0.926$), indicating the highest agreement between actual and predicted values. The t-test showed no significant difference ($p = 0.075$), and the effect size (Cohen's $d = 0.331$) was the smallest among all models, further confirming the robustness of the BiGTCN in minimizing predictive errors. This is consistent with the visualization of actual vs. predicted values, where the BiGTCN displayed the best alignment and the fewest misclassifications across the occupancy states.

This agrees with the visualization of actual versus predicted values, whereby the best alignment is for BiGTCN, which does not have many misclassifications in the series of occupancy states. To sum up, the t-test results support the conclusions from the graphs comparing actual versus predicted values. The BiGTCN clearly outperformed all other models in providing increased prediction accuracy, being more closely matched to the observed value, and it had the smallest effect size. These results illustrate how well the combination of the BiGRU with the TCN improves EV-charging station occupation forecasts. The performance of the GTCN was also very remarkable, and yielded considerable benefits from the baseline models, but BiGTCN is the most reliable and robust model on this task.

7.1.8.8. BiGTCN Model Hyperparameter Tuning

A hyperparameter tuning process was carried out to optimize the performance of the BiGTCN model using a grid search technique. The search space included six key hyperparameters: the number of filters in the TCN branch, kernel size, GRU units in the BiGRU branch, dropout rate, batch size, and learning rate. Ten random combinations, as shown in Table 7-7, of these parameters were tested to identify the optimal configuration for achieving the highest validation accuracy.

Overview of the BiGTCN Model

Table 7-7 Grid Search Combination

Tcn filters	Kernal size	GRU Units	Dropout Rate	Batch Size	Learning Rate	V-accuracy
128	2	32	0.1	32	0.001	85.77%
32	5	32	0.3	32	0.0005	83.77%
64	2	32	0.1	16	0.0001	83.99%
64	5	32	0.1	32	0.0005	85.07%
32	3	64	0.1	64	0.0005	83.68%
64	2	64	0.3	32	0.0005	84.71%
64	3	64	0.1	32	0.0005	85.03%
64	5	128	0.3	16	0.001	85.32%
32	3	32	0.1	16	0.001	83.81%
128	**3**	**128**	**0.2**	**32**	**0.001**	**86.04%**

The validation accuracy for each tested configuration varied, with the highest accuracy observed at the performing parameters (TCN filters: 128, Kernel size: 3, GRU units: 128, Dropout rate: 0.2, Batch size: 32, Learning rate: 0.001)

Configurations with fewer filters or larger dropout rates generally achieved lower validation accuracy. For example, a setup with (32) filters and a dropout rate of (0.3) achieved an accuracy of (83.77%), while increasing the number of filters and reducing the dropout rate consistently improved the results. This highlights the sensitivity of BiGTCN's performance to architectural and optimization parameters.

7.1.8.9. Model Future Prediction

In this experiment, the four models (1D-CNN, TCN, GTCN, and BiGTCN) were tasked to predict the occupancy state for the next six hours at three distinct charging stations (Loc-1, Loc-2, and Loc-3). The input interface required a specific timestamp in the format "YYYY-MM-DD HH:MM" and a station number (1, 2, or 3). The experiment aimed to evaluate the models' ability to recognize location-specific patterns and temporal variations, given that prior analysis highlighted Location 3 as the busiest among the three stations, with peak activity observed between 6 AM and 10 PM on weekdays.

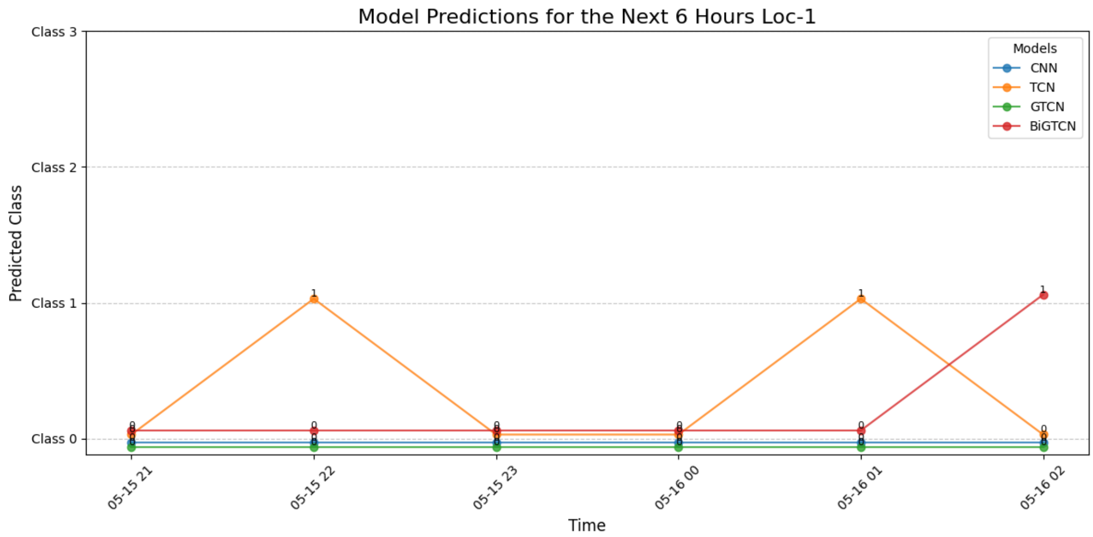


Figure 7-7 Future Predictions in Loc-1

The time window selected for illustration in this section (close to midnight) represents one of several time slots used during the broader experimental evaluation, which covered different hours of the day and days of the week. This particular example was included solely for demonstration purposes. It is acknowledged that midnight represents a relatively quiet period for Locations 1 and 2 but remains moderately active in Location 3 due to its residential surroundings. Therefore, this time slot is not intended to be representative of overall model behavior or peak periods but rather serves to exemplify how the model performs under lower activity conditions. Additional experiments were conducted at various time points to validate model robustness across different occupancy scenarios.

In Loc-1, Figure 7-7, the results for Loc-1 showed that TCN and BiGTCN exhibited intermittent predictions for Class 1 occupancy (moderate usage), while 1D-CNN and GTCN consistently predicted Class 0 (low usage). TCN occasionally predicted occupancy spikes but failed to sustain trends, whereas BiGTCN captured a peak in the final time slot.

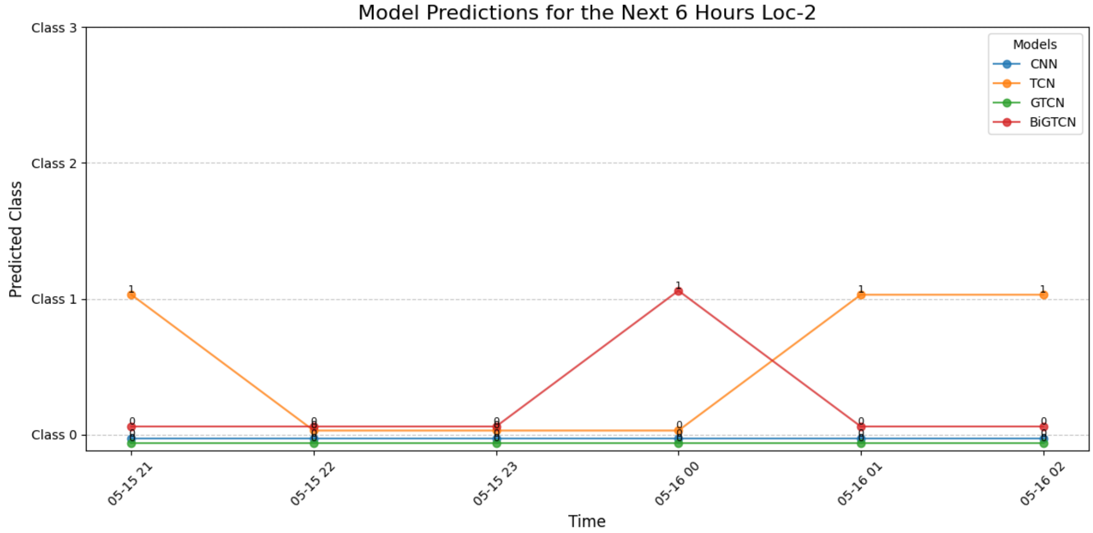


Figure 7-8 Future Predictions in Loc-2

In Loc-2, Figure 7-8, TCN and BiGTCN both captured occupancy variations, with BiGTCN predicting Class 1 during certain intervals. GTCN and 1D-CNN remained conservative, maintaining Class 0 predictions throughout. Notably, BiGTCN demonstrated more dynamic responsiveness to changes compared to TCN, particularly in recognizing moderate occupancy levels.

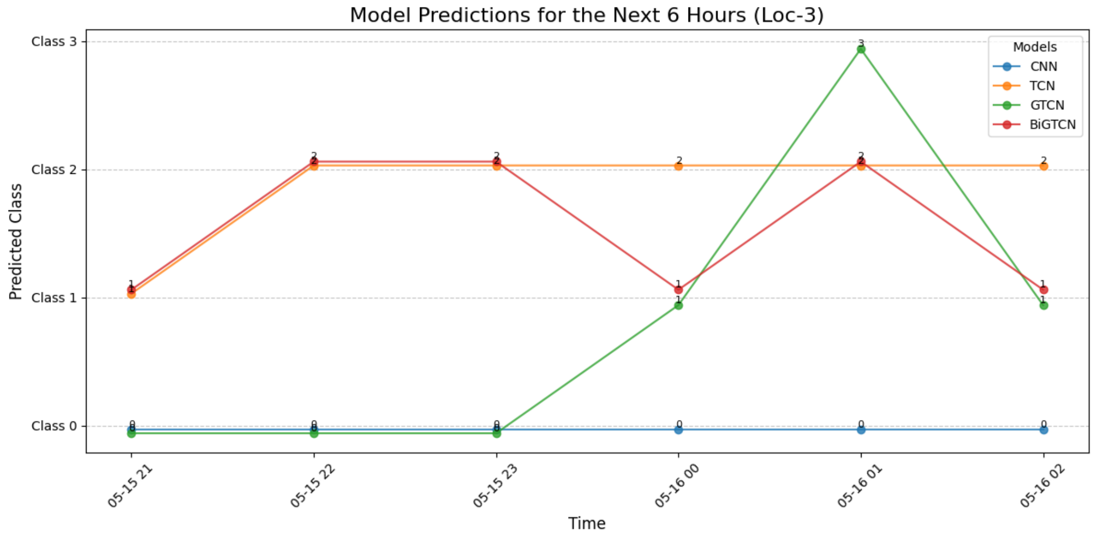


Figure 7-9 Future Predictions in Loc-3

For Loc-3, Figure 7-9, the models exhibited greater variability in their predictions. Both TCN and BiGTCN successfully captured the busier nature of this station by predicting Class 1 and Class 2 occupancy states. BiGTCN demonstrated more accuracy by detecting shifts from Class 1 to Class 2 and back. Conversely, GTCN predicted a sudden peak (Class 3) at one instance, aligning partially with the busier profile of Loc-3, though its overall trend fluctuated. The 1D-CNN consistently predicted Class 0, failing to recognize the dynamic occupancy patterns at this location.

7.2. Discussion

7.2.1. Models Performance

This section discusses the performance evaluation of the proposed architecture of the BiGTCN for EVCS occupancy prediction. The rationale behind the developed framework is to integrate the strengths of both TCNs and bidirectional GRUs. The performances of the proposed frameworks were evaluated against the baseline models: 1D-CNN, TCN, and a lightweight hybrid GTCN model. These experiments demonstrated that BiGTCN can capture the complex temporal and sequential relationships in a much better way than competing models across all evaluation metrics.

The architecture of BiGTCN was such that it captured both the long-range temporal dependencies with dilated convolutions inside the TCN block and the bidirectional sequential patterns through GRU layers. The integration of these components ensures a comprehensive understanding of both forward and backward contextual information, which is critical for accurately predicting EVCS occupancy. Compared with other models, BiGTCN demonstrated consistent performance with higher accuracy, lower losses, and better precision and recall, especially for the challenging class 3. This again justifies the incorporation of the BiGRU with the TCN to improve the learning capability regarding both the temporal and spatial patterns.

This experiment was based on the dataset that included temporal variables, meteorological conditions, and occupancy states regarding three different locations. This allowed for balanced and representative training, validation, and testing, considering the encoding of the target variable, normalization of data, and careful partitioning. All models underwent training and evaluation on an equal basis, considering metrics of classification that included accuracy, F1-score, and confusion matrix. These metrics provided an extensive analysis regarding the performance of the models on not just the general accuracy but also regarding the correct classification of each type of occupancy.

During training, the more complex models, BiGTCN and GTCN, converged faster when compared to their simpler versions, such as 1D-CNN and TCN. Both the BiGTCN and GTCN concluded training with an accuracy of (96%) and further decreases in training loss, reflecting their good learning capability. This is reflected in the fact that the generalization abilities of BiGTCN, although still sub-optimal, were better on the test dataset, while its test accuracy was at (91%) on the loss value of (0.34). The GTCN had comparable accuracy, but with a higher loss of (0.37). This comparison underlines the effectiveness of the BiGRU component in enhancing generalization by incorporating both historical and future temporal contexts.

Confusion matrices and classification reports gave further information with respect to the different categories of occupancies. Most impressively, the BiGTCN performed incredibly well for the most challenging classes, class 3, with (65%), compared to (55%) for the GTCN and (38%) for the 1D-CNN. This result indicates its better ability in capturing subtle and complex transitions between occupancy states. Similarly, the ROC curves and AUC scores also supported the fact that BiGTCN can distinguish classes with high confidence, achieving the highest AUC scores across all categories.

The paired sample t-tests and the effect size analyses further quantitatively confirmed that BiGTCN outperforms the other tested models. This was reflected in the highest correlation between the observed and forecasted values, ($r = 0.926$), the smallest effect size (Cohen's $d = 0.331$), and the lack of significant difference between the observed and forecasted values ($p = 0.075$). This is further supported by the visual depictions of the observed versus forecasted values where, out of all, BiGTCN had the best fit and the least misclassifications across all different levels of occupancy.

In contrast, the baseline models, especially 1D-CNN, showed huge divergences and anomalies, which demonstrated their inability to adequately capture temporal and sequential relations. This demonstrates that the architecture of BiGTCN outperforms others by a wide margin in terms of forecasting occupancy at an EVCS, integrating the BiGRU with TCN better for capturing long-range and sequential dependencies compared to other models. It is noted that although GTCN did very well among the baseline models, it does not generalize and classify as well as the proposed BiGTCN. All these results point to the potential of the BiGTCN to become a powerful and trustworthy method for occupancy forecasting in EVCSs.

7.2.2. The Hyperparameter Tuning Process

The hyperparameter tuning process demonstrated the significant impact of model configuration on prediction accuracy. The combination of (128) TCN filters, a kernel size of (3), and (128) GRU units allowed the model to effectively capture both short- and long-term temporal dependencies in the data. The modest dropout rate of (0.2) provided a good balance between preventing overfitting and maintaining the model's ability to generalize. Additionally, a learning rate of 0.001 ensured stable convergence during training.

The results align with previous studies emphasizing the importance of carefully tuning deep learning architectures for time-series forecasting tasks. For instance, Fan et al. (2023) highlighted that optimizing TCN parameters, such as kernel size, significantly impacts the ability of the model to capture temporal patterns. Similarly, the inclusion of Bidirectional GRU units allowed BiGTCN to enhance its bidirectional learning capabilities, which are particularly useful for highly dynamic datasets like EVCS occupancy data.

However, the process revealed trade-offs in model performance. Larger batch sizes (e.g., 64) often reduced validation accuracy, likely due to the difficulty in capturing finer temporal patterns in larger data chunks. Furthermore, configurations with lower GRU units (e.g., 32) struggled to model complex dependencies, underscoring the importance of sufficient capacity in the BiGRU branch.

7.2.3. Future Predictions

The results highlight significant differences in the predictive capabilities of the models, particularly in capturing temporal patterns and location-based variations. BiGTCN consistently outperformed the other models in recognizing occupancy trends at all three stations, particularly in Loc-3, where the busier nature required nuanced temporal learning. This improved performance can be attributed to the model's hybrid structure, which combines TCN's efficient temporal feature extraction with BiGRU's ability to capture bidirectional dependencies. As seen in Loc-3, BiGTCN effectively handled transitions between occupancy classes, demonstrating its suitability for high-demand scenarios.

TCN also showed strong performance but struggled to generalize in stations with lower occupancy (Loc-1 and Loc-2). Its unidirectional nature might have limited its ability to capture forward and backward temporal dependencies. Meanwhile, GTCN exhibited sudden peaks, particularly in Loc-3, suggesting a sensitivity to dynamic changes but lacking consistency in trend recognition. Finally, the 1D-CNN model underperformed across all locations, failing to capture variations and largely predicting Class 0 throughout, which highlights its limited capability for complex temporal forecasting.

The results align with previous findings that Loc-3 experiences higher occupancy than the other two stations, as reflected in the more dynamic class predictions. The observed peak times further validate earlier analyses that indicated higher demand during weekdays between early morning (6 AM) and evening (10 PM). These findings demonstrate that models like BiGTCN, which integrate advanced temporal learning techniques, are more effective in capturing the unique temporal and spatial dynamics of EV charging stations, particularly in high-traffic locations. Future improvements could involve incorporating real-time features or enhancing the models' adaptability to sudden spikes in demand.

Chapter 8: Quantitative Analysis – User Study A

Understanding the behaviour and preference of EV owners could go a long way toward model development with regard to EV-related use cases. Common charging behaviour among EV owners establishes the context that formulates decisions on the conditions and parameters of models, thereby improving accuracy and relevance. Quantitative results concerning EV drivers' behaviour can also inform feedback into model output assessment, through evaluating and enhancing predictive accuracy about actual user patterns.

The online survey for this research was conducted over a single round across approximately 20 days, from the 27th of August 2022 to the 15th of September 2022, and it collected responses from a total of (208) volunteers. During a two-step screening process of the collected data, responses were reviewed to ensure completeness and quality. In the initial revision step, (13) responses were identified as insufficient, with most questions left blank. These were excluded from further analysis to avoid skewing the results. In the second screening step, an additional (4) responses were removed due to incomplete ratings in the display modes section, which would have compromised the reliability of Likert scale analysis. After removing the insufficient responses, the final number of accepted responses from the volunteers became (191) participants.

During the survey preparation stage, researcher divided the structure into three main sections based on the purpose of the survey questions. The first part of the survey included demographic information, while the second part targeted at studying and extracting drivers' electric car charging habits. The third section focused on assessing various possible prediction displays and aimed to define the most suitable display form from the user perspective. The appendices of this thesis include a final version of the survey questions.

8.1. Demographic Information

The first question asks about the age group of each participant. Four choices were given in this question to describe their age cohort in years (18-25, 26-35, 36-49, and 50 and above); just over the half of the participants were aged between 26-35 (Figure 8-1). While the other age groups shared the other half, the “18-25” and “36-40” cohorts had similar numbers. The lowest number of participants were from the age group “above 50”. These results may indicate that most EV owners belong to the age group of “26-35”. This can be considered an incentive to ensure that many electric car owners are willing to accept technical means to help find charging places for their electric cars.

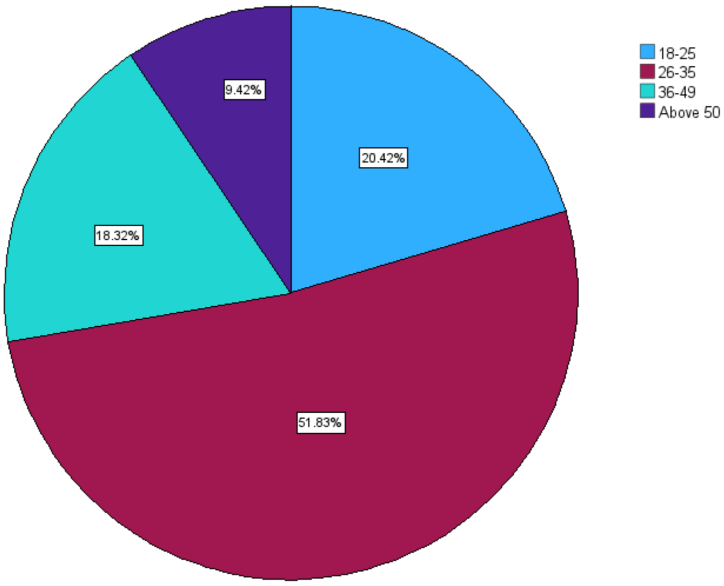


Figure 8-1 What is your age group?

The following question was about the volunteers’ gender. From the answers to this question, Figure 8-2 displays that the majority of participants identified as “female” (n = 112, 58.64%); only four participants preferred not to specify a gender, and the remainder identified as males (n = 75, 39.27%).

Demographic Information

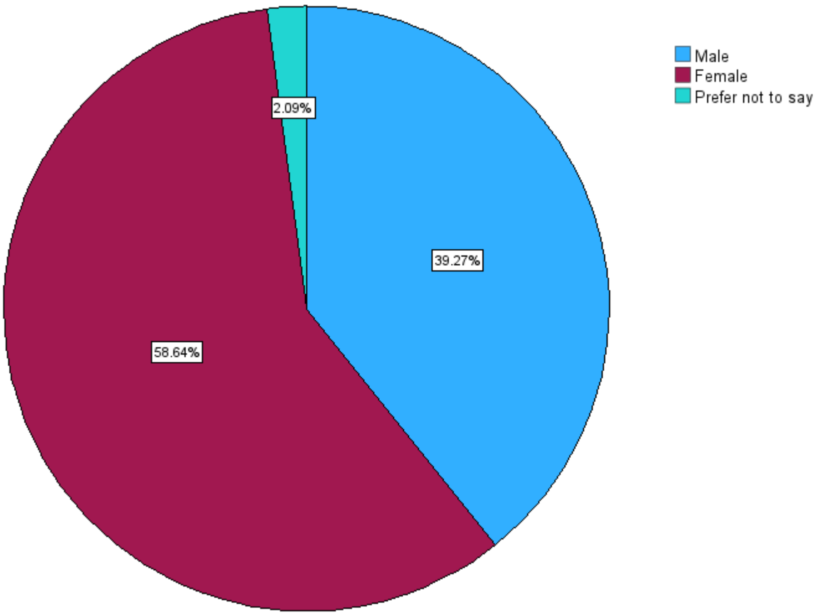


Figure 8-2 How Would You Describe Your Gender?

From Figure 8-3 Gender Representation Across Age Groups and Table 8-1, it is evident that the “26-35” age group had the highest number of respondents across all gender categories, with males being the most numerous within this age bracket. This suggests that the survey was most popular among young adults, particularly males, who are in the early stages of their careers or advanced education. The “18-25” age group shows a relatively balanced distribution between *male* and *female* respondents, with slightly more females. This could indicate a similar level of engagement with the survey topic among younger adults of both genders. The “36-49” age group shows noticeable lower number of respondents, with males leading but with a smaller margin compared to the “26-35” age group. This might reflect a decrease in the availability or interest of middle-aged adults in participating in surveys, or it could be a result of the specific population that was targeted or had access to the survey.

Demographic Information

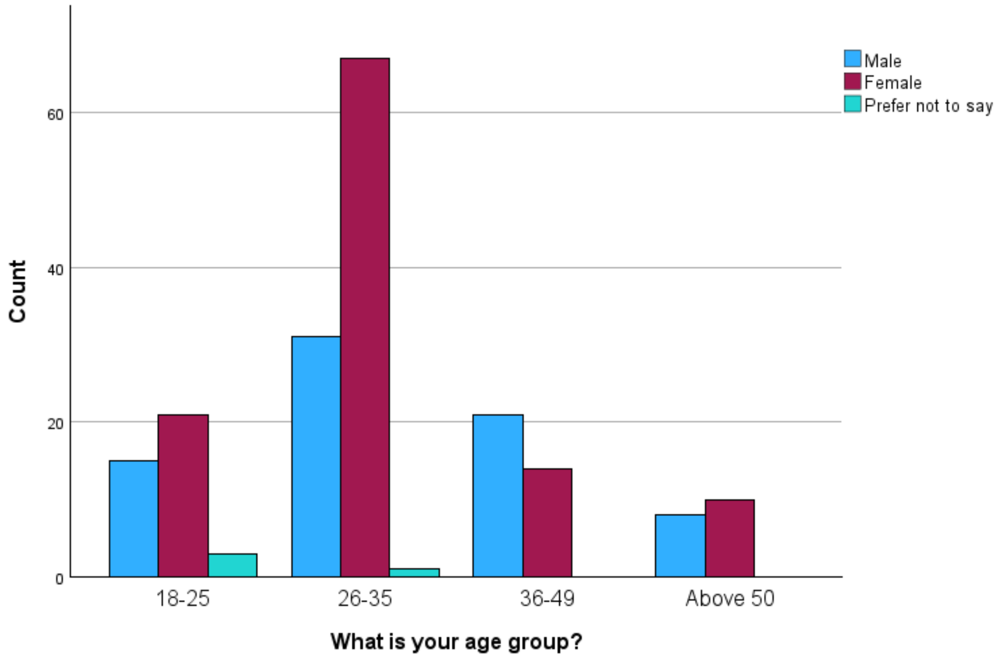


Figure 8-3 Gender Representation Across Age Groups.

