UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE INFORMÁTICA

Deep learning for reliable communication optimization on autonomous vehicles

Inês Martins Rocha

Mestrado em Engenharia Informática

Dissertação orientada por:
Prof. Doutor José Manuel da Silva Cecílio
Prof. Doutor Luís Miguel Ramos Bárbara Cunha Pinto

2025

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Prof. Doutor José Manuel da Silva Cecílio, from the bottom of my heart, for always being there for me, providing valuable guidance and assistance, and supporting me throughout the writing of this dissertation. His support and expertise were essential to the success of our work.

I also wish to thank my co-supervisor, Prof. Doutor Luís Pinto, for all the assistance he provided, the technical advice he offered, and for always being available to discuss ideas and address any concerns I had along the way. I want to thank Prof. António Casimiro for his assistance and guidance, which significantly enhanced this research. His expertise and insights greatly enhanced my dissertation. I would also like to thank Bakary Badjie, a PhD student, for his assistance with numerous technical problems and for generously offering his time and expertise. I would also like to thank the Faculty of Sciences and the Department of Informatics at the University of Lisbon for providing the academic environment and resources necessary for me to complete this work.

Many thanks to my family and close friends who have always believed in me, supported me through difficult times, and understood when challenges arose.

Abstract

Reliable network coverage prediction is essential to ensure continuous communication in autonomous vehicles, particularly in dense urban environments where buildings and other structures can significantly degrade signal quality. This dissertation proposes a hybrid deep learning architecture designed to accurately predict the Reference Signal Received Power (RSRP) in specific urban locations.

The developed model integrates three types of information: semantically encoded urban images derived from OpenStreetMap (OSM), tabular network metrics such as RSRQ (Reference Signal Received Quality), SNR (Signal-to-Noise Ratio), CQI (Channel Quality Indicator) and serving cell distance, as well as a physics-based path loss estimation based on the 3GPP TR 38.901 standard. Each data modality is processed by a specialised neural network, CNNs for spatial features and DNNs for tabular data before being fused into a unified regression model.

The approach was evaluated using real 4G LTE data collected in Cork, Ireland, following a 10-fold cross-validation scheme with strict modality alignment. Results show an average Root Mean Square Error (RMSE) of 2.77 dB with a standard deviation of $\pm$ 0.07 dB, outperforming purely physical models (e.g., 14.13 dB RMSE) and unimodal deep learning baselines (e.g., 3.81 dB with tabular-only inputs). Additional tests demonstrated spatial generalisation (RMSE ranging from 2.73 dB to 3.10 dB across regions), robustness to noise (e.g., 3.21 dB under degraded RSSI), and high sensitivity to relevant features (e.g., performance drop to 2.85 dB without CQI). The overall performance reached 2.87 dB RMSE, 1.98 dB MAE, and $R^2 = 0.93$.

By fusing physical modelling, geospatial context, and radio metrics, this hybrid approach proves effective and robust, offering a promising tool for autonomous vehicle applications and urban cellular network planning.

Keywords: Deep Learning, Autonomous Vehicles, RSRP Prediction, OpenStreetMap (OSM), Hybrid Learning

Resumo

A presente dissertação teve como principal objetivo o desenvolvimento de um modelo híbrido baseado em aprendizagem profunda (deep learning) para a previsão precisa do *Reference Signal Received Power* (RSRP), uma métrica fundamental para caracterizar a qualidade do sinal em redes móveis, em particular nas tecnologias LTE e 5G. Este tipo de previsão é especialmente crítico no contexto de veículos autónomos, cuja operação segura e eficiente depende de comunicações contínuas e fiáveis com infraestruturas externas. Em cenários urbanos densos, a presença de edifícios, árvores e outros obstáculos físicos introduz uma elevada variabilidade na propagação do sinal, tornando os modelos tradicionais baseados em fórmulas empíricas ou determinísticas insuficientes para garantir previsões robustas, especialmente em situações críticas para a segurança.

Com a crescente necessidade de soluções inovadoras para comunicação veicular e planeamento de redes móveis urbanas, propôs-se uma abordagem multimodal e informada fisicamente, que integra diferentes fontes de dados para melhorar a capacidade preditiva dos modelos baseados em *deep learning*. O modelo desenvolvido nesta dissertação explora três modalidades de dados distintas: (i) imagens semânticas urbanas derivadas de mapas de altura extraídos do *OpenStreetMap* (OSM), que capturam a geografia urbana e a morfologia tridimensional dos edifícios de forma detalhada; (ii) métricas tabulares de rede, como o RSRQ (*Reference Signal Received Quality*), SNR (*Signal-to-Noise Ratio*), CQI (*Channel Quality Indicator*), RSSI (*Received Signal Strength Indicator*), distância à célula servidora e *throughput*; e (iii) estimativas de perda de percurso baseadas no modelo físico 3GPP TR 38.901, amplamente reconhecido como referência para modelação de propagação em ambientes urbanos reais e simulados.

Cada uma destas modalidades é processada separadamente por uma sub-rede neural especializada: uma rede convolucional (CNN) para extrair características espaciais e visuais das imagens OSM; uma rede densa (DNN) para processar os atributos tabulares; e um módulo de normalização para incluir a componente baseada em modelos físicos. As saídas destes três módulos são posteriormente fundidas num bloco de regressão final, permitindo ao modelo aprender representações conjuntas que capturam tanto as regularidades estatísticas dos dados como o conhecimento prévio embutido nos modelos físicos. Esta fusão representa um dos principais contributos da dissertação, ao permitir que o modelo se beneficie simultaneamente da flexibilidade da aprendizagem profunda e da robustez das estimativas baseadas em física, tornando-se mais interpretável e confiável para aplicações críticas.

O modelo foi treinado e validado com base num conjunto de dados reais recolhido na cidade de Cork, Irlanda, em ambiente 4G LTE, que inclui medições reais de campo captadas por veículos em movimento, garantindo assim a representação de situações dinâmicas reais. A avaliação do desempenho foi realizada através de validação cruzada estratificada em 10 blocos (10-fold cross-validation), assegurando uma divisão justa e representativa das amostras, com alinhamento rigoroso entre as diferentes modalidades. Esta metodologia permitiu avaliar a capacidade de generalização espacial do modelo e medir a sua robustez em diferentes cenários urbanos, incluindo zonas com maior densidade de edifícios e áreas afastadas da célula base, onde a degradação de sinal é frequentemente observada.

Em termos de resultados, o modelo proposto alcançou um erro quadrático médio (RMSE) de aproximadamente 2.77 dB, com um desvio padrão de $\pm$ 0.07 dB. Estes valores representam uma melhoria significativa face aos modelos puramente físicos ou às abordagens baseadas apenas em métricas tabulares, cujos erros típicos se situam entre 3.04 dB e 6.3 dB, conforme reportado na literatura, nomeadamente no trabalho de Thrane et al. [1]. A análise detalhada revelou que o modelo mantém um desempenho estável mesmo em cenários desafiantes, como regiões altamente urbanizadas ou localizações a mais de 1 km da estação base, onde a variabilidade do sinal tende a ser mais elevada. Nestes contextos, o modelo conseguiu manter erros abaixo dos 2.90 dB, demonstrando uma elevada capacidade de adaptação a variações espaciais e ambientais, fator essencial em redes móveis modernas.

Adicionalmente, foram realizados vários testes complementares que reforçam a robustez e generalização do modelo. A comparação com um modelo puramente físico baseado apenas na perda de percurso revelou um erro médio de 14.13 dB, enquanto o modelo híbrido atingiu apenas 2.77 dB, evidenciando uma melhoria significativa. Testes com variações da arquitetura mostraram que a remoção de componentes como as imagens contextuais ou métricas tabulares degradam o desempenho (e.g., 3.81 dB apenas com dados tabulares; 2.85 dB com imagens preto e branco), comprovando o contributo de cada modalidade. Além disso, o modelo demonstrou resiliência face a alterações geográficas, mantendo o RMSE entre 2.73 dB e 3.10 dB em diferentes regiões espaciais, e uma estabilidade notável mesmo em cenários ruidosos, onde o erro aumentou moderadamente para 3.21 dB quando métricas como o RSSI foram degradadas. A importância de atributos como o CQI também foi evidenciada, com a sua remoção a causar um aumento no erro para 2.85 dB. Finalmente, o modelo alcançou um RMSE médio global de 2.87 dB, MAE de 1.98 dB e um coeficiente de determinação $R^2 = 0.93$, reforçando a sua eficácia preditiva em múltiplas métricas.

Para além dos resultados quantitativos obtidos, as análises complementares realizadas permitiram uma compreensão mais profunda da generalização espacial e da contribuição de cada modalidade de dados. A consistência dos desempenhos observados em diferentes regiões geográficas e faixas de distância evidencia a capacidade do modelo em lidar com cenários urbanos variados, mesmo sob condições menos favoráveis. Do ponto de vista metodológico, a avaliação isolada de cada modalidade reforça a importância da integração multimodal, demonstrando que abordagens híbridas proporcionam ganhos não apenas em desempenho, mas também em robustez e interpretabilidade. Estes resultados sustentam a relevância de modelos fisicamente informados no planeamento de redes móveis e em aplicações críticas como veículos autónomos, onde decisões fiáveis dependem fortemente da previsibilidade das condições de sinal.

Do ponto de vista técnico, a arquitetura desenvolvida foi cuidadosamente otimizada, incluindo a seleção de hiperparâmetros, regularização para evitar overfitting e utilização de técnicas modernas como early stopping e normalização de batches. O pipeline de pré-processamento foi igualmente relevante, com um sistema automatizado para converter mapas OSM em imagens semânticas com codificação RGB e preto e branco, integração dos dados tabulares e normalização dos valores físicos com base em benchmarks da literatura. Esta automatização facilita a replicação do modelo e torna o sistema mais escalável e adaptável a novos cenários.

Apesar dos resultados encorajadores, algumas limitações foram identificadas. Em primeiro lugar, o conjunto de dados utilizado foi recolhido apenas na cidade de Cork, o que limita a generalização do modelo para outras regiões com morfologias urbanas distintas. Em segundo lugar, o custo computacional do treino multimodal é elevado, devido à integração de mapas semânticos, atributos tabulares e priors físicos, o que pode dificultar a aplicação do modelo em cenários com restrições de recursos. Além disso, a abordagem depende fortemente da qualidade e atualidade dos mapas OSM, que podem conter erros ou

dados desatualizados, introduzindo imprecisões na representação espacial. A variabilidade temporal das condições urbanas, como mudanças em infraestruturas, não é capturada por mapas estáticos, o que pode causar discrepâncias entre os dados de entrada e as condições reais de propagação. Por fim, a sensibilidade do modelo a ruído nas métricas tabulares (como RSRQ e RSSI) pode comprometer a fiabilidade das previsões em cenários com medições ruidosas ou sensores imprecisos. Estas limitações reforçam a importância de investigações futuras focadas em conjuntos de dados mais amplos, melhorias na eficiência computacional e estratégias de robustez para lidar com mapas incompletos ou ruído nos dados.

Do ponto de vista prático, o modelo proposto apresenta várias aplicações potenciais. No domínio dos veículos autónomos, pode ser integrado em sistemas de planeamento de rotas e decisão de conectividade, ajudando os veículos a prever zonas de baixa cobertura e a adaptar os seus comportamentos de comunicação de forma proativa. No planeamento de redes móveis, pode servir como ferramenta de suporte à decisão para operadores, permitindo estimar zonas problemáticas e otimizar a colocação de antenas ou a configuração de parâmetros de rede. Pode ainda ser útil em contextos de planeamento urbano, como avaliação de cobertura em projetos de expansão de infraestrutura.

Conclui-se, portanto, que a fusão de dados visuais, tabulares e físicos numa arquitetura de aprendizagem profunda representa uma solução eficaz e inovadora para os desafios contemporâneos de previsão de sinal em redes móveis. A abordagem proposta nesta dissertação não só melhora significativamente a precisão das previsões como também introduz uma nova perspetiva sobre a integração de modelos determinísticos e estatísticos na era do *deep learning*. Os resultados obtidos validam a hipótese de que abordagens informadas fisicamente podem contribuir para sistemas mais robustos, interpretáveis e adaptáveis, com impacto direto na qualidade dos serviços de comunicação e na segurança de sistemas autónomos.

Palavras-chave: veículos autónomos, previsão do RSRP, deep learning, OpenStreetMap, modelo híbrido.