Table 8-1 Demographic Information Distribution

				Count	Subtable N %	Table N %
What is your age group?	18-25	How would you describe your gender?	Male	15	38.5%	7.9%
			Female	21	53.8%	11.0%
			Prefer not to say	3	7.7%	1.6%
	26-35	How would you describe your gender?	Male	31	31.3%	16.2%
			Female	67	67.7%	35.1%
			Prefer not to say	1	1.0%	0.5%
	36-49	How would you describe your gender?	Male	21	60.0%	11.0%
			Female	14	40.0%	7.3%
			Prefer not to say	0	0.0%	0.0%
	Above 50	How would you describe your gender?	Male	8	44.4%	4.2%
			Female	10	55.6%	5.2%
			Prefer not to say	0	0.0%	0.0%
	Prefer not to say	How would you describe your gender?	Male	0	0.0%	0.0%
			Female	0	0.0%	0.0%
			Prefer not to say	0	0.0%	0.0%

For those above the age of (50), the number of respondents is the lowest among all age groups. This is consistent across all gender categories, with females slightly outnumbering males. The lower response rate in this age group could be due to a variety of factors, including less familiarity with the survey medium, lower interest in the survey topic, or fewer opportunities to participate. The category for those who “prefer not to say” their gender has the fewest respondents in all age groups, which is not uncommon in surveys as this option is typically chosen less frequently. The engagement with the survey decreases with age, which may be attributed to the recruitment method, such as reliance on platforms that younger individuals are more likely to use. This observation aligns with trends suggesting that older individuals, particularly in the UK and other European countries, might be less active on certain social media platforms. However, it is worth noting that older people often have greater disposable income, potentially making them more capable of purchasing EVs. These factors highlight the importance of considering diverse recruitment strategies to capture a more representative sample of EV users. Also, there is a consistent but small representation of individuals who prefer not to disclose their gender.

8.2. EV Experience and Charging Habits

Generally, the experience of owning an EV varies among drivers, with some becoming accustomed to owning and driving EVs over time. On the other hand, owning an EV can be a novel experience within a short span of time. This section of the survey enquired about the duration of EV ownership in the participants’ lives. The poll gave participants the options to indicate the length of their experience: up to one year, one to three years, or more than three years.

The aim was to ensure that the survey outcomes should not deviate to one side by one of the given choices of the owners’ groups. From the survey’s answer to this question, as seen in Figure 8-4 more than three-quarters of the participants had EV-owning experience for more than one year, while almost a fifth (17%) had owned an EV for only up to one year. This quantity enhances the possibility of ensuring answers from those with sufficient experience using EVs.

EV Experience and Charging Habits

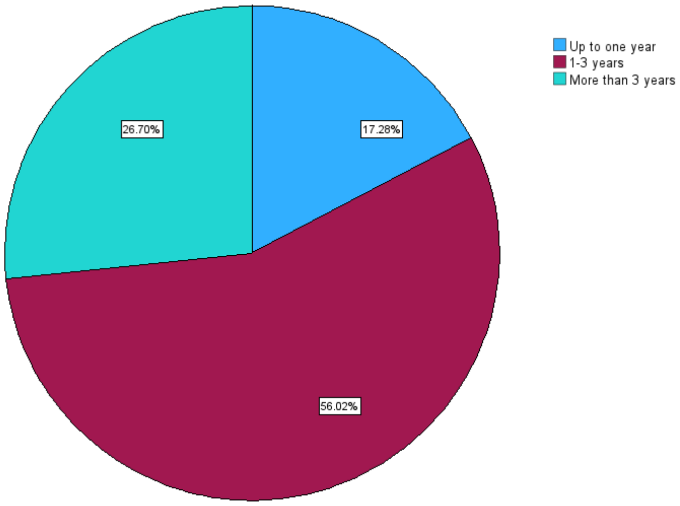


Figure 8-4 How Long Have You Owned or Driven An EV?

Charging point types could affect the charging duration, as fast-charging points lead to less charging time and vice versa. Participants were asked about the most used charging type they use to charge their vehicles. As shown in Figure 8-5 approximately (88%) participants selected either fast charging points (7kW) or (22kW) as their favourite charging option. However, a considerably smaller number of participants selected either rapid or slow chargers. This decision may be influenced by various factors, such as consensus or other tangible needs.

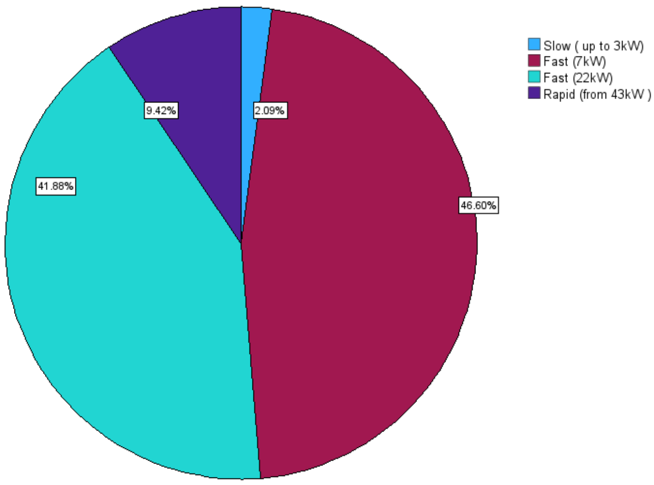


Figure 8-5 Which Charging Point Do You Mostly Prefer to Use?

Among home charging, workplace chargers, and public charging points, the option of home charging points was the most popular choice, as shown in Figure 8-6. Home chargers are typically more accessible and adaptable compared to public chargers, which are generally less adjustable. This shows similar results obtained from Anderson et al. (2023). An increasing demand for EVCSs is evident despite the widespread availability of home chargers (Viswanathan *et al.,* 2018).

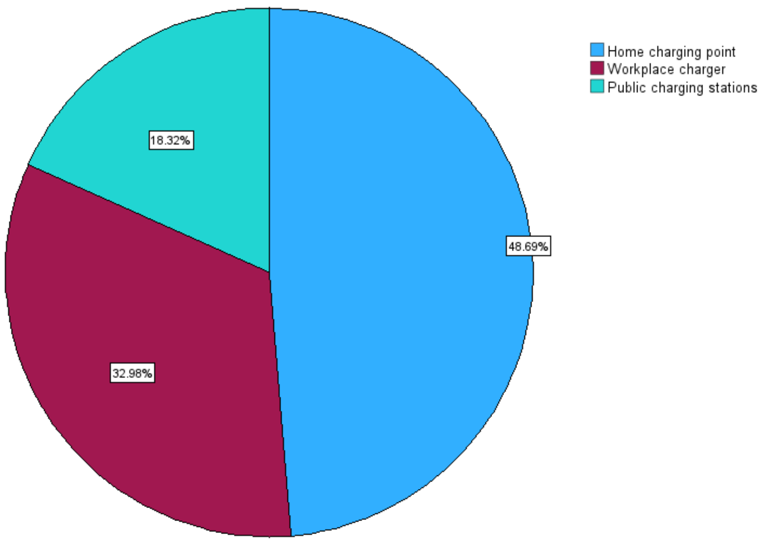


Figure 8-6 Where Do You Usually Prefer to Charge Your EV?

Depending on their charging habits, some EV drivers plan fixed time to charge their EVs during the day. The location and type of charging point used may impact determining and organizing these times. Therefore, participants in this survey were asked about their preferred time of day to charge their cars. The majority of them answered that they plan to charge times either during the middle of the day or overnight, while only just over (15%) chose the early morning option (Table 8-9).

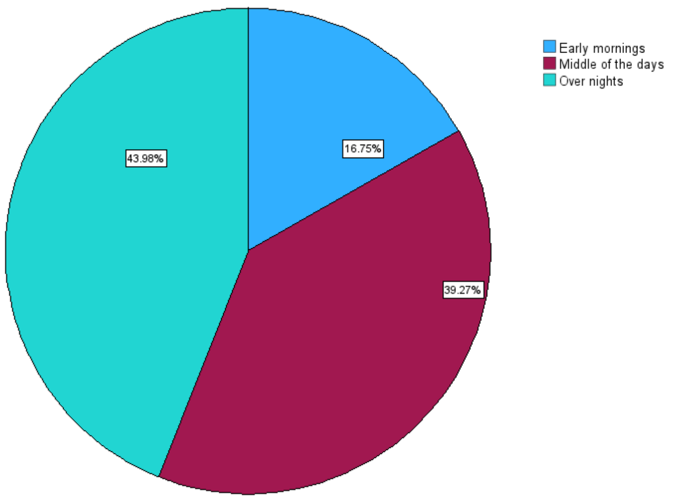


Figure 8-7 At What Time of the Day Do You Normally Charge Your Vehicle?

EV Experience and Charging Habits

This research identified both charging time and location as key attributes of the charging habits. Therefore, identifying any potential correlation between the two attributes could enrich the feasibility of this analysis. Knowing the type, strength, and direction of the relationship may be useful in highlighting a common habit among electric car owners when charging their cars. The bar chart in Figure 8-8 shows quantity of the chosen preferred charging time for each charging location as experienced by participants: *home charging points*, *workplace chargers*, and *public charging stations*. The charging times are categorized into *early mornings*, *middle of the days*, and *overnights*.

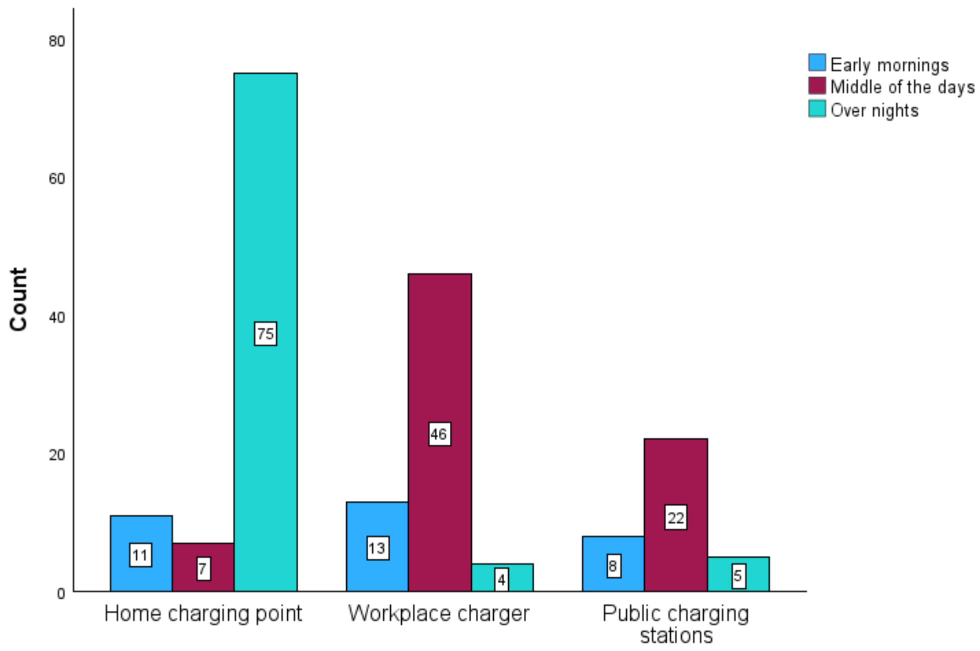


Figure 8-8 Preferred Charging Time Per Preferred Charging Location

According to the bar chart, it is clear that the majority of respondents prefer to charge their EVs overnight at home charging points, with a total of (75). This preference for overnight charging at home could be due to the convenience and lower electricity rates that are often available during off-peak hours. The significant difference between overnight charging and other times of the day at home suggests a strong relationship between the location (home) and the preferred charging time (overnight).

The bar chart indicates that the preferred charging time at workplace chargers is the middle of the day, with the highest count of (46). This preference for middle of days charging at the workplace suggests that individuals are taking advantage of the opportunity to charge their EVs while they are at work, which is both convenient and efficient, as the cars are likely parked for several hours during this time. The preference for middle of day charging at workplace chargers differs from that of home chargers, with workplace chargers favouring this time more than home chargers. This may reflect the general reality, as the majority of the study sample follows the common daily routine of being at work during the day and returning home in the evening.

Public charging stations show a more balanced distribution of preferred charging times, with a clear preference for the middle of the day, indicated by a count of (22). This may occur because drivers use public chargers when they are out during the day, either for work or personal errands, and need a quick charge to continue their travels. The lower counts for early mornings and overnights at public stations suggest a weaker relationship between these times and the use of public charging facilities.

Overall, the data indicates a strong relationship between charging location and preferred charging time, with home locations showing a strong preference for overnight charging. In contrast, public charging stations have a more pronounced preference for daytime charging. Public charging stations have a moderate preference for midday charging, reflecting the “on-the-go” nature of public charging needs. The analysis reveals that many individuals typically use workplace chargers during the middle of the day, aligning with their typical work schedule. The direction of the relationship suggests that the convenience and potential cost savings of overnight charging are significant factors for home and workplace locations, while accessibility and necessity during active hours influence the preference for public charging stations during the day.

The researcher used chi-square test to examine the correlation between the time and location of charging EVs, based on participants’ responses. The primary goal of this test is to determine whether there are statistically significant associations between the preferred time and preferred location for the survey participants to charge their EVs. Specifically, the testing was applied for the following:

H0: There is no significant association between the preferred time and preferred location to charge EVs.

HA: There is a significant association between the preferred time and preferred location to charge EVs.

The tables below show the output of running the chi-square test in SPSS.

Table 8-2 Cross-Tabulation of Distribution of EV owners' Charging Preferences Based on Location and Time of Day

Where do you usually prefer to charge your EV? * At what time of the day do you normally charge your vehicle?
Cross-tabulation

			At what time of the day do you normally charge your vehicle?			
			Early mornings	Middle of the days	Over nights	Total
Where do you usually prefer to charge your EV?	Home charging point	Count	11	7	75	93
		Expected Count	15.6	36.5	40.9	93.0
	Workplace charger	Count	13	46	4	63
		Expected Count	10.6	24.7	27.7	63.0
	Public charging stations	Count	8	22	5	35
		Expected Count	5.9	13.7	15.4	35.0
Total		Count	32	75	84	191
		Expected Count	32.0	75.0	84.0	191.0

In Table 8-3 (a), it can be seen that (Pearson Chi-Square = 105.516, P <0.001). As the p-value is below the common threshold (0.05), this means that H0 can be rejected; therefore, this indicates there is significant association between the variable. In Table 8-3(b), Phi (0.743) and Cramer's V (0.526) are both significant (p < 0.001). These values suggest a strong association between charging location and time of day preference.

Table 8-3 (a) Relationship Between Preferred Charging Locations for EVs and Preferred Time of Day: Chi-square Test Results, (b) Symmetric Measures Analysis

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	105.516[a]	4	<.001
Likelihood Ratio	121.593	4	<.001
Linear-by-Linear Association	44.320	1	<.001
N of Valid Cases	191		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.86.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.743	<.001
	Cramer's V	.526	<.001
N of Valid Cases		191	

8.2.1. Battery Depletion Risk

Recharging EV batteries is frequently necessary for exhausted batteries. Carelessness in managing the quantity of charge held in the automobile battery may result in the battery being fully drained. One concern for EV owners is the car battery depleting or nearing depletion before reaching the intended destination. This situation may have many consequences, such as anxiety, delays in appointments, and increased travel-time. The participants in this survey were asked to determine the prevalence of the phenomena by asking if they had experienced the problem before and, if so, how frequently.

The bar chart in Figure 8-9 shows the count of the responses for participants when they were asked if they have ever nearly exhausted their EV's battery before reaching a charging station. The figure shows a diverse range of experiences among EV drivers when it comes to dealing with battery depletion on the way to charging stations. A large group comprise (70) individuals indicated they have never encountered this issue. This may suggest their level of confidence in their vehicle's range or the availability of EVCI. Even a slightly larger group, making up (81) respondents, admitted to having been close to depleting their battery. This could be highlighting a potential concern for range anxiety among a significant portion of EV users. A smaller number of (36) owners experienced this problem once or twice, which may point to occasional challenges with range estimation or charging station access. Finally, only (4) participants reported they faced this issue three times or more, indicating a more frequent struggle with the vehicle's range or charging network limitations.

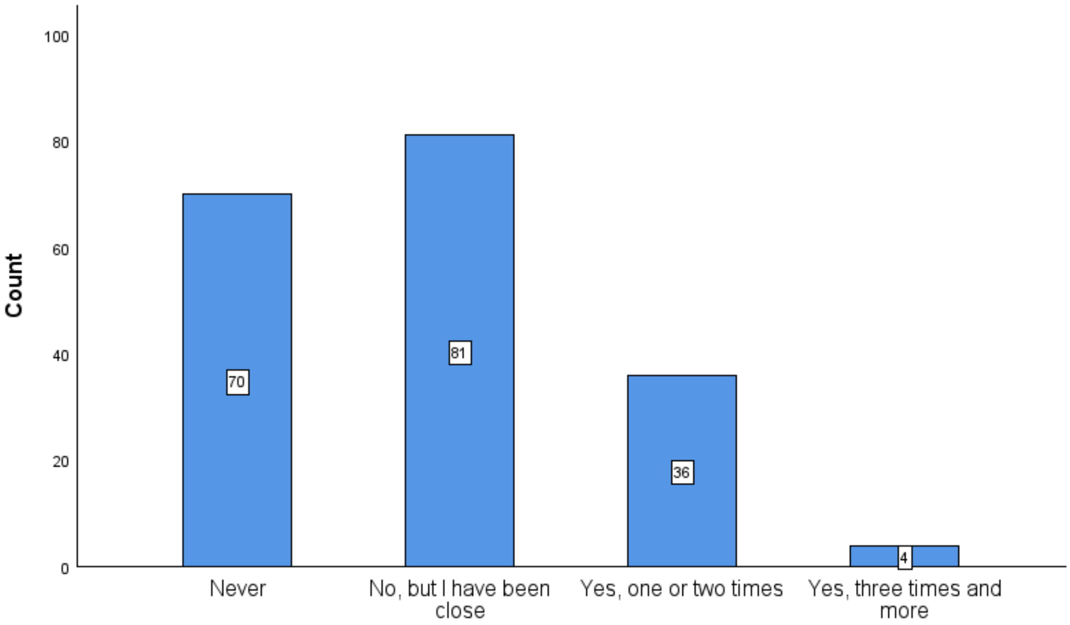


Figure 8-9 Number of Participants Who Had Experienced Their Battery Running Out

Participants were surveyed about the situations that could be good reason for encountering the risk of battery depletion. Most cited either inadequate previous planning for refuelling, or the absence of a charging outlet at the necessary time and location. The lack of organising suitable charging times and locations for EVs in both situations is seen as the key factor that caused these issues. It is crucial to focus on developing ways to decrease their frequency. The answers to the previous question of the survey were categorized into groups based on participants' experience duration with electric cars to examine the potential correlation with the frequency of car battery issues, as shown in Table 8-4.

Table 8-4 Percentage of EV owners Encountering Battery Problems for Each Category

How long have you owned or driven an EV? * Have you ever experienced or almost experienced your EV battery being drained before reaching a charging point?

Cross-tabulation

			Have you ever experienced or almost experienced your EV battery being drained before reaching a charging point?				
			Never	No, but I have been close	Yes, one or two times	Yes, three times and more	Total
How long have you owned or driven an EV?	Up to one year	Count	8	14	10	1	33
		% within battery depletion risk?	11.4%	17.3%	27.8%	25.0%	17.3%
	1-3 years	Count	27	55	22	3	107
		% within battery depletion risk?	38.6%	67.9%	61.1%	75.0%	56.0%
	More than 3 years	Count	35	12	4	0	51
		% within battery depletion risk?	50.0%	14.8%	11.1%	0.0%	26.7%
Total		Count	70	81	36	4	191
		% within Battery Depletion Risk?	100.0%	100.0%	100.0%	100.0%	100.0%

The cross-tabulation indicates that the majority of responds indicating "yes" were from the middle group, with (1-3) years of EV experience. This may be due to the fact that the majority of the responses belong to this category, as seen in Figure 8-4. Another consideration is that the first category may have less experience travelling compared to the second category, as many of them may be novice drivers. The low percentage from the third group could be attributed to their extensive experience in handling EVs for more than (3) years.

8.2.2. Prediction Model Usability

About (114) of the survey participants felt either "somewhat comfortable" or "extremely comfortable" using online platforms to help them finding charging location for their EVs (see Figure 8-10). Using an online platform can support the usability of the charging plan; this is related to the extent to which EV drivers accept to rely on them.

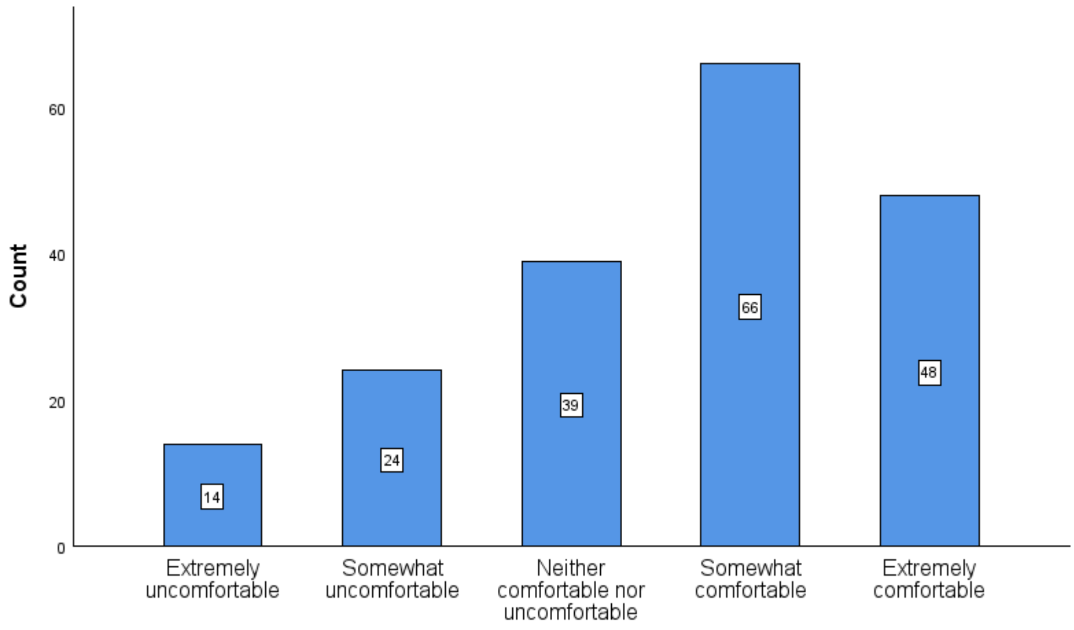


Figure 8-10 Using Online Platform for Charging Points Future Prediction

8.2.3. Preferred Predictions Scale

The predictions could be utilized for a short-term timescale in the future, or could be extended for a longer timescale; the methodology selected for producing the future forecast dictates the length (i.e., window) of future forecasting. Users may have their own thoughts about the forecasting resolution. In this survey, only around (6%) of the participants think they need to see more than (6) hours of predictions. While the remaining looking to a short term less than (6) hours, with the majority about (42%) for the choice from (3- 6) hours. This may refer to the short-term preferences over the longer one. In the following question, the majority of participants explained their choice, stating that they needed the prediction to assist them in urgent matters.

EV Experience and Charging Habits

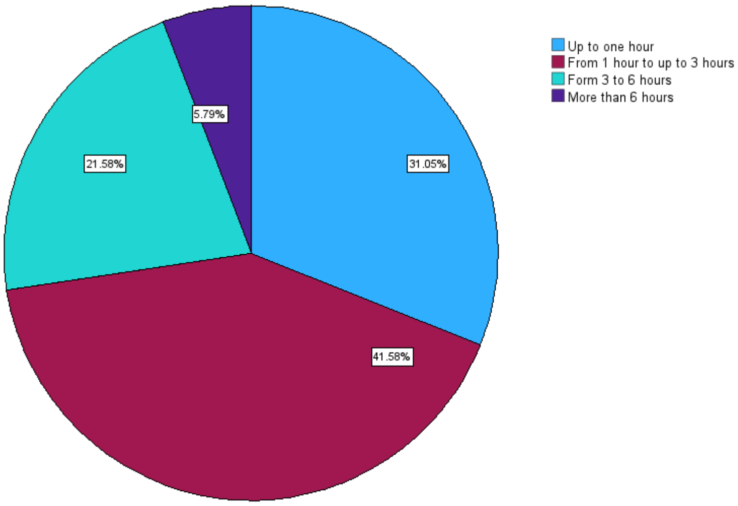


Figure 8-11 Resolution of the Future Predictions

These findings directly influenced the design and evaluation of the predictive models in this research. Specifically, the predictive models were configured to forecast occupancy up to 6 hours ahead, ensuring alignment with user expectations and practical usage contexts, particularly for users with high range anxiety or flexible travel routines.

8.2.4. Display Modes

The final part of this study aimed to investigate the more preferred display mode for the prediction outputs in terms of the clarity and usefulness. Participants were presented with a mixture of three different display modes ("continuity-based chart", "discrete-based chart", and "textual short") in separate questions, and were asked to score them using dislike great deal to like great deal five-point scale.

8.2.4.1. Group 1

The results for Group 1 are displayed in Figure 8-12, Figure 8-13, and Figure 8-14. In Group 1, the survey results show a clear indication of the participants' preferences towards a specific display mode from the Group 1 regarding the usefulness and clarity to show the results of predictions. The display modes in the Group 1 were labelled as "Continue1 graph", "line graph", "text display", "discrete graph", and "continue graph".

Continue1 Graph (Mode 1): The continue1 graph was one of the most liked displays with (28.27%) of "like a great deal" choice and (19.37%) for "like somewhat". It also had the lowest number of participants disliking it a great deal (only 9.95%), and (17.80%) for "dislike somewhat" suggesting a generally positive reception.

Discrete Graph (Output 2): The discrete graph observed more leaning towards negative feedback, with almost half of ratings for either "dislike somewhat" (25.13%) or "dislike a great deal" (20.42%). However, this option also in the position, where participants were not sure or stay balanced, as shown the most "neither like nor dislike" option with (24.08%).

Text Output (Output 3): The text output had a balanced distribution of responses across all categories, with the most rating was "like somewhat" (25.13%) then "like a great deal" (21.99%). This suggests that while the text output was considered useful by many, it did not stand out clearly to be the clearest or most useful.

Line Graph (Output 4): The line graph was generally well-acknowledged, with the majority of participants expressing a positive view, (28.27%) of the responses for "like a great deal" and (19.90%) of responses for "like somewhat" option. However, a notable percentage of participants disliked it somewhat (24.08%), indicating that while the line graph was useful for many, it did not meet the preferences of all users.

Mixed Graph (Output 5): The mixed graph received the most polarized feedback, with the highest percentage of participants selecting "dislike a great deal" (41.36%). Only (10.99%) of the participants with "like a great deal" option and (15.71%) "like somewhat".