Contents

List of Figures

List of Tables

Acronyms

- **DNN** Dense Neural Network
- **RSRP** Reference Signal Received Power
- **RSRQ** Reference Signal Received Quality
- **SNR** Signal-to-Noise Ratio
- **CQI** Channel Quality Indicator
- **RSSI** Received Signal Strength Indicator
- **CNN** Convolutional Neural Network
- **OSM** Open Street Map
- **RMSE** Root Mean Square Error
- **AV** Autonomous Vehicle
- **FCN** Fully Connected Network
- **GNN** Graph Neural Network
- **LoS** Line-of-Sight
- **LTE** Longe Term Evolution
- **DL** Deep Learning
- **3GPP** 3rd Generation Partnership Project
- **FCN** Fully Connected Network
- **RMSE** Root Mean Squared Error
- **VGG** Visual Geometry Group
- **ResNet** Residual Neural Network
- **KPI** Key Performance Indicator
- **ANN** Artificial Neural Network
- **QoS** Quality Of Service

- **IoT** Internet of Things
- **CSI** Channel State Information
- **RGB** Red, Green, and Blue
- **LeakyReLU** Leaky version of a Rectified Linear Unit
- **Stdev** Standard Deviation
- **V2X** Vehicle-to-Everything
- **V2V** Vehicle-to-Vehicle
- **V2I** Vehicle-to-Infrastructure
- **V2P** Vehicle-to-Pedestrian
- **V2C** Vehicle-to-Cloud
- **C-V2X** Cellular Vehicle-to-Everything
- **URLLC** Ultra-Reliable Low-Latency Communication
- **mMTC** Massive Machine-Type Communication
- **eMBB** Enhanced Mobile Broadband
- **CAVs** Connected and Autonomous Vehicles
- **NLOS** Non-Line-Of-Sight
- **ViT** Vision Transformers
- **MLPs** Multi-Layer Perceptrons Transformers
- **KPIs** Key Performance Indicators
- **dB** Decibels
- **MAE** Mean Absolute Error
- **GPU** Graphics Processing Unit
- **LiDAR** Light Detection and Ranging
- **GIS** Geographic Information System

Chapter 1

Introduction

Reliable wireless communication is an essential enabler for the efficient operation of AVs (Autonomous Vehicles). However, maintaining reliable connectivity in urban environments presents significant challenges due to the complex propagation dynamics of radio signals.

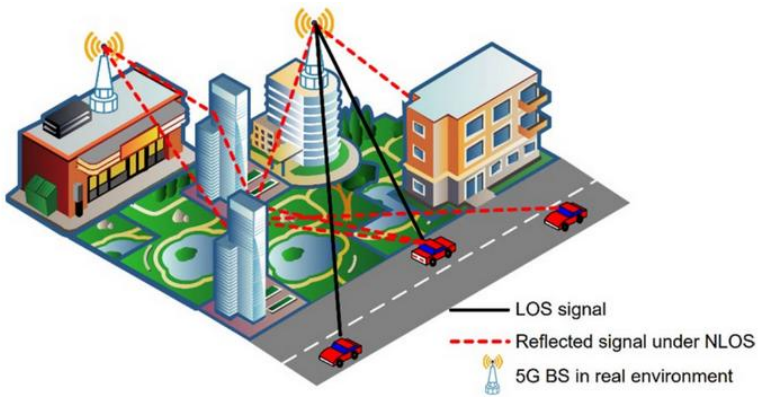

Figure 1.1: Illustration of a real-world urban scenario showing the effects of obstacles such as buildings and vegetation on signal propagation. Direct LOS paths can be obstructed, forcing signals to follow NLOS paths through reflection or diffraction, leading to signal degradation.

As illustrated in Figure 1.1, signals transmitted from a base station to a vehicle can be significantly impacted by physical obstructions such as buildings, trees, and other urban structures. These obstacles introduce shadowing, attenuation, multipath reflections, and NLOS (Non-Line-Of-Sight) conditions, all of which can severely degrade signal quality. Additionally, the distance between the transmitter and receiver further influences signal strength. Longer distances typically lead to higher path loss and greater variability, especially in dense urban settings.

These factors collectively make the task of accurately predicting signal quality particularly complex in urban scenarios, where AVs must maintain continuous communication links to support safety-critical applications such as cooperative perception, real-time decision-making, and dynamic path planning.

Challenges are not limited to urban environments: similar signal degradation can occur in suburban or rural areas due to natural obstacles such as foliage or terrain variations, which block the direct LOS between transmitter and receiver.

Given these constraints, traditional signal propagation models, such as R. Mardeni [2], Akhpashev and Andreev [3] or 3GPP-based derivations, offer only limited accuracy. These models often fail to capture the detailed spatial, contextual, and structural complexity inherent in real-world environments. In response, recent research has explored deep learning-based methods that leverage signal metrics and

contextual data to improve prediction performance [1, 4, 5]. Additionally, the integration of semantic information from open-source geographic databases such as OpenStreetMap enhances coverage prediction in scenarios where high-resolution satellite imagery or 3D maps are unavailable [6].

This dissertation extends a previous scientific study by the supervisor, deepening the theoretical foundation, expanding the experimental evaluation, and presenting a refined multimodal deep learning architecture that integrates radio, spatial, and physical features for signal strength prediction in complex urban contexts.

1.1 Motivation

The rapid evolution of wireless communication technologies has made reliable and low-latency connectivity a cornerstone of modern intelligent transportation systems. AVs, in particular, rely heavily on V2X (Vehicle-to-Everything) communication to enable functionalities such as cooperative perception, which allows vehicles to share sensor information (e.g., LiDAR (Light Detection and Ranging), camera feeds) to perceive objects beyond their own line-of-sight; collision avoidance, where vehicles exchange position, speed, and trajectory data to anticipate and prevent potential crashes; and real-time route planning, which enables dynamic path optimization based on current traffic, road hazards, or infrastructure signals.

Ensuring stable and predictable network performance in urban environments remains a significant challenge. This situation is particularly critical in densely populated urban areas, where frequent handovers and abrupt changes in channel conditions occur, making it challenging to ensure the QoS (Quality Of Service) required for efficient operation. Accurately predicting signal strength in such environments is thus a non-trivial but essential task.

Traditional empirical and theoretical models, such as R. Mardeni [2] and 3GPP-derived formulations [7], are commonly used but often fail to capture the spatial and contextual intricacies of real-world urban environments. To address these limitations, recent approaches have turned to deep learning techniques that can model complex nonlinear relationships by leveraging heterogeneous data sources, including signal measurements and semantic features from open urban datasets like OpenStreetMap.

This dissertation builds on the state-of-the-art and proposes a multimodal deep learning architecture for signal strength prediction in complex urban scenarios. By integrating domain knowledge with data-driven approaches, the goal is to design predictive models that are accurate, generalisable, and applicable to real-world deployments, paving the way for more robust and resilient V2X systems in autonomous driving contexts.

1.2 Objectives

The primary objective of this dissertation is to improve the accuracy, generalisability, and applicability of wireless signal strength prediction in autonomous vehicular environments by designing, implementing, and evaluating a hybrid deep learning framework informed by physical modelling. The proposed approach aims to overcome the limitations of conventional propagation models by leveraging multimodal data, thereby contributing to the advancement of communication-aware autonomy in intelligent transportation systems.

To achieve this, the dissertation defines the following specific objectives:

- **Analyse communication challenges in vehicular environments:** Identify and characterise key

factors that affect signal propagation in urban scenarios, such as multipath fading, building shadowing, interference, and distance-dependent attenuation and assess their impact on network reliability and QoS.

- **Design a multimodal hybrid deep learning architecture:** Develop a model that integrates: (i) semantic spatial features extracted from OpenStreetMap-derived heightmaps using convolutional neural networks (CNNs); (ii) tabular radio-frequency indicators (RSRP, RSRQ, SNR, throughput, distance to serving cell) processed by a dense neural network (DNN); and (iii) physics-based priors informed by the 3GPP TR 38.901 path-loss model.
- **Promote computational efficiency:** Incorporate model optimisation strategies such as early stopping, batch normalisation, and architecture simplification to ensure feasibility in large-scale deployments while preserving training stability and inference efficiency.
- **Evaluate performance using real-world datasets:** Conduct a comprehensive evaluation on large-scale drive-test datasets using 10-fold cross-validation. Analyse RMSE (Root Mean Square Error) performance across varying propagation conditions, distance quartiles, and spatial regions of the city.
- **Benchmark against existing methods:** Compare the proposed hybrid framework with traditional path-loss models and data-driven baselines to quantify improvements in prediction accuracy, robustness, and generalisation capacity.
- **Assess stability across spatial scales and environments:** Investigate whether the model maintains consistent predictive performance across near-field and far-field ranges, as well as across distinct urban morphologies, to demonstrate scalability and transferability.
- **Ensure reproducibility and methodological transparency:** Utilise open urban data (e.g., OpenStreetMap), clearly documented preprocessing pipelines, and modular software design to facilitate reproducibility and future extensions within the V2X research community.
- **Explore applicability to autonomous vehicle systems:** Interpret results in the context of autonomous driving, focusing on aspects such as handover reliability in dense deployments, support for V2X applications, and the ability to uphold QoS in specific use cases.

This set of objectives aims to establish a reliable, interpretable, and deployable framework for RSRP (Reference Signal Received Power) prediction that bridges domain knowledge and data-driven modelling. The proposed approach contributes toward the development of intelligent, resilient, and communication-aware autonomous vehicular networks.

1.3 Contributions

This work addresses the challenges of reliable signal strength prediction in urban vehicular environments by proposing a novel multimodal hybrid deep learning framework. The key contributions are summarised below:

1. **Semantic encoding of urban layouts:** This work introduces a novel method for encoding urban environments using semantically annotated images derived from OSM (Open Street Map), where

building heights and types are represented through a structured colour scheme. This enables efficient spatial representation without reliance on satellite imagery.

2. **Multimodal deep learning architecture:** A hybrid DL (Deep Learning) framework that integrates tabular signal data (RSRP, RSRQ, SNR, CQI, etc.), semantic spatial images, and theoretical path-loss estimates derived from the 3GPP 38.901 model is proposed. The architecture comprises a dense neural network, a CNN, and a fusion layer that jointly process signal propagation and urban features.

3. **Improved signal prediction and coverage:** The proposed architecture significantly enhances RSRP prediction accuracy compared to both traditional empirical models (e.g., 3GPP path loss) and conventional deep learning baselines. Evaluation on real-world 4G LTE (Longe Term Evolution) data from Cork, Ireland, shows a substantial reduction in RMSE, confirming the effectiveness of integrating spatial priors and domain-informed model fusion.

4. **Generalisation and practicality for autonomous vehicle systems:** The proposed method presents a general solution suitable for deployment in resource-constrained environments, making it applicable for communication-aware navigation in AVs. It eliminates dependence on high-resolution satellite imagery while maintaining physical interpretability through path-loss integration.

5. **Hybrid architecture:** The model integrates a CNN for spatial feature extraction, a DNN for tabular KPIs (Keys Performances Indicators), and a physics-based path loss baseline, achieving both interpretability and predictive accuracy.

6. **Robust validation:** The proposed model is evaluated using 10-fold cross-validation with strict alignment across modalities, achieving an average RMSE of 2.77 dB (± 0.07 dB). This demonstrates superior generalisation compared to both standalone deep learning approaches and conventional physical models.

7. **Reproducibility and Adaptability:** To ensure reproducibility, all components of the system, such as the creation of OSM-based pictures, feature pre-processing, and the deep learning architecture, are scrupulously documented and modular. The proposed methodology is flexible and can be adapted to various urban configurations, radio technologies (e.g., 5G), and geographies, thereby laying a foundation for future research in V2X communication and autonomous driving.

1.4 Document Structure

This dissertation is organised into seven chapters:

- **Chapter 2** provides the necessary background to contextualise the research problem. It introduces the concepts of autonomous vehicles, V2X communication systems, and wireless signal propagation. The chapter also covers traditional and data-driven propagation models, discusses the role of geospatial data such as OpenStreetMap, and outlines relevant deep learning techniques that have been used in similar tasks. This foundational knowledge is essential to understanding the interdisciplinary nature of the work.

- **Chapter 3** presents a comprehensive survey of related work, focusing on signal strength prediction and hybrid learning approaches. It reviews conventional empirical models, purely data-driven models, and recent physics-informed machine learning strategies. The chapter also identifies gaps in current methods, particularly in terms of scalability, generalisation, and multimodal integration, thereby motivating the proposed solution.
- **Chapter 4** details the data sources, preprocessing procedures, and overall methodological pipeline adopted in the study. It describes how multimodal data, such as semantic maps, tabular KPIs, and path-loss estimates, are collected and aligned. The chapter also introduces the proposed hybrid model and explains how each modality contributes to the predictive task.
- **Chapter 5** focuses on the technical implementation of the proposed architecture. It explains the configuration of the neural networks (CNN, DNN, and fusion layer), training procedures, and optimisation strategies employed. Tools, frameworks, and hardware setups used to support the development and training of the model are also documented in detail.
- **Chapter 6** reports the results of the experimental evaluation, analysing performance across various propagation conditions, urban morphologies, and distance ranges. It presents metrics such as RMSE and discusses the model's generalisation ability through 10-fold cross-validation. Comparative analysis against baseline models is also included to demonstrate improvements.
- **Chapter 7** interprets the results within the broader context of communication-aware autonomy. It assesses the robustness and practicality of the approach, discussing its scalability, transferability to unseen areas, and potential integration into AV systems. Limitations and potential threats to validity are also critically examined.
- **Chapter 8** summarises the key findings of the dissertation and reflects on the research contributions. It revisits the initial objectives, assesses whether they have been met, and outlines open challenges. The chapter concludes with suggestions for future work, such as adapting the framework to 5G technologies or expanding the approach to include mobility and temporal dynamics.

Chapter 2

Background

Thanks to advancements in artificial intelligence, communication systems, and geographic databases, the development of self-driving vehicles has sparked revolutionary changes within the transport industry. However, to ensure their consistent and reliable operation, autonomous vehicles must navigate a complex network of mobility, environmental diversity, and communication prediction challenges.

This section outlines the essential concepts and core technology supporting our investigation. Understanding the dependence of autonomous vehicles on effective communication networks, the importance of geographic data for network performance, and the application of deep learning methods for predicting received signal strength is essential for developing the methodology presented in this dissertation.