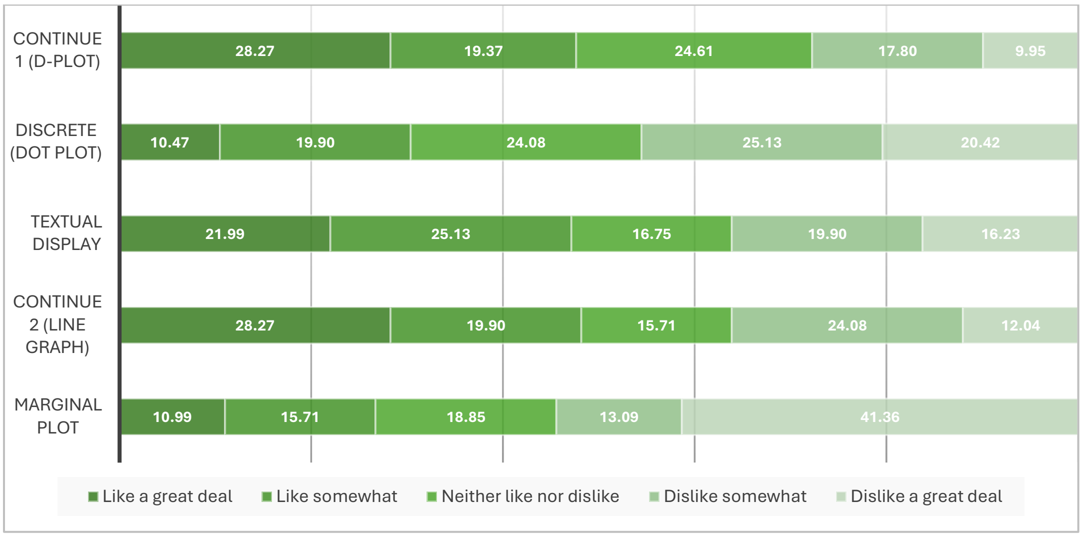
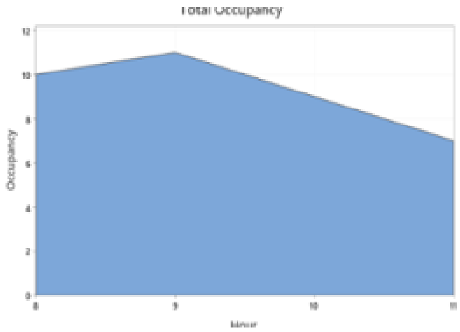
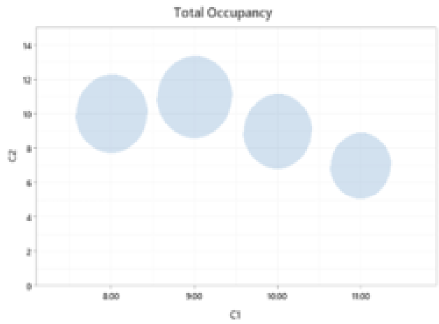


Figure 8-12 Participants' Rating Distribution for Display Models – Group 1



Display Mode (1) Continues1.



Display Mode (2) Discrete1.

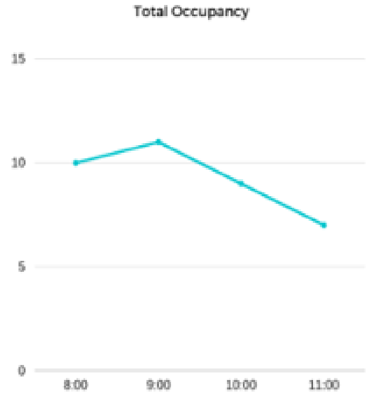
The prediction of the current charging point for the following three hours are:

1- 30% free

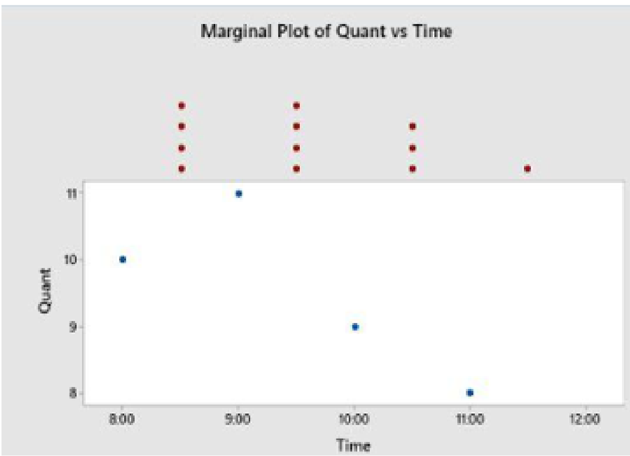
2- 50% free

3- 20% free

Display Mode (3) Textual Mode.

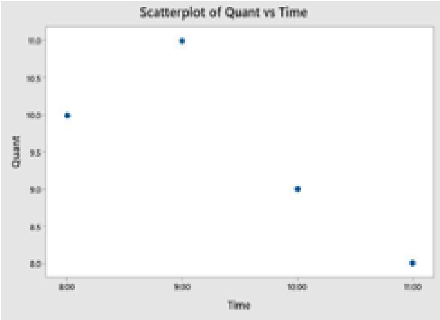


Display Mode (4) Continue 2 "Line Graph".



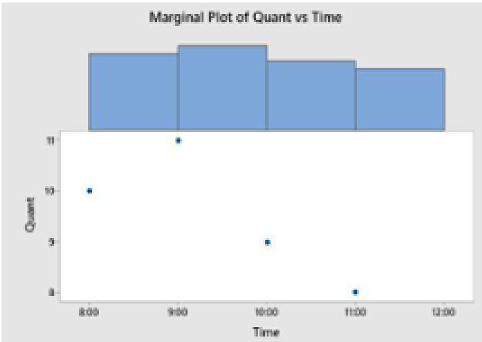
Display Mode (5) Marginal Plot - Discrete2

Figure 8-13 Different Display Modes Used in Survey Questions – Group 1

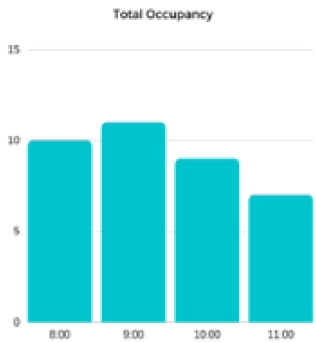


Display Mode (1) Discrete1 "Scatterplot".

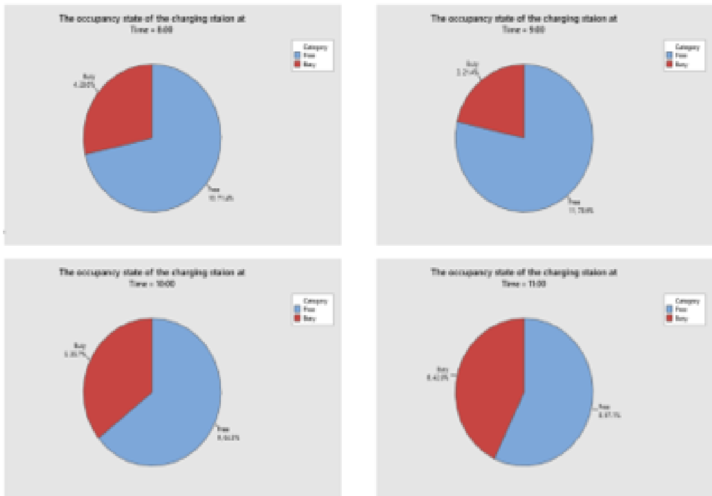
Display Mode (2) Textual Mode.



Display Mode (3) Marginal Plot.



Display Mode (4) Discrete 2 "Bar chart".



Display Mode (5) Pie Chart - Percentages

Figure 8-14 Different Display Modes Used in Survey Questions – Group 2

8.2.4.2. Group 2: Preferences for Display Modes

Figure 8-15 illustrates the diversity in participants' preferences across five display formats in Group 2: Pie Chart, Bar Chart, Marginal Plot, Text Output, and Scatterplot.. Each format was rated on a five-point Likert scale ranging from "dislike a great deal" to "like a great deal". The range of responses indicates varying levels of perceived clarity and usefulness among the different visualization types.

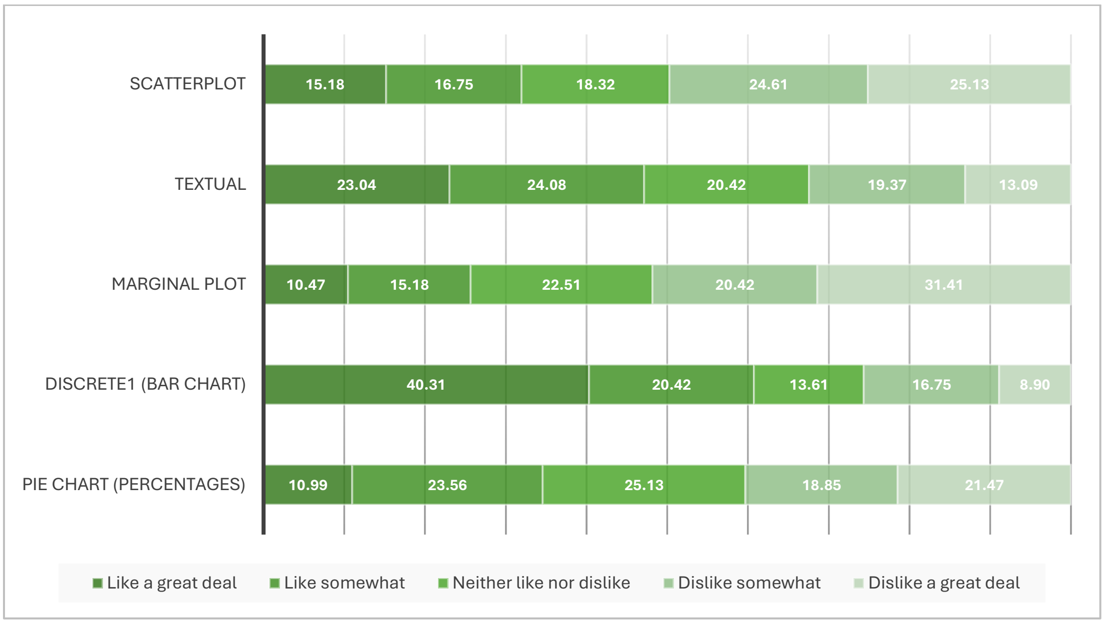


Figure 8-15 Participants' Rating Distribution for Display Models – Group 2

Generally, some display modes such as "Bar Chart" and "Text Display" tend to hold high frequencies for positive ratings such as "like somewhat" and "like a great deal". This could mean that such display formats are perceived to be clearer or better in serving their purpose in relation to the delivery of information. In contrast, the Marginal Plot and Scatterplot received more mixed ratings. These formats showed broader dispersions across both positive and negative categories, indicating greater variation in user interpretation and comfort.. A breakdown of the display formats is as follows:

"Pie Chart": This relatively even ratings, with a slight majority negative for the display. Generally, participants found that pie charts were not very helpful.

“**Bar Graph”:** This was one of the favourite formats; many participants voted for “like somewhat” and “like a great deal”. The format was found clear and, therefore, it can be considered an effective to present information.

“**Marginal Plot”:** Views were mixed, with some saying it was helpful, but the rating was not entirely positive, suggesting that not all participants found this easy to comprehend.

“Textual Display Mode”: This more traditional option was another very well-acknowledged format with strong preference in the positive direction. Several subjects found this mode clear and to the point, thus one of the better-liked options.

“**Scatterplot”:** Ratings were scattered; while there was a preference toward positive options, some negative responses showed up. While users liked it, some felt this was less helpful than the other formats.

The descriptive statistics provided in Table 8-6 support these conclusions. The Bar Chart received the highest average score (3.66), followed by Text Output (3.25) and Pie Chart (2.84). Marginal Plot and Scatterplot scored lower, reflecting the greater diversity of opinion and potentially lower effectiveness for this audience. The overall attitude towards each of the display format options was be calculated further to find out which is most favoured for EV owners. To enable a more nuanced interpretation of preferences, each Likert category will be mapped to a numerical interval scale (as shown in Table 8-7). For instance, a rating of “dislike a great deal” corresponds to a score between (1.00) and (1.80). This transformation will allow quantitative comparison of user attitudes and will aid in identifying the most user-friendly formats for EV owner engagement.The totals from the first analysis produced conflicting trends, with certain display formats being highly preferred compared to other ones. These differences in rating, therefore, indicate that the careful selection of display format will be necessary to achieve the maximum readability that the user expects. The further evaluation, based on score ranges, shall enhance this understanding by underlining which display modes will provide help to users in making inferences relative to EV usage and charging preference. Table 8-5 and Table 8-6 display the statistical descriptions for the display ratings of Group 1 and Group 2 (respectively).

Table 8-5 Display Rating Statistical Description – Group 1

Display Modes	N	**Mean**	SD	Variance
The Continue1 Graph "Density Plot"	191	**3.38**	1.328	1.764
The Discrete Graph	191	**2.75**	1.277	1.631
The Text Output	191	**3.17**	1.400	1.961
The Continue2 Graph "Line Graph"	191	**3.28**	1.408	1.983
Marginal Plot	191	**2.42**	1.434	2.055
Valid N (listwise)	191			

Table 8-6 Statistical Description for The Display Rating – Group 2

Display Modes	N	**Mean**	SD	Variance
The Scatterplot "Discrete2"	191	**2.72**	1.400	1.959
The Text Output	191	**3.25**	1.352	1.829
The Marginal Graph	191	**2.53**	1.349	1.819
The Bar Chart "Discrete1"	191	**3.66**	1.381	1.908
The Pie Chart "Percentages"	191	**2.84**	1.306	1.705
Valid N (listwise)	191			

8.2.4.3. Overall Attitude and Views Towards Different Display Formats

In the assessing attitudes and personal traits, it is challenging to convert the qualities items into quantitative measure to be used in data analysis. The study adapted the attitude scale used by Sözen and Güven (2019) to assess the attitudes and views of EV owners about various display formats and to analyse their attitudes in relation to clarity and information. The interval length is calculated as in the equation 6-1:

$$Interval\ length = \frac{R_{max} - R_{min}}{ls}$$

Equation 8-1

Where R_{max} is the maximum rating score, R_{min} is the minimum rating score, and ls is the Likert scale points.

In this study a five-point Likert rating scale was used. Therefore, from Equation 8-1 the interval length is (0.8). Table 8-7 shows the rating range used to measure the attitude of participants from the Likert scale in this study.

Table 8-7 Rating range of Likert scale of the study

Scale	Score Range
Dislike a great deal	[1:1.80)
Dislike somewhat	[1.80:2.60)
Neither like nor dislike	[2.60:3.40)
Like somewhat	[3.40:4.20)
Like a great deal	[4.20:5]

To analyse the participant responses, the mean of the rating score for the display modes of the two groups was produced from SPSS, as shown in Table 8-5 and Table 8-6, and the overall attitude was determined by comparing the mean values of each display mode with the score ranges in Table 8-7. The analysis revealed that although both Continue displays (Density-Plot and Line chart) received the highest mean rating score of (3.38) and (3.26), both display modes were placed within the “Neither like nor dislike” category along with the “Discrete” and “Text” displays with mean scores of (2.75) and (3.18) respectively. The “Mixed” method, with a lower mean score of (2.4), was categorized as “dislike somewhat”. These findings indicate a range of participant attitudes towards the different chart display methods, with “Continue” being the most favoured and “Mixed” the least.

Table 8-8 Participant General Attitude – Group 1

Display Mode	Mean rating score	Participant Attitude
Continue1 (Density-Plot)	3.38	Neither like nor dislike
Continue2 (Line chart)	3.26	Neither like nor dislike
Textual Display	3.18	Neither like nor dislike
Discrete (Dot-plot)	2.75	Neither like nor dislike
Marginal Plot	2.4	Dislike somewhat

In the Group2, “Discrete1 Bar-chart” scored the highest Mean rating score with (3.66). Therefore, this display mode assigned as “Like somewhat” participant attitude, which shows high preferences for participants. This followed by the text display with mean rating score (3.25). Similar to the first group, the marginal plot obtained overall dislike attitude with the lowest mean rating score of (2.53).

Table 8-9 Participant General Attitude - Group 2

Display Mode	Mean rating score	Participant Attitude
Discrete1 (Bar chart)	**3.66**	**Like somewhat**
Textual Display	3.25	Neither like nor dislike
Pie chart (Percentages)	2.84	Neither like nor dislike
Scatterplot	2.72	Neither like nor dislike
Marginal plot	2.53	Dislike somewhat

8.3. Discussion

8.3.1. Geographical Variation in Charging Time Preference

The survey's results have validated **H1**, that EV drivers associate several charging times with specific places: home, workplace, and public stations. It seemed that there is considerable preferred charging at home over night, which may be the result of off-peak electricity tariffs, coupled with easiness of access to home charging. This confirms the previous research of Viswanathan *et* al. (2018), who pointed out that cost-saving potential is among those factors leading most to home-charging behaviour. Moreover, this pattern aligns with results by Morstyn *et al*. (2018) who noted that uncoordinated charging habits, particularly at home during peak evening hours, contribute significantly to grid load variability. In terms of workplace charging, respondents showed a very clear preference for midday charging, reflecting common sense in charging while parked up at work.

This also corresponds to the observation of Qiao and Lin (2021) that similar charging patterns aligned with work timings. In contrast, public charging stations are distinctly flatter throughout much of the day, suggesting that public charging is used more for immediate needs, or "top-up" purposes, rather than for routinized charging. This variability reflects the findings by Ma *et al*. (2022), which noted that user preferences in urban areas are often dictated by immediate situational needs, rather than routine schedules. This geographical difference in charging preference identifies the importance of including spatial-temporal variables in predictive models for capturing the real-world active behaviour of users, which would definitely enhance its forecasting accuracy.

8.3.2. Preference of Online Platforms and Display Modes

The survey results support **H2**, showing that a majority of participants prefer using online platforms to access EV charging predictions, which highlights the increasing role of digital tools in EVI. Over 60% of participants indicated a comfort level with online platforms, emphasizing the need for accessible digital interfaces in predictive model applications. This aligns with the work of Hecht *et al*. (2021), who underscored the importance of user-friendly, accessible interfaces in EVI management.

The survey also showed characteristic partialities toward different display modes. Among the two groups, the results showed that *Discrete1* (bar chart) was rated the highest, with a score of (3.66), falling into the category "like somewhat". Followed by *Continue1* (density plot) with a mean rating of (3.38). These two highest-rated display types suggest the importance of clarity and the ease of conveying information, because the related display modes are so intuitive. This supports the argument of Soldan *et al*. (2021) that clear, visually accessible formats are the key to clearly communicating complex data to its users. Certainly, some of its more complex formats, like marginal plots, receive mixed comments; this would seem to confirm that the simplicity of display is particularly important to users when they need to get a quick sense of charging station availability.

8.3.3. Battery Depletion Anxiety and EV Ownership Experience:

The survey also shed light on **H3**, concerning the relationship between battery depletion anxiety and duration of EV ownership. Indeed, from the data, new EV owners, especially those with less than one year experience of EV, exhibit a higher likelihood of being anxious about the risk of the depletion of their batteries. This might be because feeling range anxiety is more common among newer EV owners, who are not yet accustomed to the changed charging and range aspects peculiar to EV use. Therefore, owners who had more experience with EVs, especially those with over three years of experience, reported fewer cases of battery depleting anxiety, reflecting an increased familiarity with their vehicle's range and charging options. This would indicate that experience linked with the ownership of EVs comes along with more effective planning, less concern about battery range, and predictive models may have opportunities to provide concrete guidance for beginning users. These may include alerts or recommendations regarding the best time and place to charge, which would let novice EV users build their confidence in range management.

8.3.4. Practical Implications for Predictive Model Development

The result shows that the design of predictive models needs to pay attention not only to the accuracy of the occupancy forecasts but also to the type of displays preferred by users. Choices are very strong for *Discrete1* (bar chart) and *Continue1* (density-plot), which should be considered as primary display formats in prediction interfaces. Accessibility and ease of understanding are ensured in these formats, as data is simple to understand and easy to read, which is a core aspect for a quick, informed charging decision. Accessibility and ease of understanding are supported by the survey findings, which show that display formats such as bar charts and density plots were the most preferred by participants (see Table 8-5 and Table 8-6), reflecting their clarity, simplicity, and effectiveness in enabling quick, informed decision-making. This points out a trend where predictability increases for new EV owners with a high degree of range anxiety. Thus, the predictive models must be required that have user-based adaptability in finding appropriate places and times to charge, considering the driver's experience. That would let the predictive tool help improve ownership confidence by making things as easy as possible, especially during those critical first days.

Further development and refinement of interface displays for predictive models are needed, with more research in a broader demographic setting. Moreover, features such as user experience level, favoured charging times, and commonly used locations could add more satisfaction and engagement. Qiao and Lin (2021) identified that personalization in display and functionality can greatly enhance the interaction of users with predictive models, thereby extending their effectiveness in real-world settings.

The continue graph (Output 1) emerged as the most preferred form of output among the participants, particularly among those aged 26-35 and 36-49. The mixed graph (Output 5), despite receiving the highest number of negative responses, also had a notable number of participants who found it highly useful, indicating that it may serve specific user needs well. The line graph (Output 4) and text output (Output 3) received mixed reactions, while the discrete graph (Output 2) was less favoured but still found useful by some participants. These insights can guide the selection of output forms for presenting model results to different user groups, emphasizing the need to consider the varied preferences across age demographics.

8.4. User Analysis Conclusion

Results in this chapter also confirm **H1**, whereby EV drivers associate specific charging times with particular locations (home, workplace, and public stations). The strong tendency toward home night charging, which probably relies either on off-peak electricity tariffs or comfort of home access, is shown here. Workplace charging was most preferred at noon time, although mid-morning and late afternoon charging was relatively high, reflecting parking-based charging at work. In contrast, public charging stations expressed more of a uniform daily pattern, suggesting that public charging caters to urgent needs or “top-ups” rather than day-to-day charging practices. This variation is expected and matches findings in the literature were showing that charging preference among urban users often depends on contextual needs rather than a fixed routine.

Supporting **H2**, the results showed that most of the participants would wish to use online platforms to access EV charging predictions, therefore building the role of digital instruments within EVI. Over (60%) of the participants were comfortable with online platforms; this implies that there should be friendly digital interfaces in predictive model applications. The survey also provided insights into **H3**, testing the relationship between battery depletion anxiety and EV ownership experience. The results demonstrated that newer EV owners, particularly those with less than a year of experience, have higher levels of battery depletion anxiety. This might be explained by their lack of familiarity with either the range and/or charging features associated with EV use. Inversely, owners with upwards of three years of experience felt lower levels of anxiety, indicative of greater familiarity with both their vehicle’s range and charging options available. This suggests that experience with EV ownership leads to more effective planning and reduced concern about battery range. Predictive models have the opportunity to provide guidance for novice users, including alerts or recommendations for optimal charging times and locations, thereby building confidence in range management.

Regarding display preferences, the survey indicated a strong preference for certain display modes. The *Discrete1* (bar chart) and *Continue1* (density plot) interfaces were rated the highest, suggesting that clarity and ease of information conveyance are crucial. This may refer to simpler display formats facilitate quick understanding, which is vital for making informed charging decisions. Text display mode also shows a moderate rating range among participants. Further studies could help to evaluate display mode preferences among EV owners in more detail, piloting alternative display options and aesthetics for user convenience and comfort. Therefore, in the following chapter, additional user study conducted to emphasise more the EV owners charging habits and their thought toward the prediction models including prediction display modes.

Chapter 9: Qualitative Analysis - User Study B

This chapter provides further study on the outcomes of the previous chapter to add some qualitative insights, extending on some of the quantitative survey results reported in Chapter 8. This chapter is dedicated to answering **RQ5**, identifying key factors influencing EV owners' trust in predictive models as a source to manage their charging times, and it explores further thoughts about EV users' behaviour.

9.1. Qualitative Study Analysis

For the qualitative wing of this research, semi-structured interviews were conducted by the researcher with (11) UK-based individuals who are EV owners.

9.1.1. Demographic Information

Table 9-1 shows demographic details of the participants in this qualitative study. It summarizes some of the key demographic and behavioural characteristics of the survey participants (i.e., experience of EV ownership, usual places for car charging, and level of familiarity with different digital platforms).

Table 9-1 Main Characteristics of Interview Participants

Participant number	Sex	EVO experience (years)	Most frequent charging location	Online platforms experience
P1	F	2-4	Home	Regularly
P2	M	>4	Home	Sometimes
P3	M	1-2	Public	Regularly
P4	M	>4	Home	Sometimes
P5	M	1-2	Home	No
P6	F	1-2	Home	Sometimes
P7	M	1-2	Public	No
P8	M	2-4	Home	Sometimes
P9	F	>4	Home	Sometimes
P10	F	2-4	Public	Sometimes
P11	M	1-2	Public	Regularly

The sample has a total number of eleven respondents, including six males (**P2**, **P3**, **P4**, **P5**, **P7**, and **P11**) and five females (**P1**, **P6**, **P8**, **P9**, and **P10**). This splits the group well between males and females, giving a fair comparison of both genders regarding their EV charging habits and platform usage. Such representation is crucial, given that there is substantial evidence that socio-demographic factors, such as gender, significantly affect EV ownership, charging preferences, and travel behavior. For example, according to women electric vehicle owners show different adoption and usage of electric vehicles, and therefore, their gender inclusion becomes very important in capturing comprehensive knowledge with respect to electric vehicle user behavior. In terms of experience as EV owners, participants ranged from newer owners with 1-2 years of experience to more seasoned owners with over four years of experience. Most participants were in either the (1–2) year category, or the over four-year category; only a few were in the (2-4) year bracket. For example, **P3**, **P5**, **P6**, and **P7** had experience of (1-2) years, while **P2**, **P4**, and **P9** had been using their EVs for more than four years. This strong variation in experience offers the possibility of reviewing how certain aspects of charging behaviour and the use of the platform may change due to increased familiarity and comfortability with EVs.