2.1 Autonomous Vehicles

Autonomous vehicles represent a significant technological advancement in modern transportation systems. The goal of autonomous vehicles is to reduce traffic congestion, improve road safety, and enhance mobility for vulnerable groups, including the elderly and those with disabilities, by eliminating the need for human intervention. These vehicles operate safely in dynamic environments using a complex combination of sensors, machine learning algorithms, real-time data processing, and reliable communication systems.

The ability of autonomous cars to detect and evaluate their surroundings through the integration of LiDAR, radar, ultrasonic sensors, cameras, and GPS data is crucial for their safe and effective operation. The onboard processing units analyse these sensory inputs to provide critical functions, such as collision avoidance, obstacle recognition, route planning, and lane identification. To ensure high reliability and minimal latency in changing traffic situations, the autonomous decision-making pipeline must operate within strict real-time constraints.

Autonomous vehicles use semantic environmental comprehension, localisation technology, and high-definition mapping to provide precise positioning in complex metropolitan environments. In this scenario, communication methods are crucial. V2X communication facilitates the exchange of information between V2V (Vehicle-to-Vehicle), V2I (Vehicle-to-Infrastructure), V2P (Vehicle-to-Pedestrian), and V2C (Vehicle-to-Cloud). This connectivity is essential for enabling cooperative driving, traffic control, and intelligent transportation services [8], [9], [10].

Despite the advantages of driverless vehicles, they face significant challenges when manoeuvring in busy urban areas. Environmental obstacles, such as multipath signal degradation, interference from tall structures, and fluctuating network conditions, can disrupt vehicle-to-infrastructure communication and diminish system effectiveness. These issues highlight the need for robust communication frameworks

that incorporate predictive intelligence and optimisation strategies to enhance the safety and dependability of autonomous driving systems.

2.2 Vehicular Communication Systems and Mobile Networks

Vehicular communication systems are crucial for the advancement of autonomous cars, enabling real-time data transmission among vehicles, devices, and network nodes. These systems are often categorised into three communication modes: V2X, V2V, and V2I. Each paradigm enables specific aspects of vehicular interaction, including collision avoidance and platooning in V2V communication, as well as traffic light coordination and cloud-based data exchange in V2I and V2X communication [10].

Contemporary vehicle networks predominantly utilise cellular technologies, such as LTE and the forthcoming 5G networks, which offer the requisite bandwidth and latency for safety-critical applications. C-V2X (Cellular Vehicle-to-Everything) employing 5G enables URLLC (Ultra-Reliable Low-Latency Communication), mMTC (Massive Machine-Type Communication), and eMBB (Enhanced Mobile Broadband). These competencies are crucial for addressing the intricate communication requirements of CAVs (Connected and Autonomous Vehicles) [11]. Looking forward, 6G is anticipated to bring further innovations through AI-native network architectures and the integration of high-frequency bands, such as millimetre waves.

To assess and optimise the quality of vehicular communication links, several key performance indicators (KPIs) are typically monitored. These include:

- **RSRP** – a measure of the received signal strength;
- **SNR** – representing the signal quality relative to background noise;
- **CQI** – an estimate of the link quality reported by the receiver to the base station;
- **RSRQ** – combining signal strength and quality;
- **Throughput and latency** – essential for evaluating end-to-end communication efficiency.

These metrics serve as foundational inputs for the learning algorithms presented in this dissertation and are critical to enabling adaptive, resilient, and predictive communication optimisation in real-time vehicular environments.

2.3 Propagation Models and Signal Prediction

Precise signal propagation modelling is crucial for the design and optimisation of wireless communication in automotive networks. Conventional methods, such as the COST-231 and Hata models, along with standardised guidelines like 3GPP 38.901, provide analytical frameworks for estimating route loss based on variables including distance, frequency, and environmental type. These models are often utilised because of their mathematical simplicity and broad applicability in mobile communication planning.

Nevertheless, when implemented in complex metropolitan settings, these models often fail to encompass the full range of actual scenarios. The deterioration of signal strength caused by multipath effects, building shadowing, and dynamic barriers, such as moving vehicles or pedestrians, presents considerable

hurdles. In densely populated metropolitan environments, radio frequency propagation exhibits significant non-linearity and geographic variability, rendering conventional models inadequate or excessively generalised [12].

Additionally, predicting real-time signals is an exceptionally challenging task. In situations where autonomous vehicles need to adjust to fluctuating connectivity circumstances in real-time, it is essential to predict signal attributes such as RSRP or SNR (Signal-to-Noise Ratio) with minimal latency and high dependability. Legacy models, which generally require prior calibration or set environmental factors, are ill-suited for this dynamic process.

Recent studies have initiated the investigation of data-driven approaches employing real-world datasets collected under various mobility and signal conditions to alleviate these limitations. The dataset provided by Raca et al. [13] includes contextual information such as location, speed, and environmental factors, facilitating a more comprehensive and adaptable basis for predictive modelling.

The limitations of traditional propagation models in urban vehicular scenarios underscore the need for more adaptive and learning-based signal prediction methods, an area explored in the next section through the use of deep learning strategies.

2.4 Geospatial Data and Urban Modelling

Accurate modelling of wireless signal propagation in urban vehicular environments requires detailed knowledge of the surrounding geography, infrastructure, and spatial topology. Traditional signal prediction models often treat urban space as homogeneous, overlooking critical environmental features such as building density, road orientation, terrain variation, and vegetation. These elements, however, have a direct impact on signal attenuation, reflection, and diffraction, especially in NLOS scenarios.

To bridge this gap, recent research has turned to the use of geospatial and semantic data sources. Platforms such as OSM provide access to open-source, human-annotated maps that include road layouts, building footprints, land usage categories, and other contextual metadata. Unlike high-resolution satellite imagery, OSM data offers standardised and structured information that can be easily processed by machine learning models, enabling large-scale integration into communication prediction systems [6], [14].

Geographical features derived from OSM, such as building height, street width, and semantic class (e.g., residential, commercial, industrial), can be encoded into feature maps or used as direct inputs for deep learning models. Studies have demonstrated that incorporating such contextual urban data significantly improves the accuracy and spatial resolution of signal strength and path loss predictions [15], [16].

Moreover, the semantic segmentation of urban scenes, either from satellite images or GIS (Geographic Information System) layers, enables the generation of location-specific signal prediction maps, allowing autonomous vehicles to anticipate weak coverage zones and adapt their behaviour proactively. This integration of spatial intelligence into communication models is crucial for enabling real-time adaptability and enhancing the robustness of AV operations in heterogeneous environments.

2.5 Deep Learning for Wireless Communication

The growing need for reliable wireless communication and the rising intricacy of mobile networks have generated interest in employing DL (Deep Learning) for signal prediction tasks. In many non-linear contexts, deep learning models have demonstrated considerable expertise in estimating wireless signal

parameters, including RSRP, SNR, and route loss. These tactics exceed traditional analytical models by explicitly delineating spatial and contextual patterns from actual data [4], [17], [18].

A variety of deep learning architectures have been examined to achieve this objective. Spatial feature extraction related to signal variance in urban environments is often performed using CNNs to examine geographical or satellite imagery. MLPs (Multi-Layer Perceptrons Transformers) have been utilised temporarily to model the relationships among scalar inputs, such as distance, coordinates, and network configuration. Recent research has increasingly used hybrid models that combine CNNs with fully connected layers or temporal models to improve prediction generalisation across various environments and applications [19].

The adaptability of deep learning models is a significant advantage. Deep learning-based predictors can be optimised using localised data to improve context-specific precision, unlike static propagation models. For instance, Jakob Thrane [4] devised a model-aided deep learning approach that integrates physical knowledge into the training pipeline to forecast path loss at 2.6 GHz. Cabric [17] employed 3D maps and spatial features to forecast signal strength, while Alsamhi et al. [18] employed drone-based communication to predict the optimal signal strength in dynamic IoT (Internet of Things) environments.

These developments provide a solid foundation for implementing DL-based frameworks in vehicular communication, particularly in urban settings with high signal variability. Additionally, they advocate for the implementation of deep models in the proposed solution of this dissertation to improve the connectivity of autonomous vehicles.

Chapter 3

Related Work

Accurately forecasting the RSRP in urban environments remains a challenging task due to the intrinsic unpredictability caused by dense infrastructure, complex urban layouts, and dynamic mobility. These elements introduce rapid variations in signal propagation, leading to frequent disruptions in connectivity. Such disruptions are particularly problematic for AV systems, which rely on stable, low-latency links to make timely and safety-critical decisions in real-time traffic scenarios.

Empirical models like the R. Mardeni and 3GPP 38.901 path loss equations are widely used in wireless planning for their simplicity and low computational cost [3, 4]. However, their rigid formulations limit adaptability in diverse urban contexts. To address this, several studies have proposed model calibration strategies aimed at improving accuracy in dense environments. For instance, R. Mardeni [2] recalibrated the COST-231 model to improve alignment with LTE measurements in highly urbanised scenarios. Other works, such as Thrane et al. [1], Alsamhi et al. [18], have integrated 3GPP priors into deep learning models to enhance physical interpretability and generalisation. Thao T. Nguyen [5] proposed a hybrid framework that combines 3GPP-based equations with CNNs and FCNs (Fully Convolutional Networks) to predict 7 GHz path loss. Their method incorporates satellite imagery and geometric features to improve accuracy, achieving sub-5 dB RMSE. However, it relies heavily on high-resolution imagery and precise geolocation data, which impedes scalability in under-instrumented regions.

Muhammad Brata [20] explored CNN-based path loss estimations using transfer learning (VGG-16, ResNet-50) for sub-6 GHz and mmWave frequencies. While their models outperform empirical baselines, they are constrained by exclusive dependence on satellite data and omit physical signal-level indicators (e.g., RSRQ, SNR). In contrast, the current approach combines semantic maps from OSM, structured signal features, and deterministic models to create a unified, scalable prediction pipeline that enhances spatial resilience.

Several authors have adopted machine learning models that integrate signal characteristics and spatial priors. Majd Alkorabi [21] provides a broad survey of deep learning applications in vehicular communication, identifying ongoing challenges such as the need for large annotated datasets, limited model generalisation, and real-time processing complexity in AV platforms. Our proposed multimodal framework addresses these concerns by fusing structured radio metrics, semantic urban maps, and physics-informed models, enabling effective deployment even in resource-constrained environments.

Other CNN-based architectures, such as those by Jakob Thrane [4], Minovski et al. [22], have shown improvements compared to traditional models but remain reliant on satellite and 3D data, limiting both semantic depth and deployment flexibility. In contrast, our methodology leverages publicly accessible OSM maps to provide a structured but lightweight semantic representation of urban environments.

The use of open-source geographic data presents a promising alternative. A pertinent contribution is

provided by Jamie Mersh [23], who investigated the application of geographic information and statistical learning methods to forecast wireless coverage in urban settings. Their method integrated topographical and architectural data to estimate signal propagation, showing significant improvement over empirical models. However, their reliance on high-quality datasets reduces the feasibility of applying them in less instrumented environments. Our work builds on this by employing semantic OSM maps, which preserve urban semantics without the need for high-resolution input.

Cabric [17] used semantic maps from OSM to depict urban configurations, including structural components such as building footprints. Although less resource-intensive than satellite data, most of these models still handle visual and tabular inputs separately, hence lacking multimodal integration.

Recent researchers have investigated digital twin architectures to enhance handover reliability in 5G environments through virtual replication of the network infrastructure. Umar D. Maiwada* [24] applied digital twins to emulate network states and facilitate proactive resource distribution. While these systems highlight semantic environment modelling and multi-source integration, their emphasis lies on macro-level orchestration and often requires real-time telemetry and centralised control. Conversely, our model leverages lightweight, OSM-derived spatial abstraction suitable for scalable and practical AV applications.

Multi-task learning and uncertainty-aware models have also been proposed to enhance spatial prediction. Foliadis et al. [25] introduced a deep fusion architecture that combines image-based and sensor-derived inputs. Their uncertainty-driven strategy improves resilience by selectively using modality-specific information based on confidence metrics. However, reliance on panoramic visual input complicates deployment in limited-resource settings. Our solution prioritises semantically encoded OSM data, offering a more efficient alternative.

Hybrid learning strategies have been explored to incorporate environmental priors with learned signal representations. Claussmann et al. [9] emphasised the importance of combining spatial and temporal inputs in vehicular networks. However, their models primarily focus on signal throughput estimation and often overlook semantic environmental cues. This omission limits their applicability in scenarios requiring robust spatial awareness, such as autonomous vehicle navigation. In contrast, our framework explicitly integrates structured semantic information from urban environments, ensuring better context-awareness and supporting more resilient communication in real-world conditions.

From a system-level perspective, Badue, Guidolini, and Carneiro [8] investigated communication failures in AVs, highlighting the need for resilient signal prediction in the presence of network instability. These findings support integrating spatial, physical, and signal-level information into a unified prediction framework.

Taheri et al. [26] reconceptualised RSRP prediction as an optimisation task using real-world drive-test data and simplified physical models (e.g., Friis parameters). Although accurate, this method demands extensive data collection and neighbourhood calibration, limiting its applicability. Our framework avoids these limitations by using semantic and deterministic priors to guide learning.

In lightweight modelling, Vishaka Basnayake [27] proposed a GNN-based method that infers spatial dependencies via attention mechanisms, eliminating the need for geometric maps or complex propagation modelling. Their unstructured approach supports generalisation in instrumentally sparse regions, aligning with our model's goal of scalable and interpretable signal prediction.

In summary, prior work demonstrates a spectrum of modelling strategies, ranging from empirical equations and DL to GNNs and digital twins. Many still depend on data-heavy inputs or lack urban semantic integration. Our proposed multimodal framework fills this gap by fusing semantically encoded OSM imagery, deterministic physical models, and tabular KPIs into a compact, scalable, and

interpretable architecture designed for real-world AV communication scenarios.

Name/Reference	Goal	Machine Learning Model	Results
Jakob Thrane [4]	Improve path loss prediction in mobile systems	CNN	Errors of 1 dB (811 MHz) and 4.7 dB (2630 MHz)
Akhpashev and Andreev [3]	Calibrate COST 231 Hata model using LTE urban data	N/A	Correction factor of 13.3 dBm applied; improved coverage prediction accuracy
Minovski et al. [22]	Benchmark network slices' throughput performance in 5G and evaluate QoS in industrial IoT use cases	Random Forest, Support Vector Machine (SVM), Neural	93% R^2 accuracy for LTE and 84% R^2 accuracy for non-standalone 5G networks
Cabric [17]	Predict spatial signal strength in urban environments using minimal prior data	DNN	MAE: 6.95 dBm (uniform) and 11.11 dBm (trajectory)
Umar D. Maiwada* [24]	Propose a digital twin network to improve handover performance in Malaysian 5G scenarios	N/A	Improved handover, energy efficiency, and network reliability
Foliadis et al. [25]	Joint image- and sensor-based positioning with uncertainty-based fusion	Multi-task deep learning + confidence fusion	High accuracy and robustness
Claussmann et al. [9]	Review motion planning in highway AVs	N/A	Organized taxonomy; highlighted real-time/safety issues
Badue, Guidolini, and Carneiro [8]	Survey autonomous vehicle architectures and techniques	Multiple (DL in perception and decision layers)	Improved understanding of AV modules; guidance for robust deployment strategies
Taheri et al. [26]	Predict RSRP from test data	N/A	Low error; no cell location needed
Vishaka Basnayake [27]	Evaluate 5G V2N for V2X	N/A	Good for low DoA; not reliable for high-speed V2X
R. Mardeni [2]	Optimization of path loss parameters	N/A	Errors $< 4.3\ dB$ (suburban) and $< 1.9\ dB$ (urban open areas)
Alsamhi et al. [18]	Predicting signal strength based on drone parameters like altitude, path loss, distance	ANN	$R^2 = 0.96$–0.98
Thrane et al. [1]	Integration of OSM with expert knowledge for path loss prediction	DNN, Model-aided deep learning	RMSE of 6 dB in signal strength prediction accuracy
Muhammad Brata [20]	Predict path loss at sub-6 GHz and mmWave using satellite images	CNN (VGG-16, ResNet-50), transfer learning	RMSE reduced by 5.5 dB vs. 3GPP; VGG-16 best
Thao T. Nguyen [5]	Path loss prediction at 7 GHz in urban settings	CNN + FCN with 3GPP priors	RMSE $\approx$ 4.5 dB; sub-5 dB after empirical tuning
Majd Alkorabi [21]	Reliable communication prediction in V2X for AVs	CNN-LSTM hybrid deep learning	Outperformed traditional models, achieving higher accuracy and more robust predictions of RSRP/SNR in dynamic urban scenarios
Jamie Mersh [23]	Reliable communication prediction in V2X for AVsAnalyze sensitivity of coverage prediction to input data resolution and environmental factors	N/A	Showed coverage accuracy highly dependent on data resolution; limited generalisation across diverse scenarios
Kehai Qiu [19]	Develop a deep learning model for predicting outdoor path loss maps using synthetic ray-tracing data	SegNet-based Convolutional Neural Network (PPNet)	Achieved RMSE of 0.0507 and MAE of 0.0290 on unseen urban test sets; inference 30–100× faster than commercial ray-tracers

Table 3.1: Summary of Related Works

Table 3.1 summarises the key studies in the area, highlighting their goals, the machine learning models employed, and the main results obtained by the authors.

Additionally, Table 3.2 outlines the primary limitations identified in each study. These constraints

Name/Reference	Limitations
Jakob Thrane [4]	Limited testing area, reliance on high-resolution satellite data
R. Mardeni [2]	Limited to WiMAX band (2360–2390 MHz), applicability to other terrains and environments not studied
Alsamhi et al. [18]	Computational complexity in real-time large-scale drone-IoT setups
Thrane et al. [1]	Limited by the complexity of local radio effects; performance varies with geographical region
Akhpashev and Andreev [3]	Limited generalisation across environments; correction factor derived from a single LTE urban dataset and may not adapt to varying urban morphologies or frequency bands
Minovski et al. [22]	Lack of data diversity in some test environments; reliance on specific hardware features and metrics
Cabric [17]	Lack of real-world testing; limited to simulated environments; needs further refinement in complex dynamic environments
Umar D. Maiwada* [24]	High computational complexity for real-time simulation; dependency on real-time telemetry; integration and security challenges in existing 5G infrastructure
Foliadis et al. [25]	Requires high-quality synchronized data; less effective with sparse CSI or sensor noise
Claussmann et al. [9]	Methods often lack scalability, real-time viability, or handling of uncertainties in highway environments
Badue, Guidolini, and Carneiro [8]	No empirical results; broad scope without implementation detail
Taheri et al. [26]	Needs dense measurements nearby
Vishaka Basnayake [27]	Non-Standalone-only, phone-based tests, region-limited
Muhammad Brata [20]	Relies solely on satellite images; lacks signal-level data and physical priors; limited scalability in resource-constrained setups
Thao T. Nguyen [5]	Requires high-res satellite imagery and accurate location data; limited scalability
Majd Alkorabi [21]	Urban dataset dependency; high computational cost; limited generalization to rural/highway scenarios; requires large datasets; low interpretability; potential sensitivity to mobility and interference.
Jamie Mersh [23]	Depends on data resolution; ignores dynamic factors (vehicles/pedestrians); limited scalability; only urban validation
Kehai Qiu [19]	Trained exclusively on synthetic ray-tracing data; model performance on real-world measured data not evaluated; generalization to diverse propagation conditions remains unverified

Table 3.2: Limitations identified in the Related Works

often stem from the specific characteristics of the deployment environment. Common issues include dependence on high-quality data, difficulty in adapting to diverse urban contexts, and limited transferability across settings. Such factors hinder the broader applicability and generalisation of the proposed techniques.

Several recurring patterns emerge from the analysis of Tables 3.1 and 3.2. Existing methodologies face persistent challenges in integrating heterogeneous inputs, such as three-dimensional topographical data or spatiotemporal features, which are essential for robust environmental modelling. Moreover, many approaches rely heavily on high-resolution satellite imagery or precise geolocation data, limiting their practicality in under-instrumented or resource-constrained regions.

In contrast, our proposed framework addresses these limitations by leveraging lightweight semantic data from OpenStreetMap, combined with structured signal metrics and deterministic models. This multimodal integration supports accurate and interpretable signal prediction without depending on dense or costly inputs, enabling deployment in a wide range of real-world environments.

Chapter 4

Coverage Prediction Approach

To incorporate spatial urban structure into the predictive modelling of signal strength, a semantic geospatial encoding strategy was developed. This approach leverages building footprint data extracted from OSM and processed via GeoJSON files to represent urban morphology as image inputs. The geographical images are subsequently converted into numerical tensors and input into the convolutional component of the hybrid model.

Two techniques were analysed to assess the effectiveness of spatial information encoding: a colour-coded semantic representation and a black and white encoding. Each method individually encodes urban structure and provides varying levels of abstraction for the CNN to process. The subsequent sections provide a detailed examination of various approaches.

4.1 Model Design

This study used a hybrid deep learning architecture that combines a CNN and a DNN to address the challenges of predicting RSRP in complex urban vehicular environments. This architecture is chosen due to the multimodal attributes of the input data and the need to articulate both geographical and non-spatial factors that affect wireless signal propagation.

While ViT (Vision Transformers) has recently demonstrated competitive performance in various computer vision tasks, its effectiveness often relies on large training datasets and substantial computational resources Dosovitskiy et al. [28]. In contrast, CNNs remain highly efficient for tasks with limited data availability, particularly when local spatial patterns are essential for the prediction problem. In this study, the spatial features extracted from geospatial semantic maps benefit from the strong inductive biases of CNNs, such as translation invariance and local connectivity, which enable them to learn fine-grained spatial structures more effectively and with reduced training time compared to ViTs. Prior comparative analyses, including Stéphane Cuenat [29] and José Maurício [30], indicate that although ViTs can attain comparable accuracy, CNNs frequently exhibit enhanced robustness under constrained conditions, rendering them an appropriate selection for our multimodal late-fusion architecture.

Semantic height maps of urban areas are represented as RGB images, encapsulating physical attributes such as building heights and obstructions. The analysis employs a convolutional neural network with multiple convolutional blocks, followed by batch normalisation and pooling layers. This approach enables the model to recognise spatial hierarchies and detect local patterns that influence signal attenuation, especially in densely populated metropolitan areas where obstructions and reflections pose significant challenges.

The choice of late fusion, rather than early fusion, was driven by the heterogeneity of the input modalities. Processing image-based spatial features and tabular signal/context features in separate specialised branches allows each network component to optimise for its respective data type before combining their representations. This strategy reduces the risk of modality interference during training and has been shown to improve performance in multimodal prediction tasks.

The DNN component processes structured tabular characteristics, including signal quality metrics such as RSRQ, SNR, CQI (Channel Quality Indicator) and RSSI (Received Signal Strength Indicator), together with throughput and the distance to the serving cell. Furthermore, we integrate a physically calculated path loss value based on the 3GPP 38.901 model, obtained from the frequency and distance characteristics. This model was selected because it is a widely adopted and validated standard for urban macro and microcell scenarios, offering realistic propagation estimations for autonomous vehicle communication systems.

The features are subjected to preprocessing via standard normalisation techniques to ensure numerical stability during training. The DNN has multiple fully linked layers with ReLU activations and batch normalisation, enabling it to capture intricate non-linear correlations in the data while enhancing generalisation.

The outputs from the CNN and DNN branches, in conjunction with the path loss scalar, are concatenated and input into a final regression module that forecasts the RSRP. This fusion technique enables the network to integrate geographical information, actual signal data, and theoretical propagation estimations, resulting in a more resilient and precise model.

The training employs a 10-fold cross-validation method to improve generalizability and reduce overfitting. We utilise early stopping predicated on validation performance, retaining the model that attains the optimal validation RMSE in each fold for final assessment. The employment of RMSE as the primary performance metric aligns with prior studies in signal forecasting. All forecasts are denormalised before assessment to produce interpretable results in dB units.

The implementation of a hybrid CNN and DNN model is a prudent decision that acknowledges the variability of input data and the intricate dynamics of signal transmission in urban traffic environments. This architecture precisely predicts signal strength in real-world communication networks for autonomous vehicles by incorporating spatial comprehension, learning signal patterns, and domain-specific propagation theory.

4.2 Dataset Description and Custom Preprocessing

This research employs the public dataset created by Raca et al. [13]. The dataset comprises high-frequency LTE cellular measurements collected in real-world mobility scenarios, encompassing many attributes relevant for training data-driven models in vehicular communication systems.

The dataset used in this work is the Radio Access Characterisation for Autonomous vehicles, which contains field measurements of LTE signal quality and contextual parameters collected across different mobility scenarios. It was chosen because it provides high spatial resolution measurements in various real-world environments, including static, pedestrian, bus, car, and train scenarios. This diversity enables the evaluation of communication performance under multiple movement conditions, making it highly suitable for modelling urban vehicular signal propagation.

The dataset includes measurements of RSRP, RSRQ, SNR, CQI, RSSI, throughput, serving cell identifiers, and GPS coordinates. Data was gathered via specialised mobile network measurement apparatus, incorporating GPS logging at each sampling location. The dataset encompasses numerous days and

times, with all measurements conducted under standard network operating settings, including authentic fluctuations in signal quality attributable to network congestion, environmental influences, and mobility.

Customised preparation was executed to ensure the dataset's suitability for deep learning. The fundamental steps were:

- Outlier detection to remove physically implausible values, such as RSRP readings outside the valid range (e.g., below –140 dBm or above –40 dBm) or GPS coordinates outside the study area.
- Removal of missing or corrupted records to prevent introducing bias during training.
- Standardised path loss value computation based on a physical model (3GPP 38.901).
- Geographical filtering to retain only measurements within the Cork case study boundaries.
- GPS-to-OSM tile matching at zoom level 17 to ensure precise spatial correspondence between measurements and OSM-derived semantic maps.
- Normalisation of all numerical features to stabilise model training.

Engineered features were also added to enhance model input:

- **Type -** a categorical variable denoting the mobility scenario (Static, Pedestrian, Bus, Car, or Train). This label was derived from the original dataset's metadata and enables stratified analysis of model performance under different mobility conditions.
- **ServingCell_Distance -** a numerical attribute representing the geographic distance (in meters) between the measurement location and its associated serving cell tower. This was computed using the Haversine formula, based on the latitude and longitude of both the measurement point and the serving cell coordinates.