On the question of favourite charge places, most respondents said that they preferred to charge at home (**P1**, **P2**, **P4**, **P5**, **P6**, **P8**, and **P9**), revealing that they have a significant orientation toward convenience regarding home charging. Four others (**P3**, **P7**, **P10**, and **P11**) expressed a preference for public charging, indicative of a fair number of users dependent on the public charging infrastructure. This might be because they face restricted choices regarding home charging, or they simply have different usage habits. Finally, there is a great variety in participants' experiences concerning the use of online platforms for managing or investigating EV charging availability. Those who use online platforms on a regular basis are **P1**, **P3**, and **P11**; however, most only used online platforms once in a while (**P2**, **P4**, **P6**, **P8**, **P9**, and **P10**).

At the same time, **P5** and **P7** have no experience with online platforms, which flags another segment of users either without needs or interests in such kind of tool. Different use of the platform points at a range of digital engagement and perhaps different expectations on how EV owners' access and use the charging information. taken together, the participants in this study comprise a broad and representative sample that can be utilised to understand a wide perspective on the experiences of EV owners in driving EVs, their charging habits, and use of digital platforms to satisfy their needs, thus shedding light on varied behaviours and preferences among EV users.

9.1.2. Perceptions of Charging Accessibility and Acceptance of Digital Solutions

Rapid growth in the adoption of EVs has made charging infrastructure an important topic in talks on sustainable transportation. As more drivers are switching to electric mobility, their experiences with the options available for charging significantly influence their satisfaction and hence continued use of the vehicles. Experiences gained are of paramount importance to stakeholders interested in improving charging networks and developing supporting technologies. This section explores the perceptions of EV owners in terms of charging accessibility and their acceptance of digital solutions devised to complement the charging experience. Participants were asked to rate their level of their acceptance of the following statements in relation to EV charging using a five-point Likert-type scale:

- **Charging Practices**

"I am always keen to perform regular stops to charge my EV".

- **Charging Challenges During Long Trips**

"I always struggle to get a charging point to charge my EV during my long trips".

"If I don't plan ahead, I always have difficulty finding a suitable time to charge my electric car during long trips".

- **Experience with Digital Solutions**

"I feel comfortable using online platforms, if they exist, to direct me to the most suitable charging point".

These statements were prepared to assess the level of practical problems faced by EV owners and also their attitudes and readiness to use technology for overcoming these difficulties. An analysis of their responses provides insight into:

Frequency and convenience of recharging: This would imply the regularity and accessibility of charging stations for drivers, thus reflecting the existing condition of the charging infrastructure.

Planning behaviours: Rating the relative importance of planning a trip reflects the extent to which or not drivers need to plan ahead in order to facilitate their travel.

Acceptance of online platforms: Evaluating comfort levels with digital tools indicates the potential success of implementing technological solutions to improve charging accessibility.

The results from this analysis are illustrative of EV owners' charging infrastructure and technology adoption perceptions and experiences, offering insights for policymakers, charging network providers, and technology developers in understanding where they should intervene based on the problem perceptions amongst the respondents for improving the e-mobility charging experience. Addressing these challenges and embracing supportive digital solutions will help to move toward a better, smooth, friendly charging ecosystem, therefore triggering seamless growth in electric mobility.

9.1.3. Analysis of Results

In analysing the responses of the participants to the statements by means of a Likert scale, measures of central tendency (mean, median, and mode) of each statement were computed. The analysis and the calculation of the measures were performed in Qualtrics and Microsoft Excel, to identify the general tendencies and attitudes of participants in relation to EV charging habits, challenges, and comfort with digital solutions. Several studies have used the central tendency measure to analyse the Likert-scale data in qualitative research (Boone and Boone, 2012; Joshi *et al.*, 2015). Although Likert scales produce ordinal data, using quantitative summaries such as mean, median, and mode is quite an accepted practice that can be used to complement qualitative findings. This mixed-methods approach develops a more complete insight into participants' attitudes and perceptions.

9.1.3.1. Charging Habits

Statement: "*I am always keen to perform regular stops to charge my EV*".

Responses and frequencies: In Table 9-2, the responses are moderately variable, with an SD of about (1.23). While the majority of the respondents slightly agreed with the statement, there was a dispersed spread across the scale, showing different extents of keenness toward performing the regular charging stops.

Table 9-2 Frequencies of the Responses Options About the Charging Habits Statement

Response Option	Numerical Value	Frequency
Strongly Disagree	1	1
Somewhat Disagree	2	2
Neither Agree nor Disagree	3	1
Somewhat Agree	4	5
Strongly Agree	5	2
Total		11

Mean (μ) = 3.45
SD (σ) ≈ 1.23

9.1.3.2. Charging Challenges During Long Trips

Statement: "*I always struggle to get a charging point to charge my EV during my long trips*".

Responses and frequencies: In Table 9-3, the SD of about (1.15) indicates a moderate average variability; the response is somewhat closer to the mean. Most responses lie in the disagreement and somewhat disagreement regions, which means fewer struggles were faced in finding charging points during long trips.

Table 9-3 Frequencies of the Responses Options About the Charging Challenges During Long Trips Statement

Response Option	Numerical Value	Frequency
Strongly Disagree	1	3
Somewhat Disagree	2	4
Neither Agree nor Disagree	3	1
Somewhat Agree	4	3
Strongly Agree	5	0
Total		11

Mean (μ) = 2.36
SD (σ) ≈ 1.15

9.1.3.3. Planning and Difficulty Finding Charging Times

Statement: "*If I don't plan ahead, I always have difficulty finding a suitable time to charge my electric car during long trips*"

Responses and frequencies: In Table 9-4, the SD of about (0.96) indicates that less variability has occurred compared to the preceding statements. In fact, responses are closely clustered around the mean, and several of the participants somewhat agreed that not planning ahead results in some difficulties with charging on longer trips.

Table 9-4 Frequencies of the Responses Options About Planning and Difficulty Finding Charging Time Statement

Response Option	Numerical Value	Frequency
Strongly Disagree	1	1
Somewhat Disagree	2	1
Neither Agree nor Disagree	3	3
Somewhat Agree	4	6
Strongly Agree	5	0
Total		11

Mean (μ) = 3.27
SD (σ) ≈ 0.96

9.1.3.4. Comfort with Digital Solutions

Statement: "*I feel comfortable using online platforms, if they exist, to direct me to the most suitable charging point*".

Responses and frequencies: In Table 9-5, the SD of about (0.64), showing that responses are not very variable; responses are bunched together around the mean, which engendered a high consensus on the feeling of comfort in using online facilities to find charging points.

Table 9-5 Frequencies of the Responses Options About Comfort with Digital Solutions Statement

Response Option	Numerical Value	Frequency
Strongly Disagree	1	0
Somewhat Disagree	2	0
Neither Agree nor Disagree	3	1
Somewhat Agree	4	5
Strongly Agree	5	5
Total		11

Mean (μ) = 4.36
SD (σ) ≈ 0.64

9.2. Thematic Analysis of Qualitative Interviews

9.2.1. Thematic Analysis for the Assessed Scenarios

This section presents the thematic analysis of the data arising from interviews conducted with the 11 EV owners, aiming to explore users' beliefs concerning predictive models for EVCS occupancy, focusing on themes related to trust, data display, and the challenges associated with predictive accuracy. The thematic analysis followed the six-phase approach of Braun and Clarke (2006), ensuring a rigorous and systematic examination of the data.

9.2.1.1. Familiarizing with the Data

The analysis began with an in-depth study of the interview transcripts. This stage involved iterative reading to identify significant patterns and initial areas of interest. Perceptions of participants were recorded, namely those that demonstrated varying degrees of confidence in the predictive model, preferences for data presentation, and difficulties associated with EV charging. Furthermore, upon concluding this phase, similarities and differences among participants' opinions were emphasised, in order to adequately prepare for the subsequent phase.

9.2.1.2. Generating Initial Codes

The data was systematically coded, with codes such as "trust in predictions", "graphical preference", "planning importance", and "accuracy concerns" emerging. These codes served as the building blocks for the development of broader themes. Initially, the researcher generated as many codes as possible, then carefully sorted and reduced the codes to fully represent the participants' views and answers.

9.2.1.3. Searching for Themes

The initial codes were organized into potential themes. For example, codes related to trust and conditional acceptance of the model's predictions were grouped into the theme "Conditional Trust in Predictive Models". Similarly, codes on data representation preferences were grouped under "Preference for Data Representation".

9.2.1.4. Reviewing Themes

The themes were refined to ensure they accurately captured the data's essence. Some themes were merged, while others were further subdivided. For instance, the theme "Challenges with Predictive Accuracy" was refined to capture specific concerns about the model's ability to handle unexpected events.

9.2.1.5. Naming Themes

Three key themes were identified and defined, each of which has some sub-themes as discussed below.

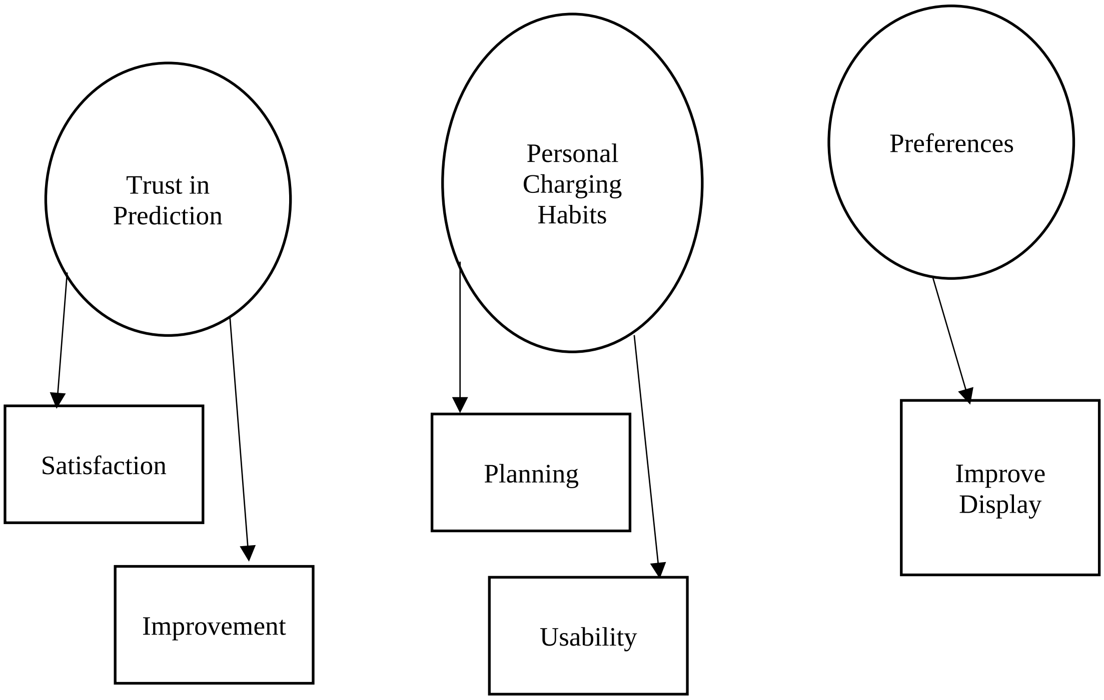


Figure 9-1 The Final thematic Map of The Performed Thematic Analysis

9.2.1.6. Illustrative Coding Process

To enhance the transparency of the thematic development process, this section presents illustrative examples of the initial codes and how they were grouped into the final themes. For instance, during the coding phase, separate codes such as "Trust in prediction output," "Skepticism toward long-term forecasts," and "Willingness to rely on imperfect data" were identified. These were later synthesized into the sub-themes High Level of Confidence, Partial Confidence, and Weak Confidence, under the broader theme Trust in Model Prediction. This grouping allowed differentiation between full, cautious, and minimal trust based on contextual reasoning provided by participants.

Similarly, initial codes like “Graphs show trends clearly,” “Text is easy to understand while driving,” and “Percentages help quick reading” were grouped under the theme Preference for Data Representation. These were then further categorized into sub-themes such as Graph Display, Text Display, and Categorical Display, depending on participants’ rationale for preferring each format.

In the theme Charging Habits, the initial codes “Routine trip planning,” “Unplanned charging difficulties,” and “Long journeys need preparation” were merged into sub-themes like Planned Stops and Unplanned Struggles to represent the contrast in experiences. This coding evolution ensured that participant experiences were faithfully represented while maintaining clarity in thematic structure.

A summarized view of these themes and their linked evidence is shown in Table 9-6 and mapped conceptually in Figure 9-1. The diagram is simplified for clarity but reflects the underlying thematic logic developed through this process.

9.2.2. Analysing Defined Themes

9.2.2.1. Charging Habits

In the second section on this interview, participants showed various perceptions towards regular charging habits, as alluded to in provided statements in section (9.1.2). During the interviews, researcher asked an open question to participants about the reasons for their agreement or disagreement with some statements about the charging habits in that section. About the regular stops for charging. **P1** agreed with the statement with regard to long journeys, and he considered that regular stops could be used for having a break as well as charging.

> “I guess somewhat agree. I like to do so, especially for the long journeys, for comfort. That was the same with my internal combustion car. I do not like to let it get down to 10%, so I top it up while I am in my break time. But I would not stop for the sake of it”. **[P1]**

Some participants expressed that they needed to perform some charging stops when necessary, but they did not consider this to be onerous. As **P2** mentioned, his experience with charging stops in long journeys always occurred with a smooth charging process and no untoward inconveniences. This was supported by **P3** and **P8**. On the other hand, **P5** recalled charging struggles in cases of travelling further than the battery capacity could handle, which pertains to unexpected extended journeys.

> "I would say I just stop when I need to charge. I do not plan any extra stops. So, yes, I never encountered any problem not like what we see in the press such queuing up. So, it is very occasionally. You know, we can get a rapid charger around if we really need to top up quickly". **[P2]**

> "Not really. I mean, somewhat disagree. It never has been a challenge. You know, there are options if it happens". **[P3]**

> "I am going to strongly disagree with that one [statement], to be honest. The majority of people will buy the EV with the battery to cover the majority of their trips. From my point of view, which was how I have found everybody else I know. Other people that get completely worried about charging tend to go for the biggest battery from doing short runs because all they bothered about is running out of juice". **[P8]**

> "It is very unlikely. Yes, public charging network is very poor. Though if I was travelling further than what my battery capacity was then, and if I had not planned, I would struggle". **[P5]**

9.2.2.1.1. Preference for Data Representation

This section further investigates the results in Section (8.2.4), which shows the result of display modes rating. Each participant in this interview asked to try the following scenarios.

Scenario 1: Participants were shown predictions for the occupancy state in an EVCS for a certain time in the following day. These predictions were produced by the chosen model from the comparison experiments in Chapter 6.

Scenario 2: Participants were shown a short-term prediction for the occupancy state in a certain charging point within the next six hours. The predictions, in this time, were shown in a form of a text expression.

Scenario 3: One of the previous scenarios was randomly chosen to be repeated, but this time the predictions were presented in a categorical description (Full, Moderate Full, Moderate Empty, Empty).

Each participant received the scenarios in a randomised ordering. In each scenario, Researcher asked participants the following questions:

- *What information you can get from this result, was this way of presenting predictions clear and simple?*
- *Is there any further information that you think of that should be included to make the display clearer and more understandable?*
- *What was your initial decision when you saw these predictions?*

Scenario 1: Chart Display

Participants found the predictions chart largely satisfactory in terms of simplicity. Almost all the interviewees found the graph easy to interpret and clear in informing them of predictions. Only one of the participants **[P2]** said they would have preferred more clarification to understand the information provided by the chart. Most of them also agreed about the importance of these predictions and the usefulness of havening them to plan their EVs charging times in advance. Despite this clear agreement on the efficiency of the chart in giving a clear picture of the forecasts produced, some argued that it is important to include more information to support the predictions' graph. For example, **P3** mentioned that adding information such as the type of the available charging plugs may support the presented predictions. The participant justified this by the fact that the faster the charger was, the shorter the time it would remain reserved.

> "I think prediction only is not enough. This model implies that as cars charge and leave new cars arrive at the same rate. I would like to see information about the charging type if rapid or fast". **[P3]**

> "When I plug it in, if I'm very low on power and on battery and I'm doing a long journey, the fast charger you know, is a maximum of 22K. It will take hours to charge the car. You would use a rapid charger if you could 50 or 100 kilowatts, and then the time would be less than an hour. So, 150-kilowatt rapid charge and my car would charge in about 40 minutes. So, what's missing here? How fast?" **[P3]**

> "The point is that, if this is a rapid charger, then a car would arrive at time zero and be fully charged at time one. Yeah. And it would leave, and the new car would arrive to take its place. The number is eight available spaces. But if this is only a fast charger, the six cars that are there at time zero could still be there at time two". **[P3]**

Similarly, **P5** believed that including details about the type of charging point enhances the value of the forecasts. Likewise, he considered that his decision might change, depending on the type of charger available.

> "So, knowing this kind of information, even though it is just a prediction, it is still valuable for my trip. But I believe that if this information is talking more about the charging plug type, whether it is rapid or slow, will be even more helpful". **[P5]**

> "You know, I might change my decision according to the type of the available plugs". **[P5]**

Accepting this viewpoint is possible, but it is important to remember that some people might continue to charge their cars even after their EV's battery is full, possibly due to other obligations like extended rest periods. Almost identical to the previous opinions, **P6** said that adding hints about how quickly and how much these predictions vary between time periods could be more useful.

> "If we know how fast this predictions changes by time. So, does it take long time to change? You know sometimes the numbers I mean the actual numbers here does not change sharply". **[P6]**

Scenario 2: Text Display

Textual display of the occupancy state predictions was not the preferred choice for most participants. In contrast, few have seen that textual presentation provides a simple and convenient means. **P2**, for example, directly preferred textual display of predictions over the chart.

> "Well, obviously the second one 'the text display' more straightforward and clearer so I think it is better". **[P2]**

Similarly, **P4** consider the text display mode better to express the occupancy state especially when the information provided for a short-term scenario while traveling.

> "Text! I think would be really useful. Or I think somewhat on balance. I would probably, for ease of use, prefer a text. Especially if you are using it while you are travelling, you cannot really look at a graph, but where if you get a text. The text is easier". **[P4]**

P8 added that the text also can be useful specially if supported by showing the percentages of the occupancy state instead of numbers.

> "I think informatively wise it would be a percentage of busyness. I would look at it so if it was (90%) busy, I would not go. And if it was (20%) busy, I would go. So, I would look on a percentage of availability rather than anything else on number of spaces". **[P8]**

Scenario 3: Categorical Display

In this section the aim was to explore the extent of which participants agree to see predictions as categorical level of occupancy rather than numbers. To achieve this, the researcher asked participants about their perception of seeing the forecasts of the occupied spaces in charging stations that take the form of categories rather than numerical values. Although most of participants were able to translate the categorical form of the predictions, some required support to clarify the categories associated with the different occupancy states. Most participants agreed that representing the forecasts in the form of numbers is better and clearer than representing them in the form of categories. The most prominent reason for this preference was that knowing the numbers gives a more accurate interpretation of the predictions, which helps in making their decision.

> "I personally prefer exact number because sometimes it might say like (quite empty or quite quiet), but maybe there is only three stations there. And then it completely fills. But if it was quite quiet and it was one with 18 stations. Then there was definitely going to be loads of space on there. So, I prefer the number". **[P1]**

> "I think predictions are predictions. Anyone uses prediction has to take it with 'a grain of salt'. So, I am happy with the numbers rather than the categories. Because it is a bit vague in terms of numbering. It will really help understanding how many numbers of slots are available in total or how much is the predictions saying". **[P3]**

> "I think the numbers better. From the numbers we know how busy, so numbers are ok". **[P6]**

P4 believed that knowing only the category of the availability state may not be useful if there are varied changes in a short period.

> "I think if it is possible, knowing numbers is better. I mean if you think it is a low occupancy state or a high occupancy state and you don't know how many charges are there if there were only four and it was a high occupancy rate, you might find you get there and there is nothing available. But if they were 14, you might find that there are three available. So, knowing an occupancy would not be so useful as knowing number of spaces". **[P4]**

P5 and **P9** considered that both representations are similar, and provide the necessary information, but they had a slight preference for presenting predictions in numbers.

> "There is not much difference between the value or category, I think. Both will be good to know about the station's occupancy. However, numbers can give better ideas to describe the case". **[P5]**

> "I think there is a slight difference between knowing a number of spaces available and knowing only how the occupancy state looks like. Both can be useful but, in my opinion, the more useful is to know how many spaces as a number". **[P9]**

P2 hinted that, if he saw the predictions in numbers, he could translate them into a category, but the opposite may not be true.

> "I prefer to see the actual data and make my own decision than to be told by somebody else that was medium busy because I do not really know things in their mind". **[P2]**

P2 added that categories might be more useful if they were accompanied by numbers when presenting forecast results. This was similar to the thought articulated by P10:

> "But maybe if you said how many numbers were at the station and then 'quite busy' you will know that this station quite busy". **[P2]**

> "Well, as I said numbers will work nice, but no problem if combined with its descriptive category so this may give supportive answer about the occupancy state". **[P10]**

Comparing Preferences for Chart or Textual Expression Display

This section reports on participants' responses when they were asked to decide which display mode in the two scenarios was more informative and clear (in terms of clearly expressing predictions). The aim mainly focuses on investigating the preferred display, whether text expression or graph display, for EV owners to see the predictions. Most participants in this study chose the chart presentation method. Some of them strongly preferred this method over textual expression, while others preferred it with some comments. P1, for example, clearly preferred the diagram over the textual presentation. He justified his choice by saying that the chart clearly gives an overview of the forecasts for a given period of time.

> “I prefer the graph. Maybe because it makes you think that you are having lots of availability at night, and then you know the overall sense of busyness at a glance”. **[P1]**

> “It reminds me about the ‘Tesla app’, they use a graph to show how busy their charges are, and it is really easy to see. You can see when the clear spikes are”. **[P1]**

Likewise, P10 definitely preferred a diagram to text, because it is clearer, as he describes it. He added that in the textual display, it is easy to lose some details about the predictions.

> “I think the graph show more information than text it is even easier. I would prefer to see graph. Because in the text form here you will be more likely to miss out some information”. **[P10]**

P7 and P11 indicated that individuals may like charts over text due to their clarity in displaying overarching trends and patterns.

> “What my answer is, it does not make a difference in this case particularly, because the data is simple, it shows ‘8,8,8,8’ then ‘9’. It is not a challenging task to interpreting the numbers. While if the numbers went up and down, ‘6,7,9,5’ then ‘4’, then the graph is a better way of explaining what is going on than a column of numbers. People hardly see a pattern in the numbers if it is just a column, whereas the graph will give a clearer picture of what is going on”. **[P7]**

> “I think graph better to show the possible trend and patterns. So, I think graph is better”. **[P11]**

From another point of view, preferring the chart is considered better in this case because it displays future predictions and not real data.

> “Not the numbers of course. Because it is just predictions. It is not real data. So, the graph better here”. **[P3]**

Again, the need to observe the overall trend of the occupancy state could be the justification of preferring a graph over text display specially when the need to the charge is less urgent. According to this principle, **P6** explained why he preferred the diagram to the text. It was considered that the chart was easier and quicker to compare between periods.

> "Either are ok for me, but you know the graph quite good and easier to find out the difference between the time slots in a certain time scale". **[P6]**

9.2.2.1.2. Trust in Model's Predictions

This theme covers the participants' answers on a scenario when they have been shown the predictions along with the actual recorded value in a past situation. Participants expressed varying levels of trust in the predictive model, often contingent on the situation. Trust was higher when the predictions aligned closely with participants' prior experiences or when the situation was not critical.

High Level of Confidence

Some of the participants accepted the model's level of accuracy, including accurate and inaccurate predictions, and considered them acceptable, which did not negatively affect their confidence in the model predictions.

> "Well, the predictions not too far and at some points are similar. So yes, I will stay with the same decision in both scenarios". **[P1]**

> "Well, I think the model is consistently a bit more pessimistic, isn't it? So yeah, but it's still telling me that there's availability. So, I would trust the model because let us think of the worst scenario, I would go there, and I would find that all the charges were occupied. This will be similar to what will be if I ignored the predictions". **[P2]**

> "The difference is quite large. But happily, accept the model as because it's better than other cases". **[P7]**

> "I already know the system gives a prediction. So yes, I think even there are some points predicted incorrect, but it is acceptable as long as we are aware that it is predictions". **[P9]**

> "Yes, as I can see from this example, there is not much difference between the actual and predicted value. I think the algorithm somewhere helps to predict the available spaces. So yes, I will be happy to take what given to me". **[P10]**

All of the participants quoted above expressed satisfaction with the predictions generated based on the scenarios they encountered. The predictions were clearly acceptable as a means to help them make their decision to search for charging places for their EVs. Despite some apparent deviations when comparing the forecast chart to the actual value chart, they considered it fairly acceptable.

Partly Confident

Other participants expressed less confidence in the model's predictions. Some of them said that they would accept these predictions, but with caution, and additional evaluation in some circumstances (P11). Some others said that they might accept these predictions in certain situations, and may reject them in others, depending on when the predictions are needed (P3, P4), or the extent of the need to charge the car battery (**P4**, **P6**). Also, P5 stated that he would prefer to combine what the model offers with his prior experience with charging stations to make the best decision.

> "As I can see, there are some inaccurate predictions at some points within the time slot. But I think it's still reasonable. So, I will stay with the same decision for the first scenario, while in the second scenario I may try to find somewhere else". **[P3]**

> "As I can see, the model exactly predicted two of the four hours which is ok. The other two hours are slightly different, but I think both still close to the correct value. So, I would say depend on the situation or how really need to charge. If the trip just for fun with no rush, I will take it as a plan and go for it based on what the model predicted". **[P4]**

> "Again, here I will use both what the model saying and what my own experience with charging my car telling me". **[P5]**

> "Yes, I think that graph shows how close these are predictions to the real-life rates. I take these predictions into account, but I would say the predictions still predictions, I may take them with care. I mean it depends on the situation. If there is any risk, I would prefer to rely on the live data from there at that time instead on using these predictions". **[P6]**

> "I will follow the predictions that were given to me, but there are times when I may need to think a lot about them before I use them. You are well aware that these are only guesses and that you could be wrong". **[P11]**

Weak Confidence

On the other hand, P8 expressed a lack of confidence in the predictions' outcomes. The model's inability to predict unexpected events near charging stations was evident. P8 said that he prefers to rely on direct, real numbers available about the capacity and status of nearby charging stations for some electric car technologies.

> "I would not trust the model. I would go with what my car tells me… So, where the modelling can be vastly different on that day because I don't know, there's been a motorway crash and lonely people have been diverted and they're thinking well, while I'm waiting, I might as well go and have a charge. So, everybody descends on the local charging points, floods them out an unexpected time and your predictions go through the roof". **[P8]**

Unexpected events, particularly those that occur close to EVCSs can actually affect EV owners, as P8 indicated. This is an explicit reason that may directly affect the accuracy of the resulting predictions. It is important to find a way to improve predictions in these cases, but we must realize that, in general, the percentage of electric vehicles on the road is lower compared to conventional ones, and this smaller percentage may mean that most of them do not need to charge their batteries at that time. Consequently, it is possible that these events will have less impact on charging stations. However, finding a way to deal with these events remains extremely important to increase confidence in the resulting predictions and prepare for future trends with more EV uptake. In order to include the study of expected and unexpected external events affecting forecasts, it is necessary to find and providing direct data sources with continuous updates, which may be extremely difficult at the present time.

9.2.2.2. Summary and Map of Thematic Analysis

Table 9-6 combines the recognised themes together with corresponding quotes from participants, providing a thorough comprehension of their perspectives on the predictive model. The investigation shows a diverse although mostly positive attitude towards predictive models for EV charging, moderated by concerns regarding reliability and the demand for improved data representation. **Error! Reference source not found.** displays the final thematic map that visually represents the relationships between the identified themes: It identifies the three overarching themes, and how they connect with their sub themes. It also highlights the role of planning in their EV charging experience and their concerns regarding the model's predictive accuracy.

Table 9-6 Summary of Thematic Analysis

Theme/subtheme	Description	Participant Excerpt
Charging Habits		
Planned **Stops**	Charging stops are often planned for long journeys and are seen as non-intrusive.	"I plan stops for comfort during long journeys". **[P1]**
Unplanned Struggles	Struggles arise primarily during unplanned or extended journeys.	"Struggles arise only when the trip exceeds the battery range". **[P5]**
Preference for Data Representation		
Graph **Display**	Preferred for clarity and ability to show trends over time.	"Graphs provide an overall sense of busyness at a glance". **[P1]**, **[P10]**
Text **Display**	Preferred by some for simplicity, particularly in short-term scenarios or while traveling.	"Text is easier to interpret on the go". **[P2]**, **[P4]**
Categorical Display	Less favoured; participants emphasized the need for precise numerical data.	"Numbers are clearer and help in decision-making compared to categories". **[P1]**, **[P3]**
Trust in Model Prediction		
High Level of Confidence	Participants accepted predictions, even with some inaccuracies, finding them helpful for decision-making.	"The predictions are close and acceptable as a means to help make decisions". **[P1]**, **[P2]**, **[P10]**
Partial **Confidence**	Participants used predictions cautiously, combining them with prior knowledge or considering situational urgency.	"Predictions are useful, but I would also consider live data or the situation's urgency". **[P3]**, **[P6]**
Weak **Confidence**	Some expressed distrust, particularly due to the model's inability to account for unexpected events.	"Unexpected events can make predictions unreliable". **[P8]**
Unexpected Events	Real-world variables, like accidents or demand surges, undermine prediction reliability.	"Predictions can't account for all real-world variables". **[P8]**
Lack of Contextual Information	Adding data about charging plug types and speeds could enhance decision-making.	"Knowing plug types would help refine predictions and decisions". **[P3]**, **[P5]**

9.3. Discussion

9.3.1. Charging Accessibility and Acceptance of Digital Solutions

The SDs of the participants' answers reveal substantial insights into their perceptions and experiences with EV charging. Overall, charging habits and challenges were at a medium level of variation; with an SD of about (1.15) to (1.23), participants reported different experiences and attitudes when it comes to charging on long trips. This diversity suggests that while the inconveniences of regular charging stops are minimal for some, regarding other EV owners, access to charging points is just not that easy, reflecting a wide range of experiences within the EV community.

In contrast, there was lower variability regarding the importance of planning ahead, for which the SD was about (0.96), demonstrating greater consistency among participants. Many EV owners understand that a lack of planning leads to challenges in searching for suitable times of charging when on long drives. This consensus underlines the fact that trip planning is significant in allowing EV drivers to have smooth experiences on the road, free from many stressors.

The lowest variability again was about participants' comfort in relation to digital solutions; the SD of about (0.64) is a good fit between strong agreement and high comfort in using online platforms for one's own charging needs. This shows that a general openness and readiness can be observed in the current EV owners for adopting technological tools that may provide convenience for their charging experience, like apps or online platforms showing real-time information about charging station availability.

A notable moderate variability in both charging habits and challenges; because of specific experiences and needs of different user groups, an intervention should be directed toward them. Targeted interventions can be undertaken to respond to unique challenges emanating from the different segments of EV owners with more personalized support and resources.

Since there is unity on the role of planning, this is an opportunity to move ahead with the planning tools that would facilitate preparations for the journey. This research consequently calls for creation and refinement of applications that would be used in route planning by scheduling stops for charging. Such tools could reduce some anxiety about taking long trips and further optimize charging.

The high levels of comfort with the digital solution and low variability suggest that it is a potential strategy to tap into digital comfort. Bringing online platforms for charge assistance or improving the existing ones will be acceptable for the majority of EV owners. Such online platforms can include functions such as the forecasting of filling station congestion, routing towards the closest point of availability, and even booking possibilities for reserving charging slots.

The calculation of the SD for survey responses added value to the understanding of the dispersion of the answers given by the participants, and gave deeper insights into their experiences and perceptions. With the wide variety of experiences of EV owners, especially regarding charging habits and challenges faced during longer journeys, a range of different solutions must be developed and tailored for specific needs. This all points to great potential for the related technologies to be implemented in a very successful way. These findings can serve to assist stakeholders like policymakers, charging infrastructure providers, and technology developers in prioritizing their work to address this variation in experiences. Stakeholders can then manage in ways that complement the overall EV ecosystem by focusing more on those areas of disagreement in experiences and those areas of strong agreement. Infrastructure improvements, active facilitation of plans using effective tools, and better use of digital platforms will in themselves create a more functional and seamless charging experience for all users of EVs.

9.3.2. The Qualitative User Study

In the first section of this chapter, the sample of the qualitative study presented a well demographic distribution of 11 EV owners. Although most of the interviewed individuals in this study most frequently used their home for charging, they still had experience with public charging, especially for long journeys. The majority of the sample had experience with online platforms, whether on a regular basis or for limited periods. This diversity may serve this study with its variety of different points of view pertaining to three major themes arising from the qualitative user study: “charging habits”, “preference for data representation”, and “confidence in model predictions”. The themes are also indicative of the trends in users’ behaviours, preferences, and expectations concerning EVCI and any predictive models on the availability of the charging stations.

9.3.2.1. Charging Habits

Participants exhibited diverse charging habits influenced by factors such as convenience, infrastructure availability, and trip purpose. Most participants preferred charging their EVs at home due to its convenience, aligning with findings in Morrissey *et al*. (2016), who identified home charging as the dominant preference among EV owners. However, for those without access to home charging or undertaking long journeys, reliance on public charging infrastructure was essential. Most of the respondents declared their intentions for stops that were meant to rest during long trips, showing a wider travelling behaviour in which charging was combined.

These behaviours confirm and extend the results derived by Lucas *et al*. (2019), about the optimization of satisfaction while traveling. Conversely, individuals who faced difficulties related to charging during unforeseen or prolonged journeys emphasized the shortcomings of existing infrastructure, especially in regions characterized by a scarcity of public charging stations. These observations align with research that underscores the disproportionate availability of charging facilities as an obstacle to the smooth integration of EVs (Bhagavathy and McCulloch, 2020).

Such differences in charging behaviour indicate the intense need for expanding and spreading charging options, especially in underserved communities. Real-time availability information and the inclusion of charging station suggestions in trip-planning tools could also ease a portion of range anxiety related to longer trips and unexpected stops to charge.

9.3.2.2. Interface Preference

The research indicated a strong trend towards graphical data presentation, due to clarity and ease of understanding of temporal trends being highly appreciated by respondents. Graphical formats, especially charts, were highly valued, since there is a possibility to present information on charging station availability so as to support user decisions. These findings support the results of Qiao and Lin (2021), on the role of graphical tools in enhancing the user experience of EVs users. Despite the preference toward graphical information, some participants stated a preference for text-based messages in very near, short-term contexts-especially while traveling.

Overall, participants were of the view that categorical messages, such as “Full” or “Moderate”, were not informative, thus they underscored the need for precise numerical messages underlying more subtle decision-making. Hybrid methods, which put together textual, graphical, and categorical representations, may thus be most effective in the light of diverse preferences. There were also suggestions from the participants for extra contextual information, like charging speed and plug type, to be integrated into the predictive tool.

Importantly, insights from the user study directly informed the development of the predictive modelling component in this research. Specifically, the feedback gathered regarding users’ preferred planning horizons and decision-making needs led to the selection of a short-term prediction window of up to 6 hours. This time frame was chosen as it aligns with users' real-world behaviours and expectations for EV charging availability, particularly among those who indicated a preference for same-day or near-future availability forecasts. Integrating this user-driven input ensured that the model remained not only technically sound but also practically relevant and user-centred in its design and output.

9.3.2.3. Confidence in Model Predictions

Levels of trust in predictive models varied widely among respondents. Those who trusted the models highly viewed predictions as generally reliable, even allowing for minor errors if the models were useful to inform decisions. Users showing conditional trust raised several worries regarding the ability of the model to perform in unforeseeable circumstances, like a traffic jam or an accident close to an EV charging point. These kinds of end-users often use the forecasts for supportive means, together with their experience and current measurements.

Correspondingly, Douaidi *et al.* (2023) noted that such predictive models enhance reliability and user assurance if they are complemented with real-time information and exogenous variables. Those showing low trust emphasized how necessary it was to get the latest updates, with transparency of the model forecasts. They revealed a positive attitude toward tools that can adapt to changing conditions and provide actionable insights matched to the current status. Such concerns may be overcome effectively by integrating federated learning methods, as Douaidi *et al*. (2023) suggested, which enable predictive capabilities that are private yet run in real-time.

The present qualitative research work offers rich insights into the charging behaviour, preferences, and trust of EV owners. With regard to the variety in user needs, there is a need for an extension and optimization of the charging infrastructure, hybrid, and customizable data representation format, and increasing the adaptability and transparency of predictive models. These can help improve functionalities for predictive tools to create confidence in their output for seamless facilitation of EV charging and the wider acceptance of electric mobility. It also provides practical recommendations to policymakers, technology developers, and infrastructure providers by drawing relationships with existing literature and participant responses.

Chapter 10: Conclusions

10.1. Main Outcomes

The main aim of the thesis was to answer the research questions listed in Chapter 1, querying the performance of DL techniques to predict the occupancy state of the EVCSs, and investigates the common user charging behaviour of the EV owners and their thoughts about charging occupancy predictive models. It was found in the literature that ML and DL methods have been previously showed a strong performance in variety of prediction task, but less focus has been devoted to the field of predicting the occupancy state of public EVCSs. In order to more fully answer these questions, it was important to assess a number of tasks, including to:

- Analyse charging process of EV owners at charging stations.
- Determine the most common charging patterns among the EV owners.
- Evaluate the performance of different deep learning methods in predicting the occupancy state.
- Test the impact of input data on the model generalisation.
- Understand the possible impacts of the independent variables on the dependent variables, and how they correlated with each other.
- Evaluate users' experiences with occupancy predictive models.
- Assess users' usability and trust in the model predictions.

In conjunction with great technological and information developments, Internet services, and mobile phone applications in societies, smart projects and machine learning have played an increasingly major role in providing various solutions for many different fields, including EVs. In the literature review of this thesis, an indication was presented of the major role played by deep learning algorithms in providing accurate predictions in different fields. Therefore, this research focused on the characteristics of DL, paying attention to CNN and RNN; the main difference in these neural networks pertain to their architecture and core design, operational mechanism, and applications. Usually, in practice, hybrid models like CNN-RNN architectures are used to leverage the strengths of both. Understanding these differences helps in selecting the appropriate architecture based on the nature of the data and the problem at hand.

10.2. Contributions

The introduction of this thesis presented a series of research questions that this thesis attempted to answer. The main questions were: “1. How accurately can DLMs predict the availability of EVCSs based on historical data?” and “What display expectations do EV owners prefer to see for EVCS occupancy prediction?”. Therefore, the main contribution of this thesis is a new classification structure that takes its inspiration from the hybrid combination of CNN and RNN approaches, which has been investigated in terms of performance in training and testing states, classification report, and its ability to generalising. This research understood a mixed-method investigation of user experiences via quantitative and qualitative methods dedicated to analysing user perceptions toward predictive models, addressing the serious dearth of literature exploring EV users’ perceptions. This thesis therefore investigates the gap in understanding the attitudes of EV users and enhances predictive modelling with respect to EVCS accessibility. Accordingly, the main contributions of this research are identified as follows.

Introduction of a New Classification Model

A core contribution of the current thesis is to introduce a new model, BiGTCN. This new class was inspired by the joined power of both CNN and RNN, and was analysed extensively regarding its performance during training and testing, by classification and in relation to generalization abilities. The proposed architecture improves the accuracy of occupancy prediction, hence providing a comprehensive framework for modelling temporal dependencies in the data of EVCSs. Unlike the work presented by Hecht et al. (2021), which relied on binary occupancy features, this research incorporates predefined occupancy classes to train the BiGTCN model, enabling finer-grained predictions.

Occupancy Prediction Using Deep Learning Models

The studied DLMs were evaluated to answer **RQ1**, as presented in Chapters 6 and 7. This entailed extensive experiments to empirically test and then evaluate models with historical data to enable them to predict the future occupancy status of an EVCS. The comparative study of models' performance supports the identification of those DL models best able to predict occupancy with the best accuracy, furthering current insights into this domain. This study, therefore, balances model performance with generalization compared to models in the literature such as Soldan et al. (2021), which leveraged big data streaming for real-time occupancy predictions.

Important Time and Environmental Variables Identification

In order to answer **RQ2-1**, an in-depth data analysis was performed (as presented in Chapter 5), including correlation studies and statistical tests to identify specific temporal and environmental variables that contribute significantly to the models' predictive power. Feature extraction experiments were carried out in order to quantify the relevance of various input features, which enhances the quality and relevance of the training dataset.

Impact of Data Aggregation on Model Generalization and Accuracy

For **RQ2-2**, the thesis investigated how the use of aggregated data from multiple locations versus location-specific data in relation to impacts on the generalization and accuracy of DL models (Chapter 6). Unlike the majority of studies discussed in the literature, experiments under different data formatting scenarios compared models trained on location-specific datasets against those trained on combined datasets, providing insights into the trade-offs between model generalizability and accuracy.

Occupancy Prediction (Classification vs. Regression)

In addressing **RQ3**, this thesis explores occupancy prediction as both a classification and regression task. Unlike Ma and Faye (2022), who primarily focused on classification without external contextual features, this study demonstrates the advantages and limitations of each approach. By comparing accuracy and error metrics across tasks, the findings underscore the suitability of classification for high-variability occupancy states while highlighting the potential of hybrid models for enhanced performance, as suggested by Chen et al. (2022).

Understanding Occupancy Display Expectations for EV Owners

RQ4 involved the conduct of a quantitative survey (chapter 8) in order to understand the charge behaviour of electric vehicle owners and evaluate their expectations on how best to present occupancy predictions. This study presented some quite fundamental user insights in developing user interfaces and ways of presenting information in electric vehicle charging applications.

Investigation of Factors of Trust in Predictive Models:

Qualitative user study (chapter 9) for **RQ5** were conducted to investigate in greater detail the factors that influence the level of trust of EV owners in predictive models for effective management of their charging times. This study underlined the preceding quantitative survey and gave a detailed explanation of user trust and acceptance of predictive technologies. Unlike studies such as Douaidi et al. (2023) that emphasize federated learning for privacy preservation but neglect user experience, this research incorporates user perceptions to inform display strategies. The findings align with the growing need for user-cantered designs in EVCS applications, addressing factors such as trust and usability.

All these, put together, enrich the understanding of predictive modelling regarding occupancy with respect to electric vehicle charging stations and add practical insight into the perceptions and expectations of users, which are needed for the adoption and their effective use.

10.3. Future Research Directions

Based on the findings presented in this thesis, several possible directions for further research are outlined to extend both predictive modelling capabilities and user interaction with electric vehicle charging infrastructure. These directions address current limitations and explore opportunities for practical and theoretical advancements.

Real-Time Integration and Dynamic Updates: While this research focuses on historical data for predictive modelling, future studies could incorporate real-time data streams such as current traffic conditions, live charging station occupancy, and dynamic pricing updates. Real-time integration can significantly enhance the responsiveness and precision of the models, allowing them to adapt to sudden fluctuations in demand. For instance, events like unexpected traffic congestion or changes in pricing schemes could be captured and incorporated dynamically, ensuring users receive accurate and up-to-date availability information. Advances in edge computing and IoT-based sensor networks could enable such integration while maintaining computational efficiency.

Increased Predictive Features and Contextual Data: To build more comprehensive and robust predictive models, future work should explore the inclusion of additional data sources, such as:

- Special Events: Incorporating information on local events, holidays, or festivals that can cause sudden surges in demand for EV charging stations.
- Socio-Economic Data: User demographics and economic indicators can provide deeper insights into usage patterns, as charging behavior often varies based on socio-economic contexts.
- Environmental Factors: Incorporating real-time weather patterns, such as temperature and precipitation, could further improve predictions, as adverse weather conditions influence travel and charging habits.

These additional features, combined with advanced feature engineering and selection techniques such as mutual information and attention mechanisms, would create richer model inputs. This could ultimately enhance the models' ability to predict complex patterns.

Transfer Learning for Broader Generalization: The findings of this thesis highlight challenges in balancing generalization and location-specific performance. Future studies could employ transfer learning techniques to train predictive models on data from one region and adapt them to other geographic locations. By fine-tuning pretrained models with smaller location-specific datasets, researchers can improve generalizability without sacrificing accuracy. This approach could address the aggregate versus location-based trade-off, making predictive models applicable across diverse environments. Furthermore, cross-domain studies that include urban, suburban, and rural charging infrastructure could provide valuable insights into region-specific behaviours and demand patterns.

Advanced Model Architectures and Explainability: Building upon the proposed BiGTCN model, future work could explore more sophisticated hybrid architectures that integrate additional deep learning components, such as transformer-based models or graph neural networks (GNNs). These architectures could improve the models' ability to capture spatial-temporal dependencies in more complex datasets. In addition, increasing the transparency of deep learning models through explainable AI (XAI) techniques would make predictions more interpretable for stakeholders, such as policymakers, EV operators, and users. For instance, attention-based visualizations could highlight which features most strongly influence predictions, improving trust and adoption of the predictive system.

User Interface Design and Usability Evaluation: The insights gained in this thesis regarding display expectations provide a foundation for designing user-centric interfaces. Future research should focus on:

- Prototype Development and Testing: Creating functional prototypes of interfaces that deliver occupancy predictions in real time, using visual elements like bar charts, heat maps, or notifications based on user preferences.

- Usability Studies: Conducting user-cantered evaluations to assess interface effectiveness, satisfaction, and accessibility. Metrics such as task success rate, time-to-information, and user satisfaction scores could be used to validate design decisions.

- Incorporating Feedback Loops: Real-time user feedback mechanisms can help improve prediction accuracy by capturing actual charging behaviours and preferences, contributing to a more interactive and adaptive system.

10.4. Summary

This thesis provided significant contributions to answering the research questions outlined at the start of the study. For RQ1, deep learning models demonstrated varying performance levels across different phases and scenarios. The TCN emerged as the strongest performer, consistently excelling in handling complex temporal patterns, particularly in combined datasets where data heterogeneity is more pronounced. The Ensemble model (1D-CNN + TCN) also performed well, capturing subtle variations in occupancy trends. The introduction of the novel BiGTCN structure further improved prediction accuracy by leveraging the strengths of both TCN and RNN architectures, making it an effective hybrid solution. In Phase 2, 1D-CNN and the Ensemble model (LSTM + TCN) performed best, particularly in location-specific scenarios where localized accuracy was essential.

In addressing RQ2.1, the inclusion of average occupancy as a feature proved critical in improving model predictions. By providing a benchmark of normal occupancy trends, this feature allowed the models to better capture deviations in usage patterns, particularly during high and low occupancy periods. The findings demonstrated that incorporating average occupancy improves forecast accuracy during peak demand. The use of this feature helped stabilize predictions across different temporal and spatial contexts, making the models more robust and adaptable.

For RQ2.2, the study compared the performance of models trained on location-specific datasets against those trained on combined datasets. Location-specific training yielded higher accuracy and lower loss rates, as the models captured localized patterns and unique trends, particularly for Locations 1 and 2. However, models trained on the combined dataset exhibited stronger generalization capabilities across multiple locations, despite slightly lower localized accuracy. The TCN model, in particular, showcased superior generalization performance, as highlighted in the confusion matrix results. While location-specific training reduced errors for local predictions, combining datasets improved the models' ability to handle varied patterns, providing a trade-off between precision and broader applicability.

For RQ3, the results showed that framing occupancy prediction as a classification task was more effective than treating it as a regression task. Predicting occupancy as classes was more suitable due to the high variability and weak patterns in the total occupancy data, which made precise numerical predictions challenging. Classification models proved to be more reliable, offering greater accuracy and consistency for the research objectives.

In response to RQ4, user preferences for display formats were evaluated through a quantitative survey. The findings revealed that Discrete1 (bar charts) and Continue1 (density plots) were the most preferred display options due to their simplicity, clarity, and ease of understanding. Although other forms, like the mixed graph, obtained polarised reactions, they nonetheless fulfilled specific user requirements efficiently. These findings highlight the necessity of creating predictive interfaces that emphasise usability and demographic preferences to improve user pleasure and engagement.

Finally, RQ5 highlighted the factors influencing EV owners' trust in predictive models. Trust levels varied among participants, with highly trusting users finding predictions reliable and helpful despite minor errors. Those exhibiting conditional trust emphasized the need for model adaptability during unforeseen circumstances, such as traffic congestion. Participants with lower trust levels stressed the importance of real-time updates, transparency, and actionable insights tailored to current conditions. Addressing these concerns through federated learning approaches and user-adaptive tools could significantly enhance trustworthiness and confidence in predictive models, ensuring their practical utility for EV charging infrastructure.

Overall, this research advances the understanding of predictive modelling for EV charging station availability while addressing user expectations and trust. It provides practical insights for improving model performance, user interface design, and general applicability, contributing to more efficient and user-centric EV infrastructure management.

References

Agbro, D.E., Asthana, A. and Last, G. 2021. The Future of Hybrid Electric Vehicles and Sustainable Vehicles in the UK. In: Mporas, I., Kourtessis, P., Al-Habaibeh, A., Asthana, A., Vukovic, V., and Senior, J. eds. *Energy and Sustainable Futures*. Springer Proceedings in Energy. Cham: Springer International Publishing, pp. 213–221. Available at: https://link.springer.com/10.1007/978-3-030-63916-7_27 [Accessed: 14 October 2024].

Akshay, K.C., Grace, G.H., Gunasekaran, K. and Samikannu, R. 2024. Power consumption prediction for electric vehicle charging stations and forecasting income. *Scientific Reports* 14(1), p. 6497. doi: 10.1038/s41598-024-56507-2.

Aktas, M., Yilmaz, M.B. and Karabacak, A. 2023. An Experimental Study Investigating The Effects of Data Pre-processing Methods on Deep Learning Based Time Series Data Prediction. Available at: https://www.preprints.org/manuscript/202310.1635/v1 [Accessed: 7 June 2024].

Alface, G., Ferreira, J.C. and Pereira, R. 2019. Electric Vehicle Charging Process and Parking Guidance App. *Energies* 12(11), p. 2123. doi: 10.3390/en12112123.

Ali, Z. and Bhaskar, S.B. 2016. Basic statistical tools in research and data analysis. *Indian Journal of Anaesthesia* 60(9), pp. 662–669. doi: 10.4103/0019-5049.190623.

Almaghrebi, A., Aljuheshi, F., Rafaie, M., James, K. and Alahmad, M. 2020. Data-Driven Charging Demand Prediction at Public Charging Stations Using Supervised Machine Learning Regression Methods. *Energies* 13(16), p. 4231. doi: 10.3390/en13164231.

Amara-Ouali, Y. et al. 2023. Forecasting Electric Vehicle Charging Station Occupancy: Smarter Mobility Data Challenge. Available at: https://arxiv.org/abs/2306.06142 [Accessed: 10 October 2024].

Anderson, J.E., Bergfeld, M., Nguyen, D.M. and Steck, F. 2023. Real-world charging behavior and preferences of electric vehicles users in Germany. *International Journal of Sustainable Transportation* 17(9), pp. 1032–1046. doi: 10.1080/15568318.2022.2147041.

Becerra-Rico, J., Aceves-Fernández, M.A., Esquivel-Escalante, K. and Pedraza-Ortega, J.C. 2020. Airborne particle pollution predictive model using Gated Recurrent Unit (GRU) deep neural networks. *Earth Science Informatics* 13(3), pp. 821–834. doi: 10.1007/s12145-020-00462-9.

Bengio, Y., Courville, A. and Vincent, P. 2013. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8), pp. 1798–1828. doi: 10.1109/TPAMI.2013.50.

Bhagavathy, S.M. and McCulloch, M. 2020. Electric Vehicle transition in the UK. Available at: http://arxiv.org/abs/2007.03745 [Accessed: 14 October 2024].