In addition to the engineered features, the dataset provides eight core measurement attributes essential for wireless signal modelling:

- **Mobile Position:** Current GPS coordinates of the mobile device.
- **RSRQ:** Evaluates the received signal quality by calculating the ratio between RSRP and RSSI. Expressed in decibels (dB), it reflects the signal clarity and link reliability.
- **SNR :** Represents the ratio between signal power and background noise, measured in dB. It is crucial for assessing communication reliability.
- **CQI:** Derived from the SNR, it is reported by the user equipment to assist the network in selecting an appropriate modulation and coding scheme. Values range from 0 to 15.
- **RSSI:** Indicates the total received signal power, including noise and antenna losses, measured in dBm. It serves as a general indicator of signal strength.
- **Cell Position:** Provide the geographic coordinates (longitude and latitude) of the serving cell tower associated with each measurement.
- **Cell Distance:** Indicates the distance (in meters) from the mobile device to the serving cell tower. It functions as a crucial spatial metric.

- **Throughput:** Represents the effective data rate received by the user equipment, measured in Mbps. It reflects real-time network performance and quality of service.

The final processed dataset was stored in a consolidated CSV file, containing all preprocessed measurements alongside the engineered features. This refined dataset provided the foundation for the subsequent geospatial encoding and deep learning stages.

4.3 Geospatial Semantic Encoding and Integration

Geospatial data was encoded and integrated into the hybrid learning framework to enhance the contextual understanding of the urban landscape. This encoding encompasses environmental attributes that influence signal propagation, including the existence, elevation, and spatial arrangement of structures. The geospatial data, sourced from OSM and preprocessed GeoJSON files, was transformed into semantic images that depict the spatial configuration of urban structures. These images were subsequently processed by the CNN component of the hybrid model to extract relevant spatial features.

Two encoding strategies were implemented and analysed independently:

- **Colour-based semantic encoding**, in which each building height category was assigned a distinct RGB value.
- **Black and White semantic encoding**, in which buildings were represented as binary masks indicating only their presence or absence.

4.3.1 Color-Based Semantic Encoding

Colour-based semantic encoding was used to express structural variation in urban environments visually. This technique encodes building height information into RGB colour values, enabling the CNN to extract semantically meaningful spatial patterns from rendered urban maps.

The encoding process begins by extracting building geometries from GeoJSON files using the `osmnx` [1] and `geopandas`[2] libraries. Each record in the dataset included the user and serving cell coordinates. The midpoint between these two points was used as the centre of interest for each rendered map. A dynamic radius, proportional to the distance of the serving cell, was applied to define the spatial coverage of each image.

To enrich the building data with height information, average building heights were computed for each building type based on the available metadata. In cases where height values were missing or inconsistent, the mean height per building category was used as an alternative. These estimated heights were then grouped into predefined bins, each associated with a distinct colour, as illustrated in Table 4.1.

Each building in the retrieved region was coloured according to its corresponding height bin. In cases where no height information was available, a default colour (black) was applied. These colour-coded urban maps were then rendered using the `matplotlib` [3] library.

[1] `https://github.com/gboeing/osmnx`
[2] `https://geopandas.org/en/stable/docs.html`
[3] `https://matplotlib.org/`

Height Range (m)	Color
$H = 0$	Black
$0 < H < 1$	Blue
$1 \leq H < 5$	Green
$5 \leq H < 10$	Yellow
$10 \leq H < 20$	Orange
$20 \leq H < 30$	Red
$30 \leq H < 50$	Purple
$H \geq 50$	Pink

Table 4.1: Height-to-Colour Mapping Used for Semantic Encoding.

Figure 4.1: Example of semantic encoding with short user-to-cell distance.

Figure 4.1 illustrates an example where the user is located close to the serving cell tower. Due to the shorter distance, the rendered region is more zoomed in and contains detailed information about individual buildings.

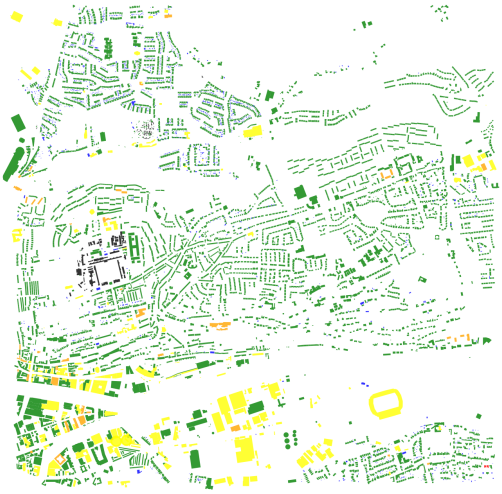

Figure 4.2: Example of RGB-encoded semantic map capturing height-based building structure.

Another example is provided in Figure 4.2, which shows the overall appearance of one of the semantically encoded maps used as model input. These RGB images were later decoded into single-channel height maps, enabling the CNN to work with structured, numerical input that captures both building presence and vertical variation.

This encoding method provides the model with critical structural context, allowing it to better understand signal obstructions, building density, and propagation paths across heterogeneous urban areas.

4.3.2 Black and white encoding

As a baseline comparison, a black and white semantic encoding strategy was also developed. Unlike the colour-based approach, this encoding does not differentiate building heights but instead focuses on the presence or absence of structures.

Each semantic image was generated using the same spatial logic: centring the map between the user location and the serving cell, with a radius scaled by the transmission distance. Building footprints were extracted via `osmnx` and rendered in black over a white background using the `matplotlib`library. No height differentiation was encoded.

These images, representing binary urban layouts and transformed into black and white tensors. Each pixel was mapped to either 0 (no building) or 1 (building), resulting in simplified spatial masks.

Figure 4.3: Example of a geographical image generated from OSM data, illustrating building footprints and spatial distribution.

Figure 4.3 illustrates an example of a black and white urban semantic image. Although this approach lacks vertical structure information, it provides proper control for evaluating the benefit of height-aware encoding.

By comparing model performance across the colour and black and white encodings, the influence of detailed spatial information on signal strength prediction can be empirically assessed.

4.4 Preprocessing pipeline

The preprocessing pipeline was designed to ensure that heterogeneous data sources, including tabular cellular measurements, geospatial semantic images, and theoretical path loss estimates, were aligned, normalised, and ready for ingestion into the hybrid neural architecture. This stage acted as the bridge

between raw data acquisition and model training, converting the collected information into a structured, machine-readable format suitable for each branch of the network.

The preprocessing pipeline begins with the selection of relevant features from the original dataset. Although the dataset contained a wide range of LTE channel and context metrics, only the attributes relevant to the proposed modelling approach were retained. Supplementary designed attributes were implemented to improve spatial and contextual representation. A novel categorical variable, `Type`, was created to categorise each sample according to the measurement scenario (*Static*, *Pedestrian*, *Car*, *Bus*, or *Train*). The variable `ServingCell_Distance` was calculated using the Haversine formula to ascertain the great-circle distance between the user equipment (UE) coordinates (`Latitude`, `Longitude`) and the serving cell coordinates (`ServingCell_Lat`, `ServingCell_Lon`).

For tabular features (e.g., RSRQ, SNR, CQI, RSSI, throughput, `ServingCell_Distance`), a z-score standardisation was applied using `StandardScaler` to ensure zero mean and unit variance across features, thus avoiding scale dominance during training. The target variable (RSRP) was normalised using the same method to stabilise the regression output.

In parallel, geospatial semantic images, both colour-encoded and black and white variants, were generated using the methodology described in Section 4.3.1. For the CNN branch, these images were resized to 150×150 pixels and transformed into numerical tensors. In the colour-based approach, each pixel's RGB value was mapped to a corresponding building height; in the black and white approach, building footprints were represented as binary intensity maps. Both encodings were converted into single-channel tensors and normalised to maintain consistent intensity ranges.

The theoretical path loss estimation, produced by the custom `PathLossModel` based on 3GPP TR 38.901 formulations, was precomputed for each sample using `ServingCell_Distance` and an estimated carrier frequency derived from the `NetworkMode` field. The resulting path loss values were standardised before being integrated as a separate numerical feature.

Finally, all data streams were aligned by truncating them to the minimum available sample count across modalities, ensuring that each index corresponded to a complete set of features:

1. A semantic height map image (CNN input);
2. A standardised tabular feature vector (DNN input);
3. A standardised path loss value (physical model input);
4. The corresponding normalised RSRP value (regression target).

This multi-modal dataset served as the direct input to the training process described in Section 5.2, enabling the hybrid architecture to jointly exploit spatial, contextual, and physical information for RSRP prediction.

4.5 Hybrid Model Architecture

The proposed framework addresses the challenge of accurately predicting RSRP in autonomous vehicle (AV) scenarios, where dynamic mobility and complex urban environments make signal estimation particularly difficult.

Figure 4.4 illustrates the proposed hybrid deep learning architecture.

Traditional models often struggle with the heterogeneity of urban landscapes and the scarcity or inaccessibility of high-quality input data. To overcome these limitations, our architecture fuses three complementary information sources:

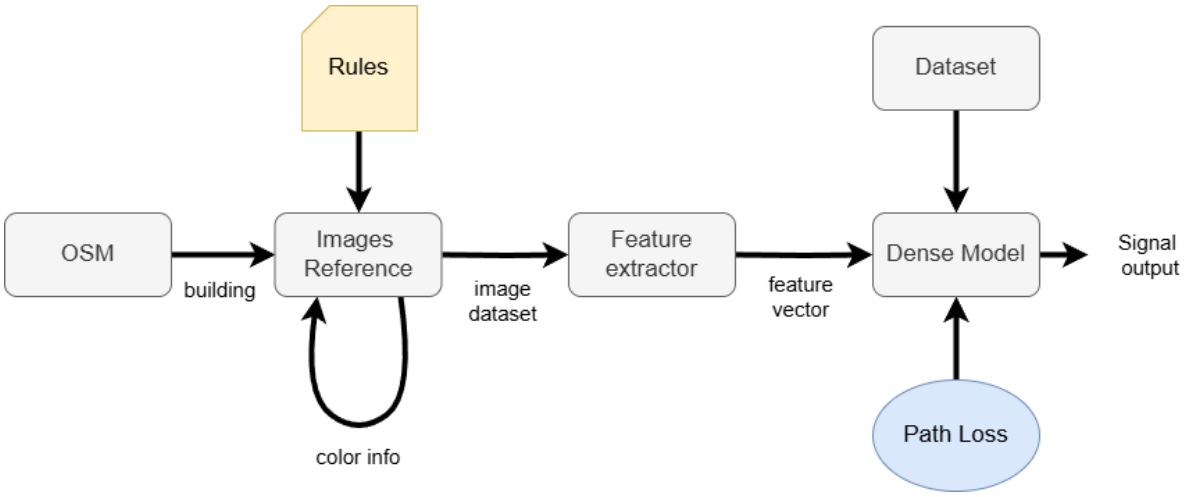

Figure 4.4: Hybrid model architecture integrating semantic OSM-based images (CNN), signal-level tabular metrics (DNN), and a deterministic path loss scalar. Outputs are fused for final RSRP prediction.

- **Semantic images from OSM:** capturing the structural layout of urban environments.
- **Signal-level tabular metrics:** encapsulating physical and contextual KPIs.
- **A deterministic path loss scalar:** providing inductive physical guidance via propagation models.

These inputs feed into three synchronised branches: a *CNN* for image-based spatial features, a *DNN* for tabular KPIs, and a *scalar input node* for the 3GPP-based path loss estimate. Their outputs are fused into a unified latent representation, enabling accurate and interpretable RSRP prediction.

To ensure the full reproducibility of the proposed architecture, detailed implementation specifications are presented. The CNN branch comprises three sequential convolutional blocks with filter dimensions of 32, 64, and 128, respectively. Each block consists of a two-dimensional convolutional layer (kernel size 3×3, stride 1, padding 1), followed by a LeakyReLU activation function (negative slope of 0.2), batch normalisation, and a 2×2 max-pooling operation. The input image is preprocessed into a single-channel tensor and normalised using a mean value of 35 and a standard deviation of 25. The resulting feature maps are subsequently flattened and passed through a fully connected layer comprising 100 neurons.

The DNN branch processes a six-dimensional vector comprising the tabular KPIs described in Section 4.2. These features are standardised using z-score normalisation via a `StandardScaler` to ensure consistent scaling. The vector is then passed through a multilayer perceptron with two hidden layers of 200 and 100 units, respectively, each followed by a ReLU activation and a batch normalisation layer.

The path loss scalar input is derived from a deterministic model based on the 3GPP TR 38.901 formulation, using fixed transmission power and subcarrier configurations. The resulting path loss values are independently standardised using a dedicated `StandardScaler` instance prior to integration into the hybrid model's fusion layer.

The outputs of the three branches namely, the CNN feature vector (100 dimensions), the DNN output (100 dimensions), and the scalar path loss estimate (1 dimension), are concatenated and passed through a fusion module. This module consists of a fully connected layer with 16 neurons followed by a ReLU activation, and a final output neuron that produces the predicted RSRP value.

The architecture consists of the following core components:

- **OSM:** Provides the primary geospatial information, including raw building footprint geometries, spatial layout, and structural location. This data is retrieved dynamically from OSM through spa-

tial queries centred between the UE and the serving cell tower, with a radius proportional to the transmission distance.