Bharath Reddy, S., Nookaraju, Y., Kiran Goud, S., Usman, S., Sandeep, B., Madhavi, K. and Viktorovna Nemova, D. 2024. ECharge – An Electric Vehicle Charging Station Finder Application. Satyanarayana, K., Bobba, P. B., Perveen, A., and Debnath, S. eds. *MATEC Web of Conferences* 392, p. 01079. doi: 10.1051/matecconf/202439201079.

Bikcora, C., Refa, N., Verheijen, L. and Weiland, S. 2016. Prediction of availability and charging rate at charging stations for electric vehicles. In: *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. pp. 1–6. doi: 10.1109/PMAPS.2016.7764216.

Boone, H. and Boone, D. 2012. Analyzing Likert Data. *Journal of Extension* 50(2). Available at: https://tigerprints.clemson.edu/joe/vol50/iss2/48/ [Accessed: 15 November 2024].

Braun, V. and Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3(2), pp. 77–101. doi: 10.1191/1478088706qp063oa.

Cahour, B., Nguyen, C., Forzy, J.-F. and Licoppe, C. 2012. Using an electric car: a situated, instrumented and emotional activity. In: *Proceedings of the 30th European Conference on Cognitive Ergonomics*. Edinburgh United Kingdom: ACM, pp. 22–28. Available at: https://dl.acm.org/doi/10.1145/2448136.2448142 [Accessed: 15 October 2024].

Casolaro, A., Capone, V., Iannuzzo, G. and Camastra, F. 2023. Deep Learning for Time Series Forecasting: Advances and Open Problems. *Information* 14(11), p. 598. doi: 10.3390/info14110598.

Chang, M., Bae, S., Cha, G. and Yoo, J. 2021. Aggregated Electric Vehicle Fast-Charging Power Demand Analysis and Forecast Based on LSTM Neural Network. *Sustainability* 13(24), p. 13783. doi: 10.3390/su132413783.

Chen, Q., Liu, S., Qu, H., Zhu, R. and You, L. 2022. TWAFR-GRU: An Integrated Model for Real-time Charging Station Occupancy Prediction. In: *2022 IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles (SmartWorld/UIC/ScalCom/DigitalTwin/PriComp/Meta)*. Haikou, China: IEEE, pp. 1611–1618. Available at: https://ieeexplore.ieee.org/document/10189531/ [Accessed: 11 October 2024].

Chen, T., Zhang, X.-P., Wang, J., Li, J., Wu, C., Hu, M. and Bian, H. 2020. A Review on Electric Vehicle Charging Infrastructure Development in the UK. *Journal of Modern Power Systems and Clean Energy* 8(2), pp. 193–205. doi: 10.35833/MPCE.2018.000374.

Chicco, D., Warrens, M.J. and Jurman, G. 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science* 7, p. e623. doi: 10.7717/peerj-cs.623.

Chung, Y.-W., Khaki, B., Li, T., Chu, C. and Gadh, R. 2019. Ensemble machine learning-based algorithm for electric vehicle user behavior prediction. *Applied Energy* 254, p. 113732. doi: 10.1016/j.apenergy.2019.113732.

Dastpak, M., Errico, F., Jabali, O. and Malucelli, F. 2023. Dynamic Routing for the Electric Vehicle Shortest Path Problem with Charging Station Occupancy Information. Available at: https://arxiv.org/abs/2305.11773 [Accessed: 10 October 2024].

De Cauwer, C., Van Mierlo, J. and Coosemans, T. 2015. Energy Consumption Prediction for Electric Vehicles Based on Real-World Data. *Energies* 8(8), pp. 8573–8593. doi: 10.3390/en8088573.

Dong, L., Chen, M., Li, J. and Wang, Y. 2018. Time and Space Distribution Prediction of Electric Vehicle Charging Load Based on Statistical Analysis. *DEStech Transactions on Computer Science and Engineering* (cmee). Available at: http://dpi-journals.com/index.php/dtcse/article/view/20023 [Accessed: 8 November 2024].

Douaidi, L., Senouci, S.-M., El Korbi, I. and Harrou, F. 2023. Predicting Electric Vehicle Charging Stations Occupancy: A Federated Deep Learning Framework. In: *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*. Florence, Italy: IEEE, pp. 1–5. Available at: https://ieeexplore.ieee.org/document/10199832/ [Accessed: 10 October 2024].

Du, S. et al. 2023. An enhanced prediction framework for coalbed methane production incorporating deep learning and transfer learning. *Energy* 282, p. 128877. doi: 10.1016/j.energy.2023.128877.

El Hafdaoui, H., El Alaoui, H., Mahidat, S., El Harmouzi, Z. and Khallaayoun, A. 2023. Impact of Hot Arid Climate on Optimal Placement of Electric Vehicle Charging Stations. *Energies* 16(2), p. 753. doi: 10.3390/en16020753.

Elton-Chalcraft, S., Hansen, A. and Twiselton, S. 2008. *Doing Classroom Research: a Step-By-Step Guide for Student Teachers*. McGraw-Hill Education (UK).

Fan, J., Zhang, K., Huang, Y., Zhu, Y. and Chen, B. 2023. Parallel spatio-temporal attention-based TCN for multivariate time series prediction. *Neural Computing and Applications* 35(18), pp. 13109–13118. doi: 10.1007/s00521-021-05958-z.

Faraz, A., Ambikapathy, A., Thangavel, S., Logavani, K. and Arun Prasad, G. 2021. Battery Electric Vehicles (BEVs). In: Patel, N., Bhoi, A. K., Padmanaban, S., and Holm-Nielsen, J. B. eds. *Electric Vehicles*. Green Energy and Technology. Singapore: Springer Singapore, pp. 137–160. Available at: http://link.springer.com/10.1007/978-981-15-9251-5_8 [Accessed: 14 May 2024].

Feng, Y., Xu, Y., Hu, Q., Krishnamoorthy, S. and Tang, Z. 2022. Predicting vacant parking space availability zone-wisely: a hybrid deep learning approach. *Complex & Intelligent Systems* 8(5), pp. 4145–4161. doi: 10.1007/s40747-022-00700-1.

Feyisa, D.W., Debelee, T.G., Ayano, Y.M., Kebede, S.R. and Assore, T.F. 2022. Lightweight Multireceptive Field CNN for 12-Lead ECG Signal Classification. *Computational Intelligence and Neuroscience* 2022(1), p. 8413294. doi: 10.1155/2022/8413294.

Fotouhi, Z., Hashemi, M.R., Narimani, H. and Bayram, I.S. 2019. A General Model for EV Drivers' Charging Behavior. *IEEE Transactions on Vehicular Technology* 68(8), pp. 7368–7382. doi: 10.1109/TVT.2019.2923260.

Franconeri, S.L., Padilla, L.M., Shah, P., Zacks, J.M. and Hullman, J. 2021. The Science of Visual Data Communication: What Works. *Psychological Science in the Public Interest* 22(3), pp. 110–161. doi: 10.1177/15291006211051956.

Frendo, O., Gaertner, N. and Stuckenschmidt, H. 2021. Improving Smart Charging Prioritization by Predicting Electric Vehicle Departure Time. *IEEE Transactions on Intelligent Transportation Systems* 22(10), pp. 6646–6653. doi: 10.1109/TITS.2020.2988648.

Genofer, R., Glyn, G.G. and Sing, M.A.S. 2018. A REVIEW ON ELECTRIC POWERED VEHICLE AND IT'S BATTERY. *Proceedings of 4th International Conference on Energy*

Efficient Technologies for Sustainability–ICEETS'18. St.Xavier's Catholic College of Engineering, Nagercoil, TamilNadu, India, from 5 th to 7th April, 2018. Available in SSRN eLibrary of ELSEVIER.

Granger, C.W.J. 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 37(3), p. 424. doi: 10.2307/1912791.

Gruver, N., Finzi, M., Qiu, S. and Wilson, A.G. 2024. Large Language Models Are Zero-Shot Time Series Forecasters. Available at: http://arxiv.org/abs/2310.07820 [Accessed: 12 December 2024].

Guetterman, T.C. 2019. Basics of statistics for primary care research. *Family Medicine and Community Health* 7(2), p. e000067. doi: 10.1136/fmch-2018-000067.

Gunarto, D.M., Sa'adah, S. and Utama, D.Q. 2023. Predicting Cryptocurrency Price Using RNN and LSTM Method. *Jurnal Sisfokom (Sistem Informasi dan Komputer)* 12(1), pp. 1–8. doi: 10.32736/sisfokom.v12i1.1554.

Guyon, I. and Elisseeff, A. 2003. An Introduction to Variable and Feature Selection. *Journal of machine learning research* (3(Mar)), pp. 1157–1182.

Hecht, C., Figgener, J. and Sauer, D.U. 2021. Predicting Electric Vehicle Charging Station Availability Using Ensemble Machine Learning. *Energies* 14, p. 7834. doi: 10.3390/en14237834.

Hill, G., Heidrich, O., Creutzig, F. and Blythe, P. 2019. The role of electric vehicles in near-term mitigation pathways and achieving the UK's carbon budget. *Applied Energy* 251, p. 113111. doi: 10.1016/j.apenergy.2019.04.107.

IEA. 2024. Available at: https://www.iea.org/reports/global-ev-outlook-2024 [Accessed: 17 May 2024].

Jin, M. et al. 2024. Time-LLM: Time Series Forecasting by Reprogramming Large Language Models. Available at: http://arxiv.org/abs/2310.01728 [Accessed: 12 December 2024].

Joshi, A., Kale, S., Chandel, S. and Pal, D. 2015. Likert Scale: Explored and Explained. *British Journal of Applied Science & Technology* 7(4), pp. 396–403. doi: 10.9734/BJAST/2015/14975.

Ke, G. et al. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html [Accessed: 21 October 2024].

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M. and Inman, D.J. 2019. 1D Convolutional Neural Networks and Applications: A Survey. Available at: https://arxiv.org/abs/1905.03554 [Accessed: 16 October 2024].

Küfeoğlu, S. and Hong, D. 2020. Emissions performance of electric vehicles: A case study from the United Kingdom. *Applied Energy* 260, p. 114241. doi: 10.1016/j.apenergy.2019.114241.

Kylén, M., Schmidt, S.M., Jonsson, O., Slaug, B. and Iwarsson, S. 2020. Awareness of and Attitudes Toward User Involvement in Research on Aging and Health: Protocol for a Quantitative Large-Scale Panel Study. *JMIR Research Protocols* 9(9), p. e17759. doi: 10.2196/17759.

Lara-Benítez, P., Carranza-García, M. and Riquelme, J.C. 2021. An Experimental Review on Deep Learning Architectures for Time Series Forecasting. *International Journal of Neural Systems* 31(03), p. 2130001. doi: 10.1142/S0129065721300011.

Lavrenova, Y.R. and Denysiuk, S.P. 2023. Optimization of power dispatching schedule of a charging station based on a micro grid with a photovoltaic module. *Applied Aspects of Information Technology* 6(4), pp. 391–403. doi: 10.15276/aait.06.2023.26.

LeCun, Y., Bengio, Y. and Hinton, G. 2015. Deep learning. *Nature* 521(7553), pp. 436–444. doi: 10.1038/nature14539.

Lee, K., Ray, J. and Safta, C. 2021. The predictive skill of convolutional neural networks models for disease forecasting. Chen, C.-H. ed. *PLOS ONE* 16(7), p. e0254319. doi: 10.1371/journal.pone.0254319.

Li, D. and Li, L. 2022. Detection of Water pH Using Visible Near-Infrared Spectroscopy and One-Dimensional Convolutional Neural Network. *Sensors* 22(15), p. 5809. doi: 10.3390/s22155809.

Li, Q. and Xiong, H. 2023. Short-term Load Prediction of Electric Vehicle Charging Stations built on GA-BPNN Model. *Journal of Physics: Conference Series* 2474(1), p. 012023. doi: 10.1088/1742-6596/2474/1/012023.

Li, W. and Law, K.L.E. 2024. Deep Learning Models for Time Series Forecasting: A Review. *IEEE Access* 12, pp. 92306–92327. doi: 10.1109/ACCESS.2024.3422528.

Li, Y., Huang, Y. and Zhang, M. 2018. Short-Term Load Forecasting for Electric Vehicle Charging Station Based on Niche Immunity Lion Algorithm and Convolutional Neural Network. *Energies* 11(5), p. 1253. doi: 10.3390/en11051253.

Lilliefors, H.W. 1967. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association* 62(318), pp. 399–402. doi: 10.1080/01621459.1967.10482916.

Lim, B., Arik, S.O., Loeff, N. and Pfister, T. 2019. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. Available at: https://arxiv.org/abs/1912.09363 [Accessed: 7 October 2024].

Lipton, Z.C., Berkowitz, J. and Elkan, C. 2015. A Critical Review of Recurrent Neural Networks for Sequence Learning. Available at: http://arxiv.org/abs/1506.00019 [Accessed: 7 October 2024].

Lu, Z., Zheng, C. and Yang, T. 2020. Application of Offshore Visibility Forecast Based on Temporal Convolutional Network and Transfer Learning. Zhang, N. ed. *Computational Intelligence and Neuroscience* 2020, pp. 1–12. doi: 10.1155/2020/8882279.

Lucas, A., Barranco, R. and Refa, N. 2019. EV Idle Time Estimation on Charging Infrastructure, Comparing Supervised Machine Learning Regressions. *Energies* 12(2), p. 269. doi: 10.3390/en12020269.

Lundberg, S. and Lee, S.-I. 2017. A Unified Approach to Interpreting Model Predictions. Available at: http://arxiv.org/abs/1705.07874 [Accessed: 21 October 2024].

Ma, D., Song, Y. and Zhou, Z. 2022. Research on the Characterization of Electric Vehicle Energy Storage Capacity and Coordination Mechanism with Distribution Network Based on Two-stage Method. *Journal of Physics: Conference Series* 2237(1), p. 012024. doi: 10.1088/1742-6596/2237/1/012024.

Ma, J. and Zhang, L. 2018. A Deploying Method for Predicting the Size and Optimizing the Location of an Electric Vehicle Charging Stations. *Information* 9(7), p. 170. doi: 10.3390/info9070170.

Ma, T.-Y. and Faye, S. 2021. Multistep Electric Vehicle Charging Station Occupancy Prediction using Mixed LSTM Neural Networks. Available at: https://arxiv.org/abs/2106.04986v1 [Accessed: 8 November 2021].

Ma, T.-Y. and Faye, S. 2022. Multistep electric vehicle charging station occupancy prediction using hybrid LSTM neural networks. *Energy* 244, p. 123217. doi: 10.1016/j.energy.2022.123217.

Mackey, A. and Gass, S.M. 2015. *Second Language Research*. 0 ed. Routledge. Available at: https://www.taylorfrancis.com/books/9781317612650 [Accessed: 14 October 2024].

Majidpour, M. et al. 2016. Forecasting the EV charging load based on customer profile or station measurement. *Applied Energy*. doi: 10.1016/j.apenergy.2015.10.184.

Mandys, F. 2021. Electric vehicles and consumer choices. *Renewable and Sustainable Energy Reviews* 142, p. 110874. doi: 10.1016/j.rser.2021.110874.

Marei, M. and Li, W. 2022. Cutting tool prognostics enabled by hybrid CNN-LSTM with transfer learning. *The International Journal of Advanced Manufacturing Technology* 118(3–4), pp. 817–836. doi: 10.1007/s00170-021-07784-y.

Mathew, S., Pulugurtha, S.S., Bhure, C. and Duvvuri, S. 2023. One-Dimensional Convolutional Neural Network Model for Local Road Annual Average Daily Traffic Estimation. *IEEE Access* 11, pp. 127229–127241. doi: 10.1109/ACCESS.2023.3332125.

McHugh, M.L. 2013. The Chi-square test of independence. *Biochemia Medica*, pp. 143–149. doi: 10.11613/BM.2013.018.

Mekky, M.F. and Collins, A.R. 2024. The Impact of state policies on electric vehicle adoption -A panel data analysis. *Renewable and Sustainable Energy Reviews* 191, p. 114014. doi: 10.1016/j.rser.2023.114014.

Morrissey, P., Weldon, P. and O'Mahony, M. 2016. Future standard and fast charging infrastructure planning: An analysis of electric vehicle charging behaviour. *Energy Policy* 89, pp. 257–270. doi: 10.1016/j.enpol.2015.12.001.

Morstyn, T., Teytelboym, A. and McCulloch, M.D. 2018. Matching Markets with Contracts for Electric Vehicle Smart Charging. In: *2018 IEEE Power & Energy Society General Meeting (PESGM)*. Portland, OR: IEEE, pp. 1–5. Available at: https://ieeexplore.ieee.org/document/8586361/ [Accessed: 7 November 2024].

Morton, C., Lovelace, R. and Anable, J. 2017. Exploring the effect of local transport policies on the adoption of low emission vehicles: Evidence from the London Congestion Charge and Hybrid Electric Vehicles. *Transport Policy* 60, pp. 34–46. doi: 10.1016/j.tranpol.2017.08.007.

Motz, M., Huber, J. and Weinhardt, C. 2021. *Forecasting BEV charging station occupancy at work places*. Gesellschaft für Informatik, Bonn. Available at: http://dl.gi.de/handle/20.500.12116/34779 [Accessed: 8 November 2021].

Nagelkerke, N.J.D. 1991. A note on a general definition of the coefficient of determination. *Biometrika* 78(3), pp. 691–692. doi: 10.1093/biomet/78.3.691.

Nannavecchia, A., Girardi, F., Fina, P.R., Scalera, M. and Dimauro, G. 2021. Personal Heart Health Monitoring Based on 1D Convolutional Neural Network. *Journal of Imaging* 7(2), p. 26. doi: 10.3390/jimaging7020026.

Nayak, B.K. and Hazra, A. 2011. How to choose the right statistical test? *Indian Journal of Ophthalmology* 59(2), pp. 85–86. doi: 10.4103/0301-4738.77005.

Nespoli, A., Ogliari, E. and Leva, S. 2023. User Behavior Clustering Based Method for EV Charging Forecast. *IEEE Access* 11, pp. 6273–6283. doi: 10.1109/ACCESS.2023.3235952.

Ostermann, A., Fabel, Y., Ouan, K. and Koo, H. 2022. Forecasting Charging Point Occupancy Using Supervised Learning Algorithms. *Energies* 15(9), p. 3409. doi: 10.3390/en15093409.

Páez, A. and Boisjoly, G. 2022. Exploratory Data Analysis. In: *Discrete Choice Analysis with R*. Use R! Cham: Springer International Publishing, pp. 25–64. Available at: https://link.springer.com/10.1007/978-3-031-20719-8_2 [Accessed: 6 August 2024].

Pan, Q., Hu, W. and Zhu, J. 2020. Series Saliency: Temporal Interpretation for Multivariate Time Series Forecasting. Available at: http://arxiv.org/abs/2012.09324 [Accessed: 7 October 2024].

Pavlyshenko, B.M. 2020. SALES TIME SERIES ANALYTICS USING DEEP Q-LEARNING. *International Journal of Computing*, pp. 434–441. doi: 10.47839/ijc.19.3.1892.

Pavlyshenko, B.M. 2022. Forecasting of Non-Stationary Sales Time Series Using Deep Learning. Available at: http://arxiv.org/abs/2205.11636 [Accessed: 7 October 2024].

Peng, H., Wu, J. and Wang, L. 2014. Electric Vehicle Charging Load Model Based on Diffusion Theory. *The Open Mechanical Engineering Journal* 8(1), pp. 954–959. doi: 10.2174/1874155X01408010954.

Picon, A. et al. 2019. Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia. *PLOS ONE* 14(5), p. e0216756. doi: 10.1371/journal.pone.0216756.

Powers, D.M.W. 2020. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Available at: https://arxiv.org/abs/2010.16061 [Accessed: 16 October 2024].

Qiao, F. and Lin, S. 2021. Data-driven prediction of fine-grained EV charging behaviors in public charging stations: Poster. In: *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*. Virtual Event Italy: ACM, pp. 276–277. Available at: https://dl.acm.org/doi/10.1145/3447555.3466567 [Accessed: 2 November 2021].

Qu, T. 2018. Short Term Prediction of Electric Vehicle Charging Load Based on Optimized Genetic Algorithm. In: *Proceedings of the 2018 8th International Conference on Management, Education and Information (MEICI 2018)*. Shenzang, China: Atlantis Press. Available at: https://www.atlantis-press.com/article/55910094 [Accessed: 8 November 2024].

Rai, A. and Mitra, M. 2021. Lamb wave based damage detection in metallic plates using multi-headed 1-dimensional convolutional neural network. *Smart Materials and Structures* 30(3), p. 035010. doi: 10.1088/1361-665X/abdd00.

Raugei, M., Hutchinson, A. and Morrey, D. 2018. Can electric vehicles significantly reduce our dependence on non-renewable energy? Scenarios of compact vehicles in the UK as a case in point. *Journal of Cleaner Production* 201, pp. 1043–1051. doi: 10.1016/j.jclepro.2018.08.107.

Rauh, N., Franke, T. and Krems, J.F. 2015. Understanding the Impact of Electric Vehicle Driving Experience on Range Anxiety. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57(1), pp. 177–187. doi: 10.1177/0018720814546372.

Reza, S., Ferreira, M.C., Machado, J.J.M. and Tavares, J.M.R.S. 2022. Traffic State Prediction Using One-Dimensional Convolution Neural Networks and Long Short-Term Memory. *Applied Sciences* 12(10), p. 5149. doi: 10.3390/app12105149.

Sadık, O. 2019. A discussion of the concepts of validity and reliability in qualitative and quantitative Research. Available at: https://open.metu.edu.tr/handle/11511/96505 [Accessed: 2 September 2024].

Salah, M., Higa, H., Ishizaka, J. and Salem, S.I. 2023. 1D Convolutional Neural Network-based Chlorophyll-a Retrieval Algorithm for Sentinel-2 MultiSpectral Instrument in Various Trophic States. *Sensors and Materials* 35(11), p. 3743. doi: 10.18494/SAM4331.

Salinas, D., Flunkert, V., Gasthaus, J. and Januschowski, T. 2020. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting* 36(3), pp. 1181–1191. doi: 10.1016/j.ijforecast.2019.07.001.

Sao, A., Tempelmeier, N. and Demidova, E. 2021. Deep Information Fusion for Electric Vehicle Charging Station Occupancy Forecasting. In: *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. pp. 3328–3333. doi: 10.1109/ITSC48978.2021.9565097.

Schon, E.-M., Thomaschewski, J. and Escalona, M.J. 2020. Lean User Research for Agile Organizations. *IEEE Access* 8, pp. 129763–129773. doi: 10.1109/ACCESS.2020.3009101.

Schuemie, M.J., Ryan, P.B., DuMouchel, W., Suchard, M.A. and Madigan, D. 2014. Interpreting observational studies: why empirical calibration is needed to correct *p* -values. *Statistics in Medicine* 33(2), pp. 209–218. doi: 10.1002/sim.5925.

Shahid, F., Zameer, A. and Muneeb, M. 2020. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons & Fractals* 140, p. 110212. doi: 10.1016/j.chaos.2020.110212.

Shaw, S., Chia, K. and Kleissl, J. 2022. Neighbor-Based Optimized Logistic Regression Machine Learning Model For Electric Vehicle Occupancy Detection. Available at: https://arxiv.org/abs/2204.13702 [Accessed: 11 October 2024].

Sheskin, D.J. 2003. *Handbook of Parametric and Nonparametric Statistical Procedures: Third Edition*. 3rd ed. New York: Chapman and Hall/CRC. doi: 10.1201/9781420036268.

Singh, H., Ambikapathy, A., Logavani, K., Arun Prasad, G. and Thangavel, S. 2021. Plug-In Hybrid Electric Vehicles (PHEVs). In: Patel, N., Bhoi, A. K., Padmanaban, S., and Holm-Nielsen, J. B. eds. *Electric Vehicles: Modern Technologies and Trends*. Singapore: Springer, pp. 53–72. Available at: https://doi.org/10.1007/978-981-15-9251-5_3 [Accessed: 14 May 2024].

Situ, L. 2009. Electric Vehicle development: The past, present & future. In: *2009 3rd International Conference on Power Electronics Systems and Applications (PESA)*. pp. 1–3. Available at: https://ieeexplore.ieee.org/document/5228601 [Accessed: 6 May 2024].

Sokolova, M. and Lapalme, G. 2009. A systematic analysis of performance measures for classification tasks. *Information Processing & Management* 45(4), pp. 427–437. doi: 10.1016/j.ipm.2009.03.002.

Soldan, F., Bionda, E., Mauri, G. and Celaschi, S. 2021. Short-term forecast of EV charging stations occupancy probability using big data streaming analysis. Available at: https://arxiv.org/abs/2104.12503v1 [Accessed: 8 November 2021].

Solís, M. and Calvo-Valverde, L.-A. 2022. Performance of Deep Learning models with transfer learning for multiple-step-ahead forecasts in monthly time series. *Inteligencia Artificial* 25(70), pp. 110–125. doi: 10.4114/intartif.vol25iss70pp110-125.

Sözen, E. and Güven, U. 2019. The Effect of Online Assessments on Students' Attitudes towards Undergraduate-Level Geography Courses. *International Education Studies* 12(10), pp. 1–8.

Stein, A. et al. 2024. Forecasting Electric Vehicle Charging Behavior in Workplace Charging Infrastructure with Limited Privacy-Restricted Real Data. In: *2024 IEEE Transportation Electrification Conference and Expo (ITEC)*. Chicago, IL, USA: IEEE, pp. 1–7. Available at: https://ieeexplore.ieee.org/document/10598951/ [Accessed: 11 October 2024].

Steinhilber, S., Wells, P. and Thankappan, S. 2013. Socio-technical inertia: Understanding the barriers to electric vehicles. *Energy Policy* 60, pp. 531–539. doi: 10.1016/j.enpol.2013.04.076.

Su, S., Li, Y., Chen, Q., Xia, M., Yamashita, K. and Jurasz, J. 2023. Operating Status Prediction Model at EV Charging Stations With Fusing Spatiotemporal Graph Convolutional Network. *IEEE Transactions on Transportation Electrification* 9(1), pp. 114–129. doi: 10.1109/TTE.2022.3192285.

Sun, D., Ou, Q., Yao, X., Gao, S., Wang, Z., Ma, W. and Li, W. 2020. Integrated human-machine intelligence for EV charging prediction in 5G smart grid. *EURASIP Journal on Wireless Communications and Networking* 2020(1), p. 139. doi: 10.1186/s13638-020-01752-y.

Sun, H., Fung, J.C.H., Chen, Y., Li, Z., Yuan, D., Chen, W. and Lu, X. 2022. Development of an LSTM broadcasting deep-learning framework for regional air pollution forecast improvement. *Geoscientific Model Development* 15(22), pp. 8439–8452. doi: 10.5194/gmd-15-8439-2022.

Tal, G., Nicholas, M.A., Davies, J. and Woodjack, J. 2014. Charging Behavior Impacts on Electric Vehicle Miles Traveled: Who is Not Plugging In? *Transportation Research Record: Journal of the Transportation Research Board* 2454(1), pp. 53–60. doi: 10.3141/2454-07.

Tan, M., Merrill, M.A., Gupta, V., Althoff, T. and Hartvigsen, T. 2024. Are Language Models Actually Useful for Time Series Forecasting? Available at: http://arxiv.org/abs/2406.16964 [Accessed: 12 December 2024].

Tang, W., Long, G., Liu, L., Zhou, T., Blumenstein, M. and Jiang, J. 2020. Omni-Scale CNNs: a simple and effective kernel size configuration for time series classification. Available at: https://arxiv.org/abs/2002.10061 [Accessed: 16 October 2024].

Taylor, J. 2022. *Electric Cars*. Bloomsbury Publishing.

Tibshirani, R. 1996. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58(1), pp. 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x.

Transport Scotland. 2017. *Switched On Scotland Phase Two: An Action Plan For Growth*. Available at: https://www.transport.gov.scot/media/39306/switched-on-scotland-phase-2.pdf.

Velarde, G. 2023. *Forecasting with Deep Learning*. Available at: https://arxiv.org/abs/2302.12027v1 [Accessed: 7 October 2024].

Velychko, O., Gordiyenko, T., and State University of Telecommunications. 2022. COMPARATIVE ANALYSIS OF TECHNICAL CHARACTERISTICS OF CHARGING STATIONS OF ELECTRIC VEHICLES. *Measuring Equipment and Metrology* 83(2), pp. 34–44. doi: 10.23939/istcmtm2022.02.034.

Viswanathan, S., Appel, J., Chang, L., Man, I.V., Saba, R. and Gamel, A. 2018a. Development of an assessment model for predicting public electric vehicle charging stations. *European Transport Research Review* 10(2), p. 54. doi: 10.1186/s12544-018-0322-8.

Viswanathan, S., Appel, J., Chang, L., Man, I.V., Saba, R. and Gamel, A. 2018b. Development of an assessment model for predicting public electric vehicle charging stations. *European Transport Research Review* 10(2), p. 54. doi: 10.1186/s12544-018-0322-8.

Wakefield, E.H. 1993. *History of the Electric Automobile: Battery-Only Powered Cars*. SAE International.

Wan, R., Mei, S., Wang, J., Liu, M. and Yang, F. 2019. Multivariate Temporal Convolutional Network: A Deep Neural Networks Approach for Multivariate Time Series Forecasting. *Electronics* 8(8), p. 876. doi: 10.3390/electronics8080876.

Wang, T., Li, Q., Yin, L., Chen, W., Breaz, E. and Gao, F. 2021. Hierarchical Power Allocation Method Based on Online Extremum Seeking Algorithm for Dual-PEMFC/Battery Hybrid Locomotive. *IEEE Transactions on Vehicular Technology* 70(6), pp. 5679–5692. doi: 10.1109/TVT.2021.3078752.

Wen, Y., MacKenzie, D. and Keith, D.R. 2016. Modeling the Charging Choices of Battery Electric Vehicle Drivers by Using Stated Preference Data. *Transportation Research Record: Journal of the Transportation Research Board* 2572(1), pp. 47–55. doi: 10.3141/2572-06.

Willmott, C. and Matsuura, K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* 30, pp. 79–82. doi: 10.3354/cr030079.

Wolbertus, R., Kroesen, M., Van Den Hoed, R. and Chorus, C.G. 2018. Policy effects on charging behaviour of electric vehicle owners and on purchase intentions of prospective owners: Natural and stated choice experiments. *Transportation Research Part D: Transport and Environment* 62, pp. 283–297. doi: 10.1016/j.trd.2018.03.012.

Xing, Q., Chen, Z., Zhang, Z., Xu, X., Zhang, T., Huang, X. and Wang, H. 2020. Urban Electric Vehicle Fast-Charging Demand Forecasting Model Based on Data-Driven Approach and Human Decision-Making Behavior. *Energies* 13(6), p. 1412. doi: 10.3390/en13061412.

Xu, T. et al. 2022. Prediction of temporal and spatial distribution of electric vehicle charging load considering traffic flow. In: Pei, Z. ed. *International Conference on Mechanisms and Robotics (ICMAR 2022)*. Zhuhai, China: SPIE, p. 110. Available at: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12331/2652465/Prediction-of-temporal-and-spatial-distribution-of-electric-vehicle-charging/10.1117/12.2652465.full [Accessed: 7 October 2024].

Yang, C., Zhou, H., Chen, X. and Huang, J. 2024. Demand Time Series Prediction of Stacked Long Short-Term Memory Electric Vehicle Charging Stations Based on Fused Attention Mechanism. *Energies* 17(9), p. 2041. doi: 10.3390/en17092041.

Yu, A.S.O., Silva, L.L.C., Chu, C.L., Nascimento, P.T.S. and Camargo, A.S. 2011. Electric vehicles: Struggles in creating a market. In: *2011 Proceedings of PICMET '11: Technology Management in the Energy Smart World (PICMET)*. pp. 1–13. Available at: https://ieeexplore.ieee.org/abstract/document/6017911 [Accessed: 15 May 2024].

Yuan, X., Tanksley, D., Li, L., Zhang, H., Chen, G. and Wunsch, D. 2021. Faster Post-Earthquake Damage Assessment Based on 1D Convolutional Neural Networks. *Applied Sciences* 11(21), p. 9844. doi: 10.3390/app11219844.

Zhao, H., Chen, Z., Jiang, H., Jing, W., Sun, L. and Feng, M. 2019a. Evaluation of Three Deep Learning Models for Early Crop Classification Using Sentinel-1A Imagery Time Series—A Case Study in Zhanjiang, China. *Remote Sensing* 11(22), p. 2673. doi: 10.3390/rs11222673.

Zhao, W., Gao, Y., Ji, T., Wan, X., Ye, F. and Bai, G. 2019b. Deep Temporal Convolutional Networks for Short-Term Traffic Flow Forecasting. *IEEE Access* 7, pp. 114496–114507. doi: 10.1109/ACCESS.2019.2935504.

Zheng, A. and Casari, A. 2018. *Feature engineering for machine learning: principles and techniques for data scientists*. Beijing Boston Farnham Sebastopol Tokyo: O'Reilly.

Zhou, D., Guo, Z., Xie, Y., Hu, Y., Jiang, D., Feng, Y. and Liu, D. 2022. Using Bayesian Deep Learning for Electric Vehicle Charging Station Load Forecasting. *Energies* 15(17), p. 6195. doi: 10.3390/en15176195.

Zhu, J. et al. 2019. Short-Term Load Forecasting for Electric Vehicle Charging Stations Based on Deep Learning Approaches. *Applied Sciences*. doi: 10.3390/app9091723.

Zou, H. and Hastie, T. 2005. Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67(2), pp. 301–320. doi: 10.1111/j.1467-9868.2005.00503.x.

Appendix A: User Study A

participant information sheet

Introductory brief My name is Adel Dadaa, and I am a PhD student at the University of Strathclyde, the department of Computer and Information Sciences. My email address is adel.dadaa@strath.ac.uk. My current research is mainly about investigating ways to provide helpful forecasting about the availability of public charging stations for electric vehicles using deep learning approaches. And at this stage of my research, we study the electric vehicle owners and their charging habits and preferences regarding the availability forecasting representation. I would like you to complete a short online questionnaire about charging habits and how mobile apps could better support charging habits. The survey will take about 10 minutes. The form does not ask for any personal details unless you choose to enter into a prize draw to win a £30 shopping voucher, but details about your electric vehicle use and charging. The survey results may be published in academic publications and included in the thesis. You can stop the survey at any point, and you do not have to answer all the questions in the survey. By the end of this survey, we will ask you If you would like to be entered into a prize draw to win a £30 shopping voucher. What is the purpose of this survey? This survey aims to obtain feedback from EV owners about their expectations in the form of the outputs of the availability predictions. Moreover, this survey aims to analyze the different charging events of EV owners and study their perspectives on using forecasting tools to plan their charging time. Do you have to take part? No, the survey is entirely voluntary; participants are free to participate in this survey. They can withdraw from this survey at any point without any consequences. Who can take part in this survey? Any EV owner or previously owned EV in the UK and over 18 years old.

Appendix A: User Study A

A Participating confirmation

☐ Please press here if you agree to take part in this survey (1)

Skip To: End of Survey If Participating confirmation != Please press here if you agree to take part in this survey

End of Block: Introductory brief

Start of Block: General Details

Q1 What is your age group?

○ 18-25 (1)

○ 26-35 (2)

○ 36-49 (3)

○ Above 50 (4)

○ Prefer not to say (5)

Q2 Which one describes your gender

○ Male (1)

○ Female (2)

○ Prefer not to say (3)

End of Block: General Details

Start of Block: EV Experience & • Charging habits

Q3 For how long have you owned or drove EV?

○ Up to one year (1)

○ 1-3 years (2)

○ More than 3 years (3)

Q4 Which charging point do you mostly prefer to use to charge your EV?

- ○ Slow (up to 3kW) (1)
- ○ Fast (7kW) (2)
- ○ Fast (22kW) (3)
- ○ Rapid (from 43kW) (4)

Q5 Where do you often prefer to charge your EV?

- ○ Home charging point (1)
- ○ Workplace charger (2)
- ○ Public charging stations (3)

Q6 In which part of the day do you charge your vehicle overwhelmingly?

- ○ Early mornings (1)
- ○ Middle of the days (2)
- ○ Over nights (3)

End of Block: EV Experience & • Charging habits

Start of Block: Thinking about doing a long journey that involved charging on route

Q7 Have you ever or you were about to encounters the problem of draining your EV's battery before reaching a charging point?

- ○ Never (1)
- ○ No, but I have been close (2)
- ○ Yes, one or two times (3)
- ○ Yes, three times and more (4)

Q8 If yes, give a reason for that? *for example: Lack of planning ahead or the availability of charging points...*

____

Q9 If it is available, do you feel comfortable using online platforms to help you choose public charging location?

- Extremely uncomfortable (1)
- Somewhat uncomfortable (2)
- Neither comfortable nor uncomfortable (3)
- Somewhat comfortable (4)
- Extremely comfortable (5)

Q10 Have you used any online platform to help you find a place to charge your EV?

- Yes, sometimes (1)
- Yes, always (2)
- Never (3)

Q11 How far do you expect to see predictions of the availability of charging stations in the future time?

- ○ Up to one hour (1)
- ○ From 1 hour to up to 3 hours (2)
- ○ Form 3 to 6 hours (3)
- ○ More than 6 hours (4)

Q11-a Please briefly explain *For example: Interested in as soon as possible urgently or for long-term planning...*

____

____

____

____

____

Q12 How do you normally plan your charging times during your journeys?

○ By performing regular charging stops. (1)

○ The charging level covers the distance of the journey. (2)

○ Intuitively by performing intermittent and random stops during the journey. (3)

○ another strategy (4)

Display This Question:

If How do you normally plan your charging times during your journeys? = another strategy

Q13 If you have another strategy, please give details.

____

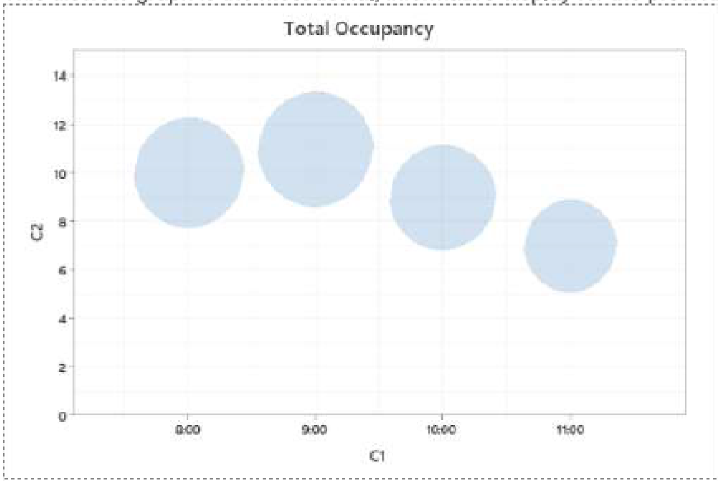
Q14

Now you will see different ways of representing the output of a prediction platform for the availability state of a charging station. Can you rate them, in your point of view, from one of the available options.

Page Break

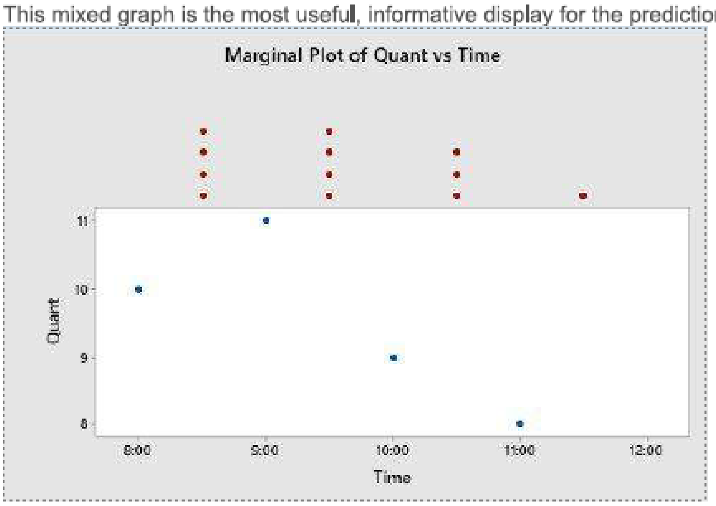
	Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal

This discrete graph is the most useful, informative display for the prediction



	Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal

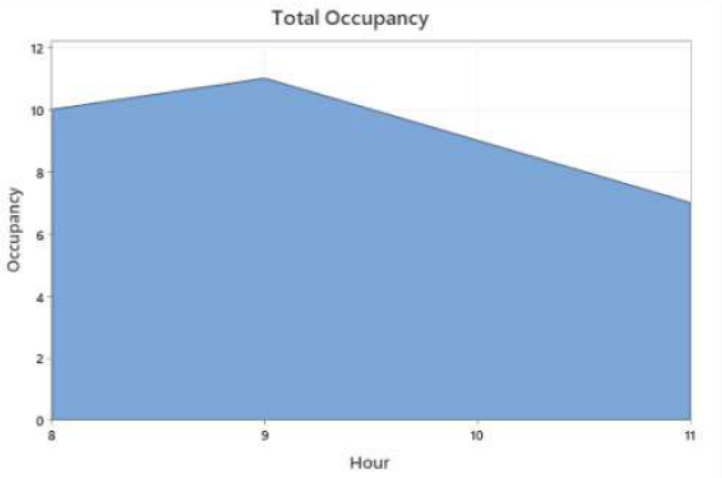
This mixed graph is the most useful, informative display for the prediction



	Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal

Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal

This continue graph is the most useful, informative display for the prediction



○	○	○	○	○
Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal

This text output is the most useful, informative display for the prediction

The prediction of the current charging point for the following three hours are:

1- 30% free

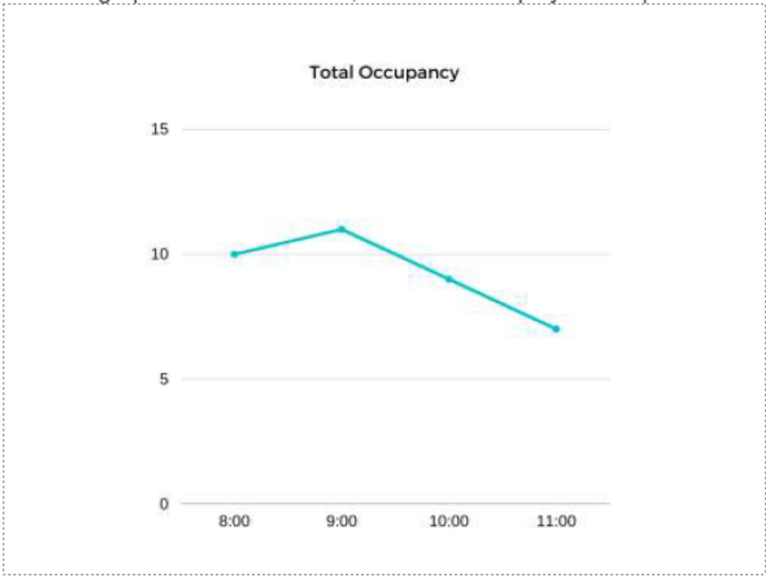
2- 50% free

3- 20% free

○	○	○	○	○
Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal

Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal

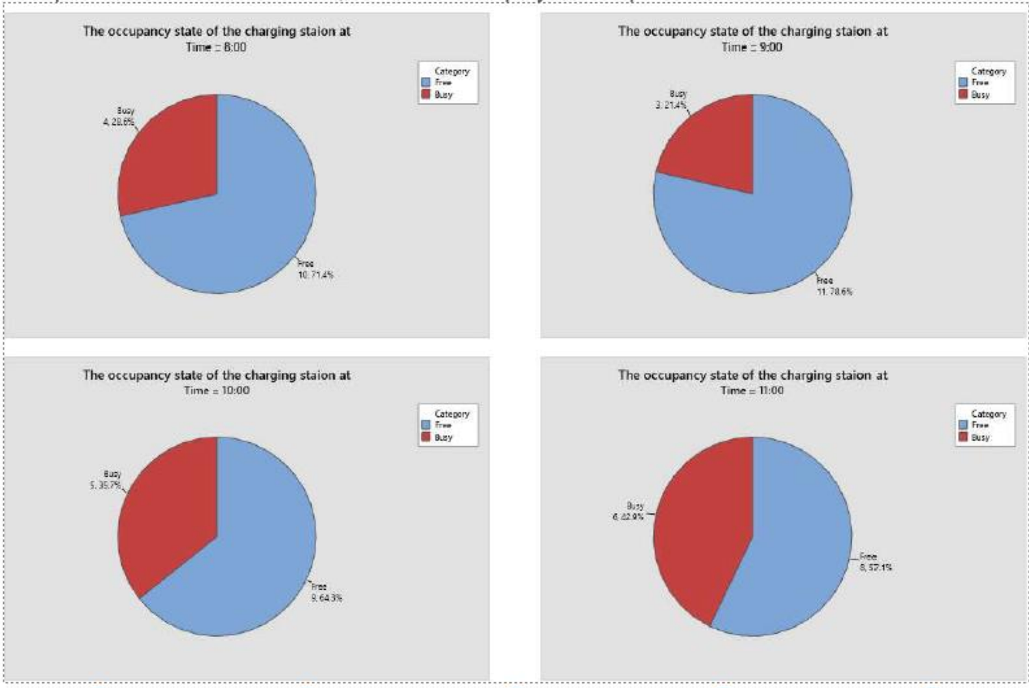
This line graph is the most useful, informative display for the prediction



For each output form below, choose either one of the specified options below to describe the way of the output in terms its usefulness and clarity:

Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal

This pie chart is the most useful, informative display for the prediction



Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal

This text output is the most useful, informative display for the prediction

The predicted occupancy rate for the following three hours for this charging point are:

1- Full / almost full.

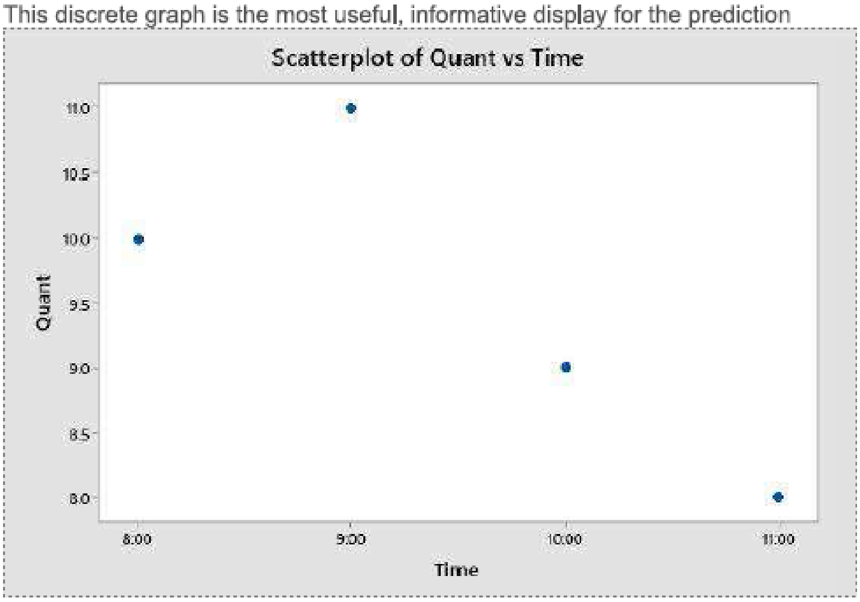
2- High occupancy rate.

3- Almost free.

Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal

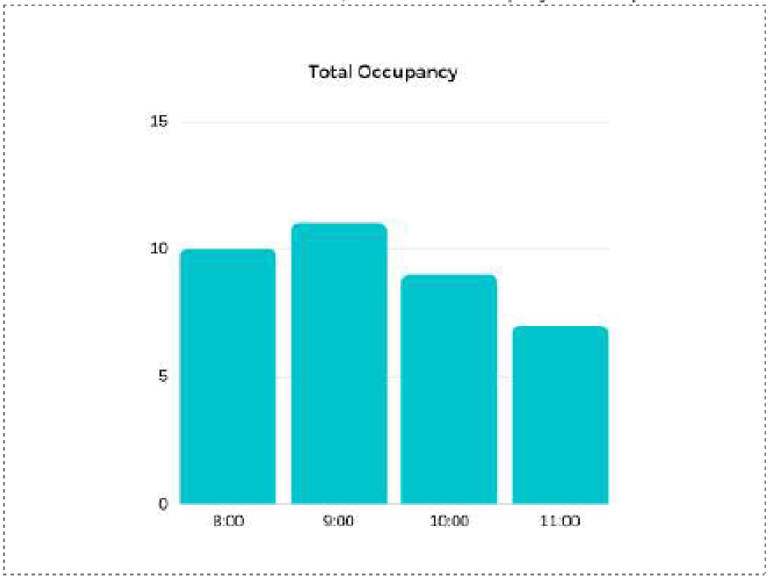
Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal

This discrete graph is the most useful, informative display for the prediction



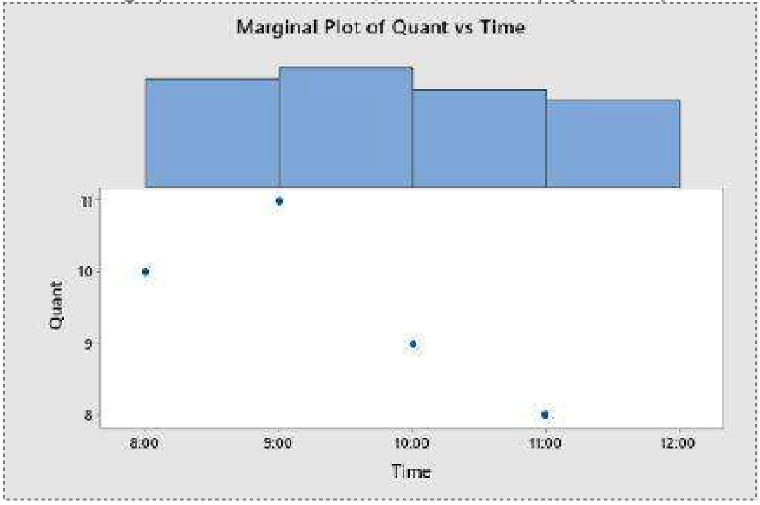
Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal
O	O	O	O	O

This bar chart is the most useful, informative display for the prediction



Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal
O	O	O	O	O

	Dislike a great deal	Dislike somewhat	Neither like nor dislike	Like somewhat	Like a great deal
This mixed graph is the most useful, informative display for the prediction	O	O	O	O	O



This is the end of this survey, thank you for taking part. If you would like to provide your name and contact details and you agree to enter in the prize draw please click agree:

Agree to enter to the prize draw

O

I do not need to be considered in the draw

O

Thanks for taking part in this survey; please fill your details in the following form at this link:

Enter Into the draw

We thank you for your time spent taking this survey.

Powered by Qualtrics

Appendix B: User Study B

Participants Information Sheet

Introduction My name is Adel Ali Dadaa, and I am a PhD student at the University of Strathclyde in the Department of Computer and Information Sciences. My research focuses on smart techniques for predicting the availability of public charging stations for electric vehicle (EV) owners. The objective is to provide this information to EV drivers to facilitate planning for long trips. **What is the purpose of this interview?** We aim to study the charging habits of EV owners and understand their preferences concerning the display of predictions. Additionally, we'll be evaluating their experience with using deep learning (DL)-based platforms. **What will you do in the project?** Once you provide your name and email, we will reach out to schedule a convenient interview time. This interview will be conducted online via Zoom and will last approximately 10-15 minutes. As a token of our appreciation, participants **will get a chance to enter the draw to win a £20 Amazon shopping voucher. Do you have to participate?** Participation is entirely voluntary. You have the right to withdraw from this study at any point without facing any repercussions. Who is eligible to participate in the project? Anyone who currently owns, or has previously owned, an electric vehicle in the UK and is above 18 is eligible. **What happens next?** If you are interested in participating, kindly leave your name and email address below. We'll get in touch to finalize the interview details. Even if you decide not to participate, I deeply appreciate your time and consideration in reading about this study.