- **Rules:** This component delineates the criteria (a specific set of parameters) for creating the *Images Reference* inputs. A customised colour gradient is used to represent the elevation of each structure, categorised as described in Table 4.1.

- **Images Reference:** This module merges the OSM building data with the rule-based colour-mapping logic (see Section 4.3.1) to generate semantically rich 2D images for model input. Each image visually encodes the spatial layout and elevation profile of urban environments using a consistent visual scheme, designed to facilitate spatial feature extraction through the CNN. To retrieve building footprints, the `osmnx` library is employed to query OSM within a dynamic search radius, defined as $1.3\times$ the serving cell distance (capped at 1 km). Non-structural geometries (e.g., `slurry_tank`) are filtered out to retain only valid `Polygon` or `MultiPolygon` types. Building heights are then assigned hierarchically:

  (a) directly from the `height` tag, if available;

  (b) inferred from the `building:levels` attribute (assuming 3 m per floor);

  (c) estimated from a precomputed GeoJSON average by building type.

  An example is illustrated in Figure 4.2.

  The left block of Figure 4.4 illustrates the CNN branch responsible for processing the semantic images generated from `OSM` data. These images use *colour* gradients to encode building heights and layouts, providing spatial information crucial to understanding signal obstruction and reflection in dense urban areas.

- **Feature Extractor:** This component processes the semantically enriched images described in Section 4.3.1, extracting spatially aware representations using a sequence of convolutional and max-pooling layers. These layers progressively distil spatial features while preserving semantic consistency. The resulting representations are flattened and passed through fully connected layers, producing a compact feature vector that captures both local and global spatial patterns. This transformation enables the model to infer how urban morphology affects signal propagation, thereby supporting accurate RSRP prediction.

- **Dense Model:** Integrates numerical parameters such as latitude, longitude, antenna elevation, tilt, and azimuth with features from the feature extractor, path loss module, and additional data branch. It processes these diverse inputs through fully connected layers to evaluate signal strength, encompassing nonlinear linkages and spatial interactions.

- **Path Loss Input:** The scalar path loss value, computed using the 3GPP 38.901 model (Equation 4.1), is introduced as a standalone input. It provides a physics-based estimate of signal attenuation, contributing a strong inductive bias toward plausible signal behaviour.

$$\mathrm{PL} = 28 + 22\log_{10}(d) + 20\log_{10}(f) \tag{4.1}$$

where d is the transmitter–receiver distance in meters and f is the frequency in GHz (estimated from network mode). The RSRP is derived by subtracting antenna gain and resource block spreading loss.

- **Dataset:** This component leverages the preprocessed mobility dataset described in Section 4.2, which includes over 135 log files collected from real UE across various transportation modes. Each sample is associated with a 1 Hz timestamp and includes multiple key performance indicators (KPIs) and contextual attributes, such as geographic coordinates, signal quality metrics (RSRP, RSRQ, SNR, RSSI, CQI), mobility parameters (speed, direction), and distance to the serving cell.

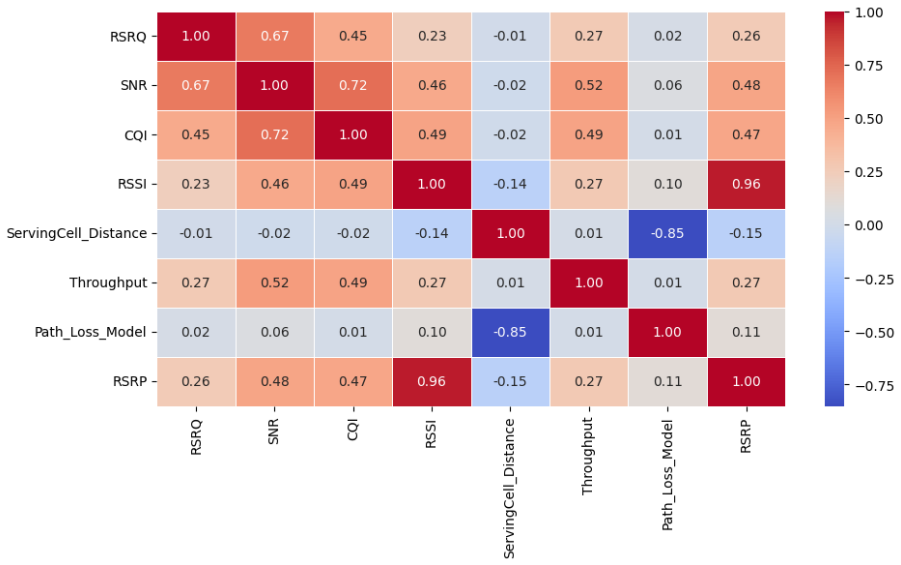

Figure 4.5: Feature correlation heatmap among selected variables including RSRP, RSSI, and path loss.

After preprocessing, a Pearson correlation analysis was applied to assess redundancy and complementarity among the input features. The study revealed a strong linear correlation between RSRP and RSSI ($\rho \approx 0.96$), as well as moderate associations between SNR, CQI, and RSSI. The path loss feature showed a substantial negative correlation with serving cell distance ($\rho = -0.85$), consistent with theoretical propagation principles. The correlation structure is summarised in Fig. 4.5, highlighting both redundancy risks and complementary feature relationships. These insights supported feature selection, ensuring that only informative and non-redundant attributes were retained for model training.

- **Fusion and Prediction Head (Signal Output):** As shown in the rightmost block, outputs from the CNN, DNN, and the scalar path loss are concatenated into a unified representation. This fused vector is then passed through a two-layer fully connected network to regress the final RSRP value. The design ensures that both data-driven patterns and domain-specific priors are leveraged in a synergistic manner.

Chapter 5

Model Workflow and its implementation

The implementation of the proposed RSRP prediction system follows a methodical workflow that encompasses data engineering, model architecture development, and iterative refinement. This phase aims to transform unrefined multimodal datasets into a trained hybrid deep learning model capable of accurately predicting RSRP values in urban environments. The overall process is summarised in Figure 5.1 and is described in detail below, following its sequential logic.

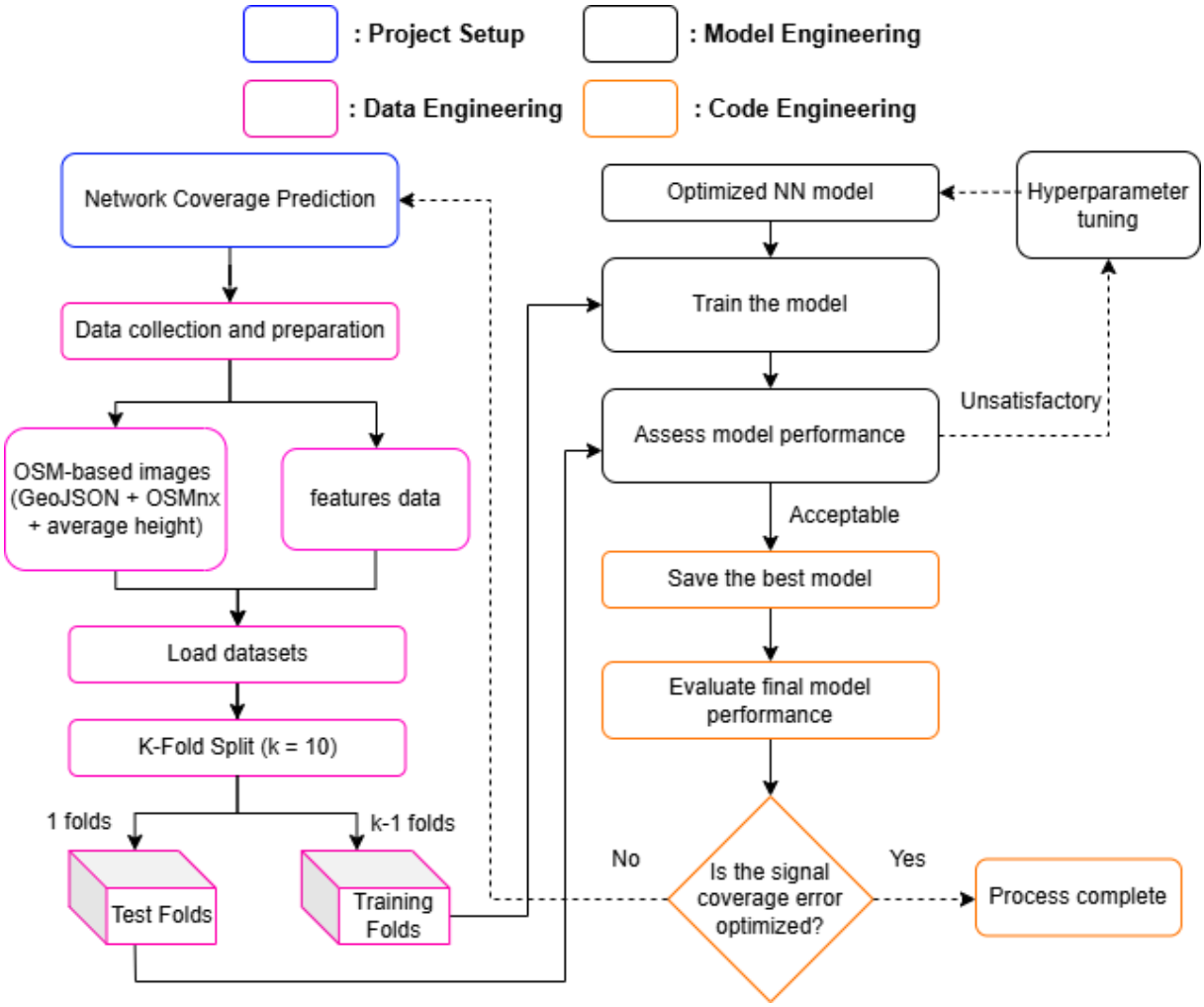

Figure 5.1: Generalized deep learning workflow illustrating the complete pipeline for RSRP prediction. The process comprises project setup, data engineering (including OSM-based spatial image processing, tabular features, and 3GPP-based path loss), model training with cross-validation, performance evaluation, and hyperparameter tuning

5. MODEL WORKFLOW AND ITS IMPLEMENTATION

The starting point was the definition of the modelling objective: to estimate RSRP values using a combination of geospatial, tabular, and physical model features, with the RMSE chosen as the primary evaluation metric. This metric reflects the magnitude of prediction errors in the natural dB (decibels) scale, which is directly interpretable in the context of cellular network performance. From the outset, the modelling approach was defined as a hybrid architecture capable of handling both image-based and tabular data in a unified framework.

The data preparation phase integrated three complementary sources:

i) cellular measurement data containing LTE/4G parameters such as RSRQ, SNR, CQI, RSSI, throughput, serving cell identifiers and coordinates, and network mode;

ii) geospatial semantic data from OSM, processed to extract building footprints, road networks, and building height information;

iii) theoretical path loss values calculated using the 3GPP TR 38.901 propagation model, based on transmitter–receiver distances and carrier frequency.

Geospatial features were converted into semantic raster images, where each building height range was encoded with a specific colour, ensuring that the convolutional network could effectively process spatial context. The images were generated with a fixed zoom level to ensure consistent spatial coverage. Later, during the CNN input preparation, their colour values were mapped back to height information.

The tabular features underwent standardisation via z-score normalisation to ensure that all variables contributed equally to the training process. The target variable (RSRP) was also standardised for training stability. In addition to the original measurements, two engineered features were included:

- **Type:** a categorical label indicating the mobility scenario (Static, Pedestrian, Bus, Car, or Train).
- **ServingCell_Distance:** the calculated distance between the serving cell and the measurement point, obtained from latitude and longitude coordinates using the Haversine formula.

The Frequency feature was obtained from the network mode through preset mappings, ensuring the physical propagation characteristics were accurately represented in the dataset.

Upon preparing all modalities, alignment was executed to ensure that the semantic pictures, tabular features, path loss values, and RSRP targets belonged to the same set of samples. The dataset was later divided utilising 10-fold cross-validation, yielding reliable performance assessments and reduced overfitting.

The model architecture comprised three parallel branches:

- A CNN branch processing the semantic height map images, extracting spatial features through successive convolutional and pooling layers.
- A DNN branch processing the tabular LTE features, using fully connected layers with batch normalisation to stabilise learning.
- A physical model branch, which ingested the standardised path loss values as a single numerical feature.

The results from these three branches were combined into a single feature vector, which was then processed by fully connected regression layers to obtain the final RSRP estimate. Hyperparameters were

established to enhance the balance between predictive accuracy and computational efficiency, comprising a batch size of 32, a learning rate of 1×10^{-3}, Leaky ReLU activation in the CNN, ReLU activation in the DNN and final layers, and an early stopping patience of 10 epochs.

The training employed the Adam optimiser and the RMSE loss function, with a maximum of 100 epochs. Data loading pipelines facilitated effective parallel processing of picture and tabular data. After each epoch, the model was evaluated on the validation set using denormalised predictions, and the best-performing model state (i.e., the one with the lowest validation RMSE) was saved. Early stopping was applied to prevent overfitting.

Upon completion of training across all folds, the RMSE scores from each fold were aggregated to compute the mean and standard deviation, providing a robust indicator of generalisation performance. If the target performance threshold was not reached, hyperparameter tuning was performed, adjusting learning rates, network depths, or preprocessing parameters until satisfactory results were achieved.