Appendix B: User Study B

Q1 After you read, please Select of the following:

☐ I am happy to take this interview (1)

☐ I am happy to use my email to contact me about the £20 voucher (2)

☐ I am happy to be contacting for other following up study in this research (3)

Q2 Name :

____

End of Block: Default Question Block

Start of Block: Block 1

Q3 For how long you have been own EV?

○ 1-2 years (1)

○ 2-4 years (2)

○ More than 4 years (3)

Q4 Where do you normally charge your EV?

- Home charger (1)
- Public charger (2)
- workplace (3)

Q5 When do you normally charge your vehicle?

- Early morning (5:00 am - 9:00 am) (1)
- Morning to noon time (9:00 am - 12:00 pm) (2)
- Middle of the day (12:00 pm - 19:00 pm) (3)
- Late night to early morning (19:00 pm - 5:00 am) (4)

Q6 Have you ever used any sort of online platforms to plan your charging sessions?

- No (1)
- sometimes (2)
- Yes, regularly (3)

End of Block: Block 1

Start of Block: Block 2

Q7 I always keen to perform regular stops to charge my EV

- Strongly disagree (1)
- Somewhat disagree (2)
- Neither agree nor disagree (3)
- Somewhat agree (4)
- Strongly agree (5)

Q8 I always struggle to get a charging point to charge my EV in my long trips?

- Strongly disagree (1)
- Somewhat disagree (2)
- Neither agree nor disagree (3)
- Somewhat agree (4)
- Strongly agree (5)

Q9 If I don’t plan ahead, I always have difficulty finding a suitable time to charge my electric car during long trips.

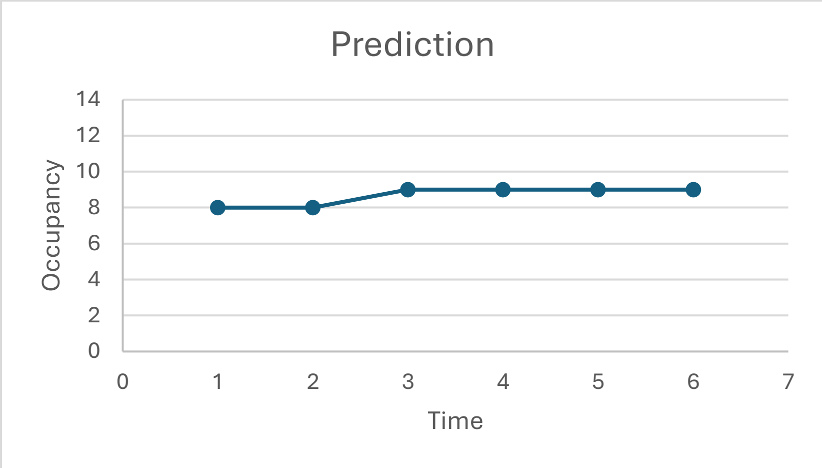
- ◯ Strongly disagree (1)
- ◯ Somewhat disagree (2)
- ◯ Neither agree nor disagree (3)
- ◯ Somewhat agree (4)
- ◯ Strongly agree (5)

Q10 I feel comfortable using online platforms, if it exist, to direct me to the most suitable charging point.

- ◯ Strongly disagree (1)
- ◯ Somewhat disagree (2)
- ◯ Neither agree nor disagree (3)
- ◯ Somewhat agree (4)
- ◯ Strongly agree (5)

End of Block: Block 2

Q11- Scen1- You are looking to plan your trip tomorrow to a city (A), and you need to plan a charging stop around a time (t) as you expect your car battery will be low. You used a predictive model to see the occupancy state for a charging point (X) and you get this results:



Scen2- You drive around the city, and you noticed that your EV urgently need to be charged. You used a predictive model to check the occupancy state for the most closer charging station and you get the following results:

```
The predictions of the occupancy state for the 13/09/2023 17:00:00 is  [8.] spaces and the following 5 hours after are :
  [[8.]
 [8.]
 [8.]
 [8.]
 [9.]]
```

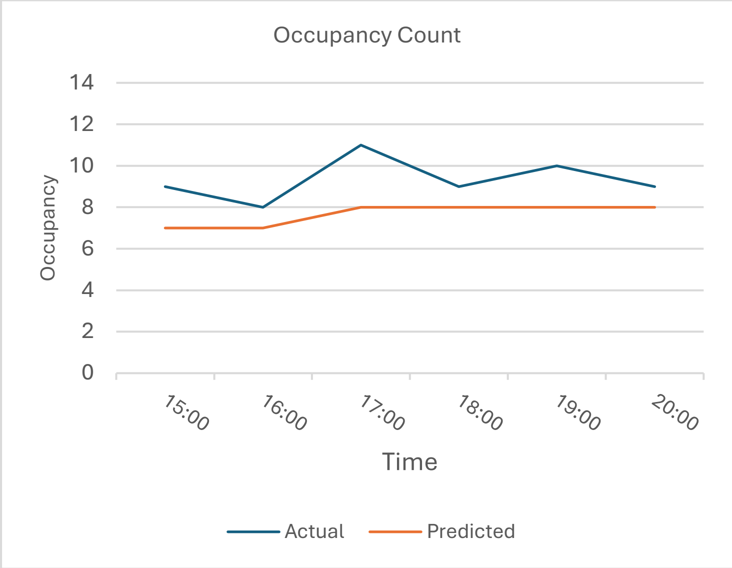
Scen3- Use occupancy category instead of value for the predicted occupancy level.

- *What information you can get from this result, was this way of presenting predictions clear and simple?*

- *Is there any further information that you think of that should be included to make the display clearer and more understandable?*

- *What was your initial decision when you saw these predictions?*

Q12- Reassess the participant decision after showing the model accuracy:



If you told that , this graph shows the overall model accuracy to predict the charging occupancy; where the blue line shows the actual occupancy and the orange line shows the model predictions. What is your thought about the model? Will you trust in the model predictions and use them in the future?

That was the end of the interview, I would like to thank you for participating in this part of my research. Thank you!

Appendix C: Models Development

```
# 1D CNN model
def build_cnn(input_shape, num_classes):
    inputs = Input(shape=input_shape)
    x = Conv1D(32, 3, activation='relu')(inputs)
    x = MaxPooling1D(2)(x)
    x = Flatten()(x)
    x = Dense(64, activation='relu')(x)
    outputs = Dense(num_classes, activation='softmax')(x)
    model = Model(inputs=inputs, outputs=outputs)
    return model

# TCN model
def build_tcn(input_shape, num_classes):
    inputs = Input(shape=input_shape)
    x = TCN(nb_filters=32, kernel_size=3, return_sequences=False)(inputs)
    x = Dense(64, activation='relu')(x)
    outputs = Dense(num_classes, activation='softmax')(x)
```

```
    model = Model(inputs=inputs, outputs=outputs)
    return model



from tensorflow.keras.layers import Bidirectional, GRU

def build_GTcn(input_shape, num_classes):
    # Input Layer
    inputs = Input(shape=input_shape)

    # (TCN)
    tcn_branch = TCN(nb_filters=64, kernel_size=3, return_sequences=False)(inputs)
    # Fully Connected Layers
    tcn_branch1 = Dense(32, activation='relu')(tcn_branch)  # Dense1

    # (BiGRU)
    bigru_branch = GRU(64, return_sequences=False)(inputs)
    bigru_branch1 = Dense(32, activation='relu')(bigru_branch)        # Dense2

    # Combine outputs of both branches
```

```
    combined = Concatenate()([tcn_branch1, bigru_branch1])

    # Output Layer
    outputs = Dense(num_classes, activation='softmax')(combined)

    # Define the Model
    model = Model(inputs=inputs, outputs=outputs)
    return model

def build_BiGTcn(input_shape, num_classes):
    # Input Layer
    inputs = Input(shape=input_shape)

    # (TCN)
    tcn_branch = TCN(nb_filters=64, kernel_size=3, return_sequences=False)(inputs)
    tcn_branch1 = Dense(32, activation='relu')(tcn_branch)  # Dense1

    # Spatial Attention Branch (BiGRU)
    bigru_branch = Bidirectional(GRU(64, return_sequences=False))(inputs)  # Spatial features
    bigru_branch1 = Dense(32, activation='relu')(bigru_branch)        # Dense2
```

```
    # Combine outputs of both branches
    combined = Concatenate()([tcn_branch1, bigru_branch1])

    # Fully Connected Layers

    # Output Layer
    outputs = Dense(num_classes, activation='softmax')(combined)

    # Define the Model
    model = Model(inputs=inputs, outputs=outputs)
    return model

input_shape = (X_train.shape[1], 1)
num_classes = y_train.shape[1]

cnn_model = build_cnn(input_shape, num_classes)
tcn_model = build_tcn(input_shape, num_classes)
GTcn_model = build_GTcn(input_shape, num_classes)
BiGTcn_model = build_BiGTcn(input_shape, num_classes)
```

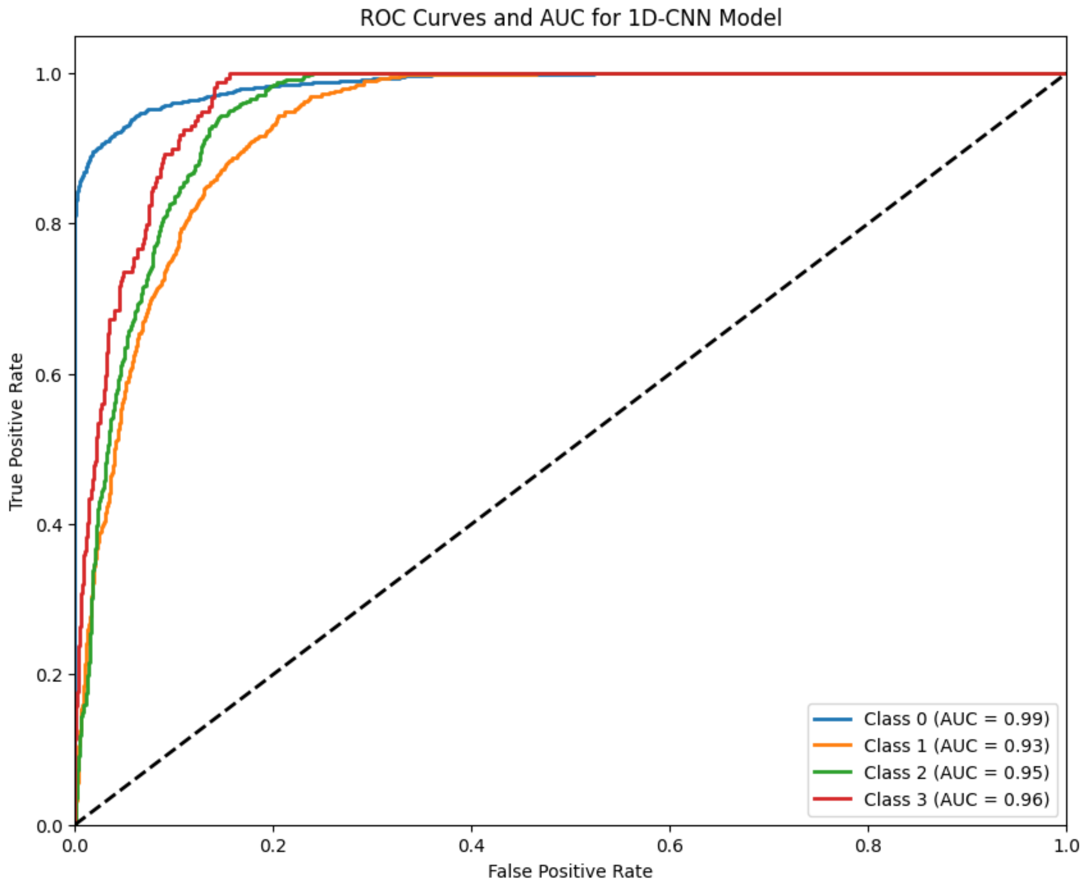
```
# Callbacks for each model
cnn_checkpoint_callback = ModelCheckpoint(filepath=cnn_model_path, save_best_only=True, monitor='val_accuracy', mode='max',
verbose=1)
tcn_checkpoint_callback = ModelCheckpoint(filepath=tcn_model_path, save_best_only=True, monitor='val_accuracy', mode='max',
verbose=1)
GTcn_checkpoint_callback = ModelCheckpoint(filepath=GTcn_model_path, save_best_only=True, monitor='val_accuracy', mode='max',
verbose=1)
BiGTcn_checkpoint_callback = ModelCheckpoint(filepath=BiGTcn_model_path, save_best_only=True, monitor='val_accuracy',
mode='max', verbose=1)

# Compile and train the models
epochs =
batch_size = 20

# Compile and train 1D-CNN model
cnn_model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
history_cnn = cnn_model.fit(X_train, y_train, validation_data=(X_val, y_val), epochs=epochs, batch_size=batch_size,
callbacks=[cnn_checkpoint_callback])
```

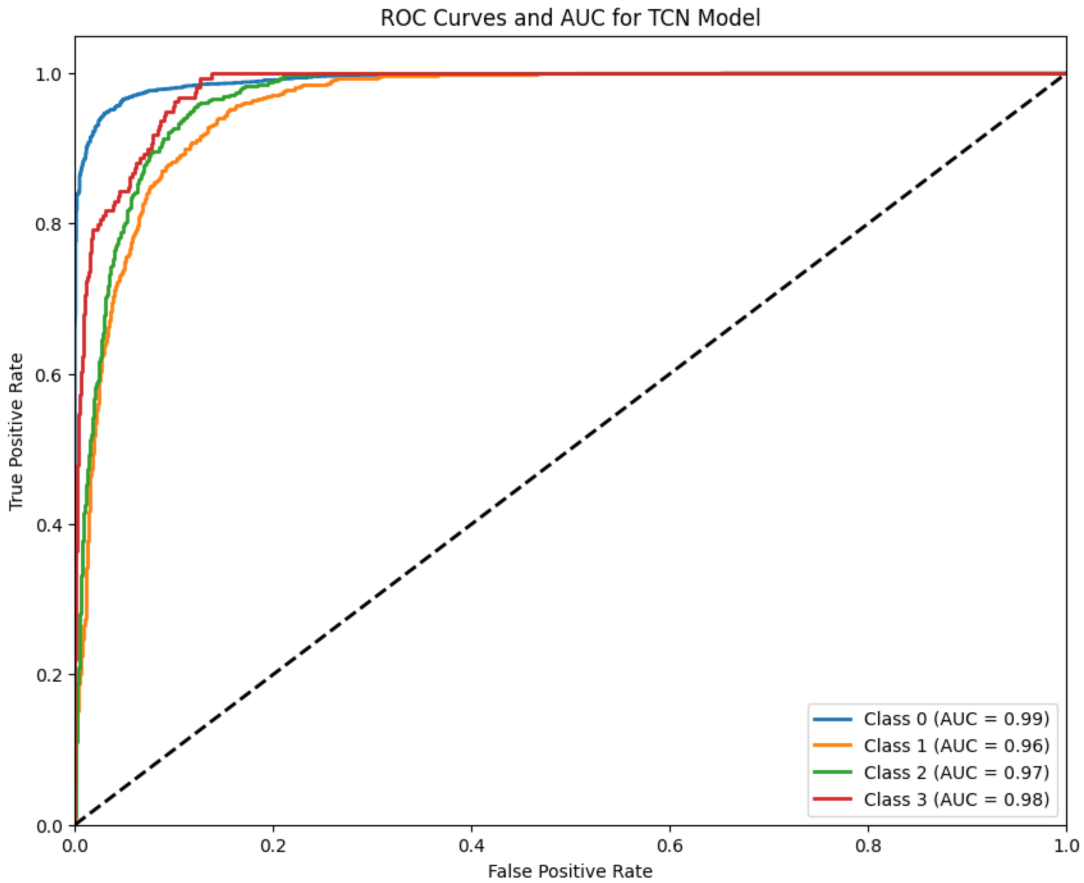
```
# Compile and train TCN model
tcn_model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
history_tcn = tcn_model.fit(X_train, y_train, validation_data=(X_val, y_val), epochs=epochs, batch_size=batch_size,
callbacks=[tcn_checkpoint_callback])

# Compile and train GTCN model
GTcn_model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
history_GTcn = GTcn_model.fit(X_train, y_train, validation_data=(X_val, y_val), epochs=epochs, batch_size=batch_size,
callbacks=[GTcn_checkpoint_callback])

# Compile and train BiGTcn model
BiGTcn_model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
history_BiGTcn = BiGTcn_model.fit(X_train, y_train, validation_data=(X_val, y_val), epochs=epochs, batch_size=batch_size,
callbacks=[BiGTcn_checkpoint_callback])

# Evaluate models on the test set
def evaluate_models(models, model_names, X_test, y_test):
    for i, model in enumerate(models):
        print(f"Evaluating {model_names[i]} on the test set:")
```

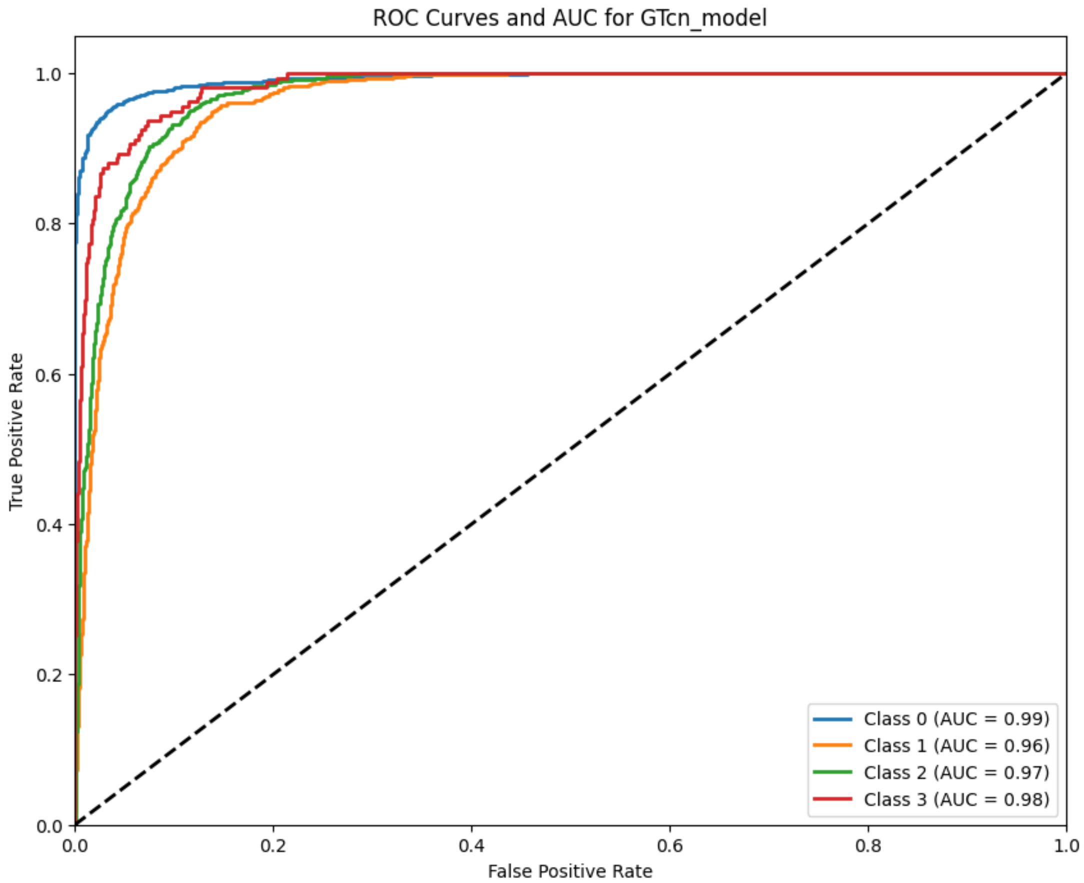
```
        test_loss, test_accuracy = model.evaluate(X_test, y_test, verbose=1)
        print(f"Test Loss: {test_loss:.4f}, Test Accuracy: {test_accuracy:.4f}\n")

# List of models and their names
models = [cnn_model, tcn_model, GTcn_model, BiGTcn_model]
model_names = ['1D-CNN Model', 'TCN Model', 'GTcn Model', 'BiGTcn Model']

# Evaluate all models on the test data
evaluate_models(models, model_names, X_test, y_test)

# Import necessary libraries for evaluation
from sklearn.metrics import classification_report, confusion_matrix, roc_auc_score
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np

# Function to predict on test data and compute classification report for all models
def evaluate_classification_report(models, model_names, X_test, y_test):
    for i, model in enumerate(models):
        y_pred = np.argmax(model.predict(X_test), axis=1)
```

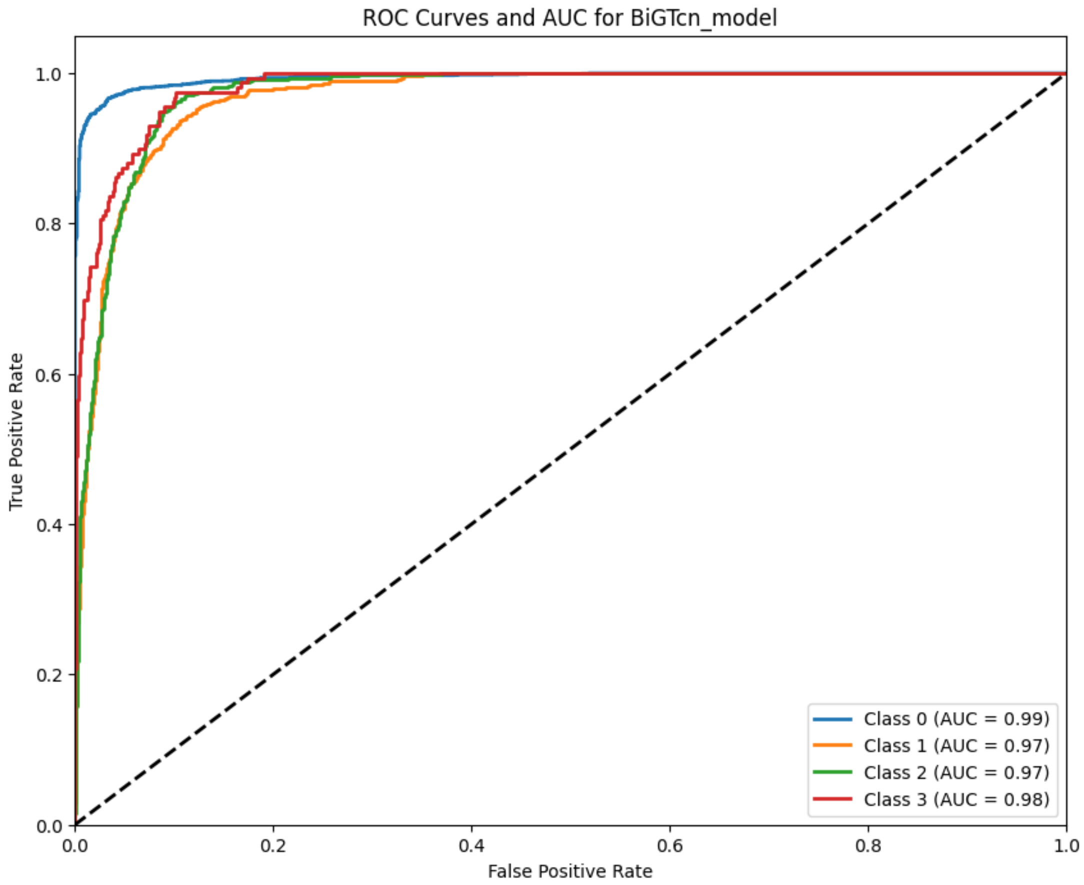
```
        y_true = np.argmax(y_test, axis=1)
        print(f"Classification Report for {model_names[i]}:")
        print(classification_report(y_true, y_pred))
        print("\n")

# Function to plot confusion matrix for all models
def plot_confusion_matrices(models, model_names, X_test, y_test):
    for i, model in enumerate(models):
        y_pred = np.argmax(model.predict(X_test), axis=1)
        y_true = np.argmax(y_test, axis=1)
        cm = confusion_matrix(y_true, y_pred)

        plt.figure(figsize=(6, 4))
        sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
        plt.title(f'Confusion Matrix for {model_names[i]}')
        plt.ylabel('True label')
        plt.xlabel('Predicted label')
        plt.show()
```

```
# Function to compute AUC for all models
def compute_auc(models, model_names, X_test, y_test):
    for i, model in enumerate(models):
        y_pred_proba = model.predict(X_test)
        auc = roc_auc_score(y_test, y_pred_proba, multi_class='ovr')
        print(f"AUC for {model_names[i]}: {auc:.4f}")

# List of models and their names
models = [cnn_model, tcn_model,GTcn_model, BiGTcn_model]
model_names = ['1D-CNN Model', 'TCN Model', 'GTcn_model', 'BiGTcn_model']

# Perform (Classification Report and Confusion Matrix)
evaluate_classification_report(models, model_names, X_test, y_test)
plot_confusion_matrices(models, model_names, X_test, y_test)

# Perform (AUC Calculation)
compute_auc(models, model_names, X_test, y_test)
```


ROC Curves and AUC for 1D-CNN Model
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
Class 0 (AUC = 0.99)
Class 1 (AUC = 0.93)
Class 2 (AUC = 0.95)
Class 3 (AUC = 0.96)


ROC Curves and AUC for TCN Model
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Class 0 (AUC = 0.99)
Class 1 (AUC = 0.96)
Class 2 (AUC = 0.97)
Class 3 (AUC = 0.98)


ROC Curves and AUC for GTcn_model
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Class 0 (AUC = 0.99)
Class 1 (AUC = 0.96)
Class 2 (AUC = 0.97)
Class 3 (AUC = 0.98)


ROC Curves and AUC for BiGTcn_model
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Class 0 (AUC = 0.99)
Class 1 (AUC = 0.97)
Class 2 (AUC = 0.97)
Class 3 (AUC = 0.98)