This systematic execution guaranteed that all phases of data preparation, model construction, and evaluation were meticulously integrated, facilitating the creation of an optimised hybrid model for accurate RSRP prediction. The subsequent subsections provide an in-depth examination of the model architecture, hyperparameters, and training methodologies.

5.1 Model Configuration and Hyperparameter Setup

The proposed hybrid deep learning model was configured through an iterative process of experimental evaluation, guided by both empirical performance and theoretical considerations in deep learning and wireless communication modelling.

Tables 5.2 and 5.1 summarise the final set of hyperparameters and architectural design parameters used in the proposed model, which were selected to optimise prediction accuracy while maintaining computational feasibility.

Hyperparameter	**Value / Description**
Image input size	150×150 pixels
CNN output dimension	100 (flattened vector)
DNN tabular layers	[200, 100] neurons
CNN convolutional channels	[32, 64, 128]
Activation function	Leaky ReLU (CNN), ReLU (DNN and head)
Batch normalization	After each CNN/DNN layer
Fusion vector	Concatenation of CNN, DNN, and path loss
Final regression layers	[16, 1]

Table 5.1: Model architecture and design configuration.

A batch size of 32 was adopted as a compromise between computational efficiency and gradient stability. Smaller batch sizes tended to produce noisier gradient estimates, whereas larger batches increased memory usage without significant improvements in validation performance. The learning rate was fixed at 1×10^{-3}, which proved optimal in initial grid search experiments: lower values slowed convergence excessively, while higher values risked overshooting minima and destabilising training.

Hyperparameter	Value / Description
Batch size	32
Learning rate	1×10^{-3}
Optimizer	Adam
Loss function	RMSE
Number of epochs (max)	100
Early stopping patience	10 epochs
3GPP Path Loss offset	0 dB (default)

Table 5.2: The proposed model's Hyperparameter Configurations

The `Adam optimiser` [1] was chosen for its adaptive learning rate capabilities and ability to handle sparse and noisy gradients, which is advantageous given the heterogeneous nature of the input features (spatial images, tabular metrics, and physical model predictions). Adam updates the model parameters according to:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t, \quad v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1-\beta_2^t}$$

$$\theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}$$

where g_t is the gradient at time step t, η is the learning rate, and β_1, β_2 are exponential decay rates for moment estimates.

The loss function was defined as `RMSE` [2] to align with the evaluation metric directly, ensuring that model optimisation focused on reducing the primary performance measure:

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

where y_i is the true value, $\hat{y}_i$ is the predicted value, and n is the number of samples.

Training was limited to 100 epochs, with an early stopping patience of 10 epochs set to mitigate overfitting by halting training when no improvement in validation loss was observed.

The CNN component, responsible for extracting spatial features from OSM-derived images, consisted of three convolutional layers with channel widths of [32, 64, 128]. These employed the `Leaky ReLU activation` [3]:

$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \alpha x, & \text{if } x < 0 \end{cases} \quad \text{with } \alpha = 0.2$$

to retain small negative outputs, facilitating gradient flow and avoiding dead neurons. Batch normalisation was applied after each convolution to stabilise learning and accelerate convergence. The CNN output was flattened into a 100-dimensional vector, representing a compressed spatial embedding.

The DNN branch processed tabular features via two fully connected layers of 200 and 100 neurons,

[1] `https://docs.pytorch.org/docs/stable/generated/torch.optim.Adam.html`
[2] `https://docs.pytorch.org/ignite/generated/ignite.metrics.RootMeanSquaredError.html`
[3] `https://docs.pytorch.org/docs/stable/generated/torch.nn.LeakyReLU.html`

each followed by `ReLU activation` [4]:

$$f(x) = \max(0, x)$$

and batch normalisation. This architecture was selected after ablation studies showed that deeper or wider networks increased the risk of overfitting without significant gains in validation accuracy.

A third input branch incorporated the physical path loss estimate from the 3GPP TR 38.901 urban macrocell model. This scalar feature, computed from transmitter–receiver distance and carrier frequency, was normalised and concatenated with the CNN and DNN outputs to form the fusion vector. The fusion vector was passed through two fully connected regression layers [16, 1] to output the predicted RSRP in dB. The 3GPP path loss offset was fixed at 0 dB, consistent with standardised modelling assumptions.

This configuration was the result of iterative experimentation, achieving a balance between predictive accuracy, generalisation capability, and computational efficiency.

5.2 Training and Validation Strategy

To ensure robustness, generalisation capability, and unbiased model evaluation, this work employed a ***k*-fold cross-validation** strategy. Cross-validation is a resampling procedure in which the dataset is divided into k equally sized, non-overlapping subsets called *folds*. At each iteration, the model is trained on $k-1$ folds and evaluated on the remaining fold. This process is repeated k times so that every sample appears precisely once in the evaluation fold. The final performance is computed as the average across all folds, which mitigates the variance and bias introduced by any single train–validation split, thus providing a more reliable estimate of the model's out-of-sample error.

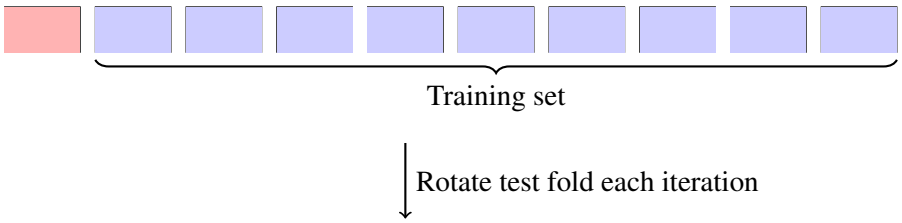

Figure 5.2: Illustration of the k-fold cross-validation process with $k = 10$, where one fold is used for testing and the remaining folds for training. This process is repeated until all folds have served as the test set.

In this work, $k = 10$ was selected as a trade-off between computational efficiency and statistical reliability, a choice widely adopted in deep learning studies with moderate dataset sizes. Figure 5.2 illustrates this procedure as implemented in our experiments:

- **Red fold** $\rightarrow$ the external *test set* for that iteration, which rotates across the dataset until all folds have served as the test set.
- **Blue folds** $\rightarrow$ the *training set*, used exclusively for model training.
- The diagram explicitly clarifies that no separate validation fold is carved from the training data. Instead, in each iteration, the test fold from the k-fold split is used both for validation and evaluation purposes, meaning the validation subset is not fixed but rotates dynamically with the test fold.

[4] `https://docs.pytorch.org/docs/stable/generated/torch.nn.ReLU.html`

5. MODEL WORKFLOW AND ITS IMPLEMENTATION

The data were randomly shuffled before splitting, using a fixed random seed to ensure reproducibility. All partitions maintained **perfect alignment across the three input modalities**, OSM-derived heightmap images, tabular signal quality metrics, and physical path loss estimates, ensuring that each multimodal triplet corresponded to the same geographical and temporal measurement instance in all folds.

The primary evaluation metric was the **RMSE**, selected for its direct interpretability in the context of RSRP prediction, as it expresses the average prediction error magnitude in dB, the same units as the target. The RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{5.1}$$

where $\hat{y}_i$ is the predicted RSRP for the i-th sample, y_i is the ground truth RSRP, and n is the total number of evaluation samples.

For each fold, the RMSE obtained on the external test set was recorded. The *mean* and *standard deviation* of RMSE across all folds were then computed, providing both a central tendency measure of predictive performance and an estimate of variability across different data partitions.

Chapter 6

Evaluation and Results

This chapter presents the experimental evaluation of the proposed hybrid deep learning architecture for RSRP prediction in urban vehicular scenarios. The assessment is conducted using real LTE drive-test datasets to analyse the model's performance under various urban morphologies and signal propagation conditions. To ensure a reliable and reproducible evaluation, we detail the computational infrastructure, software environment, and geospatial tools adopted throughout the experiments. The model's predictive capacity is analysed using standard regression metrics, such as RMSE and MAE (Mean Absolute Error), and benchmarked against conventional path-loss models and recent data-driven baselines. Additionally, we explore the model's robustness through spatial and distance-based performance analyses, highlighting its capacity to generalise across heterogeneous scenarios. The experiments follow a 10-fold cross-validation scheme, guaranteeing statistical rigour and consistent modality alignment throughout the evaluation process.

6.1 Experimental setup

The execution of the suggested hybrid deep learning framework was facilitated by a selection of hardware and software tools designed to guarantee computational efficiency, reproducibility, and scalability of experiments. The development environment incorporated GPU (Graphics Processing Unit) acceleration, geospatial data management, and sophisticated deep learning frameworks.

6.1.1 Hardware Infrastructure

All experiments were conducted on a system equipped with an **NVIDIA L40S GPU**[1], featuring **48 GB of GDDR6 memory**, and running **CUDA version 11.8**. The computing environment was based on **Ubuntu 24.04 LTS**, with an **AMD Ryzen 9 5900X** 12-core processor and **64 GB of DDR4 RAM**, facilitating rapid loading of extensive datasets and concurrent data preparation. Storage relied on a **2 TB NVMe SSD** to minimise I/O bottlenecks during image loading, dataset shuffling, and model checkpointing.

[1] `https://resources.nvidia.com/en-us-briefcase-for-datasheets/nvidia-rtx-a5000-dat-1`

6.1.2 Software Environment

The codebase was implemented in **Python 3.9.10**[2], with deep learning operations handled by **PyTorch 2.2.2**, leveraging **CUDA 11.8**[3] for GPU execution. **Visual Studio Code** (version 1.79.2) served as the integrated development environment (IDE). Key auxiliary packages included:

- **torchvision**[4] for data transformations and image preprocessing routines.
- **scikit-learn**[5] for standardisation of numerical features, K-Fold cross-validation, and performance metrics such as RMSE.
- **Pandas**[6] and **NumPy**[7] for dataset manipulation and efficient numerical computation.
- **Pillow** [8] for image file handling .
- **Matplotlib** [9] for visualisation of training curves and evaluation results.

6.1.3 Geospatial Data Processing Tools

Urban geospatial layers were obtained from OpenStreetMap via **OSMnx**[10] and **Overpass Turbo**[11]. The **Pillow** library was used to generate and process two types of spatial representations:

- Colour-encoded heightmaps, where discrete RGB values correspond to predefined altitude ranges.
- Black and White elevation maps providing continuous elevation encoding.

All images were resized to 150×150 pixels for consistency in CNN input.

6.1.4 Experiment Tracking and Version Control

All experiments, including model training logs, intermediate validation results, and per-fold metrics, were systematically recorded. Model checkpoints were saved after each fold in the cross-validation process, ensuring reproducibility and allowing post-hoc analysis. Version control was maintained using **Git**, enabling incremental development, safe experimentation with alternative architectures, and rollback capability in case of regressions.

6.2 Evaluation Metrics

- **Mean Absolute Error (MAE)**
  MAE quantifies the average magnitude of the errors in a set of predictions, without considering their direction. It is defined as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{6.1}$$

[2] https://docs.python.org/3/
[3] https://docs.nvidia.com/cuda/archive/11.8.0/
[4] https://docs.pytorch.org/vision/stable/index.html
[5] https://scikit-learn.org/stable/
[6] https://pandas.pydata.org/docs/
[7] https://numpy.org/doc/stable/
[8] https://pillow.readthedocs.io/en/stable/
[9] https://matplotlib.org/stable/index.html
[10] https://github.com/gboeing/osmnx
[11] https://overpass-turbo.eu/index.html

where y_i is the ground truth and $\hat{y}_i$ is the predicted value. MAE provides an interpretable measure of average error in the same units as the output variable.

- **Root Mean Squared Error (RMSE)**
  RMSE emphasises larger errors by squaring the differences before averaging, which is particularly useful when larger errors are more significant. It is calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6.2}$$

  Lower RMSE values indicate better performance, with a value of zero representing a perfect prediction.

- **Coefficient of Determination (R^2)**
  R^2, also known as the coefficient of determination, measures how well the predicted values approximate the actual data. It is given by:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6.3}$$

  where $\bar{y}$ is the mean of the ground truth values. An R^2 value of 1 indicates perfect prediction, while values below 0 suggest that the model performs worse than simply predicting the mean.

- **Physics-Based Baseline (3GPP TR 38.901 Path Loss)**
  Although not an error metric itself, a deterministic path loss model, as defined by the 3GPP TR 38.901 specification, is used as a baseline for comparison. This model provides a theoretical estimation of signal attenuation based on transmitter–receiver distance and frequency. It is defined as:

$$PL(d, f) = 28 + 22\log_{10}(d) + 20\log_{10}(f) \tag{6.4}$$

  where d is the distance between transmitter and receiver (in meters), and f is the carrier frequency (in GHz). Assuming a fixed transmission power $P_{\text{tx}} = 43$ dBm, the predicted RSRP is computed as:

$$P_{\text{RSRP}} = P_{\text{tx}} - PL(d, f) - 10\log_{10}(12N) \tag{6.5}$$

  where $N = 100$ is the number of LTE resource blocks. This expression provides a physics-informed estimate of signal strength and is used in our experiments as a reference point to evaluate the added value of the deep learning models. It allows quantifying the gain achieved by learning-based methods over traditional propagation models.

  In our implementation, a small scalar *offset* term is also included for empirical adjustment, allowing fine-tuning of the predicted values in case of systematic bias in the propagation model. This offset does not alter the theoretical structure of the 3GPP TR 38.901 formulation, but it enables better alignment with real-world measurements when necessary.

6.3 Comparative Analysis with State-of-the-Art

To assess the performance of the proposed multimodal framework against state-of-the-art methods, we conducted a comparative evaluation with the model introduced by Thrane et al. [1], one of the first works to leverage OSM-based semantic imagery and deep learning for RSRP prediction. Their original study reported an average RMSE of approximately 6 dB across heterogeneous measurement campaigns, with improved generalisation relative to classical empirical and ray-tracing models. However, despite these gains, their methodology was constrained by the reliance on visual inputs alone, without the integration of physical priors or signal-level features.

To enable a fairer evaluation, we retrieved the authors' publicly available implementation from GitHub[12], fine-tuned it using the vehicular dataset collected by Raca et al. [13], and re-evaluated its performance. This reimplementation achieved an RMSE of 3.0.4 dB with a standard deviation of 0.21 dB, significantly above the initial published findings. This enhancement underscores the model's sensitivity to dataset attributes and hyperparameter optimisation.

Model	Dataset	RMSE (dB)	Stdev (dB)
Proposed approach	Raca et al. [13]	2.77	0.07
Thrane et al. [1]	Satellite dataset	4.3	-
Thrane et al. [1]	GER Campus	6.3	3.6
Thrane et al. [1] (optimized for our dataset)	Raca et al. [13]	3.04	0.21

Table 6.1: Performance comparison between proposed and Thrane et al. [1] approaches.

Table 6.1 delineates the prediction errors of our model in comparison to those derived from the Thrane baseline. Although the latter exhibits competitive performance, our multimodal approach consistently surpasses it across folds, resulting in lower RMSE values and less variance. This confirms the advantages of jointly incorporating semantic imagery, signal-domain indicators (RSRQ, SNR), and deterministic path loss priors into a unified architecture. By embedding physical knowledge alongside spatial representations, our method achieves stronger generalisation in urban vehicular environments, narrowing the prediction error to well below 3 dB in most cases.

These findings emphasise that although Thrane's work established the value of OSM-based semantics for signal modelling, the integration of heterogeneous modalities, as realised in our framework, provides a step forward in prediction accuracy and robustness.

6.4 Model Configurations Analysis

To elucidate the function of each element within the proposed hybrid architecture, we assessed various model configurations utilising 10-fold cross-validation, obtaining the mean RMSE and standard deviation across folds (Table 6.2). The results confirm the incremental benefits of progressively integrating multimodal data into the predictive model.

The baseline configuration, reliant only on deterministic path loss calculations from the 3GPP 38.901 model, demonstrated the least effective performance, with a root mean square error of 14.81 dB. This error highlights the fundamental limitations of empirical models in complex urban settings, where the

[12] `https://github.com/jakthra/PseudoRayTracingOSM`

Model	RMSE (dB)	Stdev (dB)
Path Loss only	14.81	0.17
Tabular data (dataset + DNN)	3.81	0.10
Path Loss + Tabular data + context-aware Information (Black and white Images)	2.85	0.11
Path Loss + Tabular data + context-aware Information (RGB Images)	2.77	0.07

Table 6.2: Performance comparison of different model configurations.

unpredictability arising from multipath, diffraction, and shadowing cannot be adequately captured by distance-based equations alone.

The path loss baseline was substituted with a deep neural network exclusively trained on tabular data, resulting in an error reduction to 3.81 dB. This result highlights the anticipated effectiveness of measured radio parameters, such as RSRP, RSRQ, and SNR, confirming their importance for accurate coverage evaluation.

The integration of context-aware information through the CNN branch improved the model's precision. The utilisation of black and white images derived from OpenStreetMap data decreased the RMSE to 2.85 dB, demonstrating that even simplified spatial representations of the urban environment provide substantial insights into the effects of obstacles and building layouts on signal propagation. The optimal performance was achieved using the comprehensive hybrid model, which integrates path loss, tabular data, and RGB semantic images, yielding an RMSE of 2.77 dB and a minimal standard deviation of 0.07 dB. This design yielded the most accurate predictions and exhibited the highest stability over folds, signifying its resilience and generalisation abilities.

These findings validate the benefits of multimodal fusion. Path loss estimations, when integrated with data-driven attributes and semantic representations of the urban landscape, enable the model to attain both physical coherence and contextual variety. The improvement over baseline methods is substantial, confirming the effectiveness of the proposed hybrid framework for urban coverage prediction in autonomous vehicle networks.

6.5 Cross-Validation Results

To assess the performance of the proposed framework, we conducted a 10-fold cross-validation. Table 6.3 presents the results obtained regarding the RMSE, MAE, and the coefficient of determination (R^2). Metrics are expressed in dB.

Metric	Average Value (dB)	Std Dev (dB)
RMSE	2.8658	$\pm$0.1475
MAE	1.9750	$\pm$0.0459
R^2	0.9277	$\pm$0.0069

Table 6.3: Average prediction results for RSRP obtained from 10-fold cross-validation using the proposed hybrid model with image data, tabular features, and estimated path loss. Metrics include RMSE and MAE (in dB) and the coefficient of determination (R^2).

The findings exhibit robust predictive efficacy. Across all the folds, the RMSE remains consistently

below 2.8658 *dB*, while the MAE is 1.9750 *dB*. Furthermore, the R^2 values of approximately 0.93 – – 0.94 *dB* indicate that the model explains over 93% of the variance in the measured RSRP. These values compare favourably with state-of-the-art approaches reported in the literature, many of which achieve RMSE values above 3 *dB* even when employing high-resolution satellite imagery. The results, therefore, highlight the advantage of integrating semantic OSM-derived representations with structured radio indicators and deterministic path loss models, yielding both accuracy and scalability.

The narrow variance range indicates model stability and resilience to data distribution shifts across folds, confirming the reliability of the predictions beyond isolated test cases. These results validate the hybrid framework as a practical and scalable solution for RSRP prediction, combining semantic richness, measurement robustness, and physically-informed reasoning.

6.6 Hybrid Physics-Informed Deep Learning for Signal Prediction

The goal of this experiment was to evaluate the effectiveness of a hybrid deep learning architecture in predicting **RSRP** values, while benchmarking its performance against a purely theoretical path-loss model derived from the **3GPP TR 38.901** standard.

The hybrid model integrates three distinct complementary sources of information:

- **Environmental context** derived from semantic height maps processed by a CNN, capturing urban structures, obstacles, and terrain variations;
- **Measurement-based features** such as RSRQ, SNR, CQI, RSSI, throughput, and the distance to the serving cell, processed by a DNN;
- **Physics-based prior knowledge**, specifically path-loss estimates computed from the 3GPP 38.901 model, which serve as an auxiliary input during learning.

The rationale behind this hybrid design is to leverage the generalisation strength of theoretical models and the local adaptability of data-driven learning. Although the path loss model captures global propagation trends, such as distance- and frequency-based attenuation, it cannot adapt to local irregularities caused by buildings, foliage, or interference. In contrast, neural networks are well-suited to learning such residual effects from data but may lack physical grounding. By combining both approaches, the hybrid model is designed to learn corrections to the theoretical model using spatial and measurement-based features.

Model	Mean RMSE (dB)	Std Dev (dB)
Hybrid Model	2.77	0.05
Physical Model	14.13	0.23

Table 6.4: Average prediction error (RMSE) and standard deviation per fold for the proposed hybrid model and the traditional physical model.

To evaluate the model performance, a **10-fold cross-validation** was employed. The training process incorporated an early stopping strategy based on validation *RMSE*, with a patience of 10 epochs. In other words, training was stopped if no improvement in RMSE was observed for 10 consecutive epochs. The final model in each fold was selected based on the lowest validation RMSE achieved. The results across folds are summarised in Table 6.4. The hybrid model consistently achieved low RMSE values,

averaging 2.77 *dB* with a standard deviation of 0.05 *dB*. At the same time, the physical-only baseline yielded significantly worse performance with an average RMSE of 14.13 *dB* and standard deviation of 0.23 *dB*.

This substantial reduction of more than 10 *dB* in RMSE across folds highlights the hybrid model's ability to capture fine-grained propagation effects that are absent in the theoretical model alone. Specifically, these results support the following conclusions:

1. Physics-informed priors enhance the model's ability to generalise, acting as a grounding mechanism in areas with sparse or noisy data;
2. The CNN-based spatial encoder effectively captures urban topology and elevation, which correlates with signal degradation in cluttered environments;
3. Tabular features provide strong predictive signals about instantaneous link quality, complementing the spatial and physical components.

This experiment confirms that the proposed hybrid framework provides a robust and accurate solution for RSRP prediction in complex urban environments. It successfully bridges the gap between purely theoretical and purely data-driven models, offering a more generalisable and interpretable approach to wireless signal prediction.

It is also worth noting that the high RMSE observed in the physical-only baseline (14.13 *dB*) results from its inability to capture local irregularities such as buildings, foliage, or interference, since it only models the global attenuation trend with distance and frequency. The hybrid model, by contrast, achieved a substantially lower RMSE (2.77 *dB*) by learning residual effects from spatial and tabular features. Moreover, the inclusion of the theoretical path-loss model is not redundant; it acts as a physics-informed prior, guiding the learning process and improving generalisation in scenarios with sparse or noisy data. Unlike purely data-driven methods, the proposed approach also preserves interpretability, as the learned corrections can be explicitly compared with theoretical predictions, thereby offering transparency and physical grounding.

6.7 Performance Analysis by Distance Quartiles

To further investigate the robustness and generalisability of the proposed hybrid deep learning model, we analysed its prediction accuracy across different distance intervals between the user and the serving base station. The motivation for this test stems from a known challenge in wireless propagation modelling: traditional empirical models often degrade in accuracy as the distance between transmitter and receiver increases due to increased path loss, shadowing, and multipath effects. Therefore, evaluating the model's stability across distance ranges is essential to ensure its reliability in real-world, dynamic environments.

In this evaluation, the dataset was partitioned into four quartiles based on the user's distance from the serving cell, and the RMSE of the predicted RSRP values was computed separately for each interval. The use of quartiles was chosen to ensure a balanced number of samples per group, thereby avoiding intervals with very few data points that could lead to unstable error estimates. It is also important to emphasise that the reported values are not the received signal powers themselves, but rather the average prediction error (RMSE, in dB) within each distance interval. This design allows us to investigate whether the model's predictive accuracy is disproportionately affected by user proximity to the cell tower while maintaining statistical robustness across all distance ranges.

Distance Interval	Distance (m)	RMSE (dB)
[26.07, 400]	213.04	2.76
[400, 740]	570.00	2.88
[740, 1200]	970.00	2.72
[>1200]	6030.93	2.90

Table 6.5: Average RMSE values for RSRP prediction grouped by distance intervals. The reported distance corresponds to the midpoint of each interval and is included for reference only. The RMSE values are computed across all samples within each interval.

As shown in Table 6.5, the hybrid model exhibited consistent prediction errors across all quartiles. The RMSE values ranged from 2.77 dB to 2.90 dB, with the lowest error observed in the $\mathbf{[740-1200]\ m}$ range and the highest in the $\mathbf{[>1200]\ m}$ interval. It should be noted that the reported 6030 **m** value corresponds to the average user distance within this bin and does not represent a strict cutoff. Despite covering a wide range of distances from urban cells to far-field rural conditions, the variation in prediction accuracy remained below 0.3 dB, which is well within acceptable limits for RSRP estimation.

It is worth noting that RMSE values in the range of $2-3$ dB are generally considered acceptable for RSRP estimation, as such variation lies within the natural fluctuations introduced by fading and multipath propagation.

These findings confirm that the hybrid model generalises well across a variety of propagation scenarios. Unlike traditional models, which often struggle with increasing distance due to missing local features or signal obstructions, the proposed approach benefits from a combination of data-driven learning and physics-based priors. The semantic image encoder captures spatial structure and environmental features such as building height and density, while the tabular inputs provide real-time measurements that reflect link quality. Together, these allow the model to adjust effectively to both near-field and far-field conditions.

Notably, the consistently low RMSE values across quartiles demonstrate the model's resilience, even in long-range prediction scenarios. The minimal variation suggests that distance has only a marginal effect on prediction accuracy. This behaviour is particularly advantageous for applications such as real-time vehicular communication and adaptive handover systems, where signal quality must be estimated reliably across highly dynamic environments.

6.8 Spatial Performance Analysis

To assess the spatial generalisation capabilities of the proposed hybrid model, we conducted a detailed performance evaluation across different geographical regions. This analysis is essential for determining whether the model can maintain consistent predictive accuracy in locations with varying propagation conditions, a necessary aspect for real-world deployment, particularly in vehicular networks.

The evaluation involved partitioning the dataset spatially and measuring the RMSE within each region using a 10-fold cross-validation strategy. Two types of partitioning strategies were employed:

- **1D partitioning**: based solely on **longitude**.
- **2D partitioning**: based on both **latitude and longitude**.

After training the hybrid model architecture described in Section 5.2, we computed the RMSE values for each region in each fold. The average RMSEs per region are reported in Table 6.6 for the 1D

longitude-based partitioning, and in Table 6.7 for the 2D (latitude + longitude) partitioning.

Longitude Interval	Midpoint (°)	RMSE (dB)	Std Dev (dB)
$[-9.5, -8.4]$	-8.95	2.73	0.08
$[-8.4, -7.4]$	-7.90	2.91	0.20
$[-7.4, -6.3]$	-6.85	3.10	0.30

Table 6.6: Expected RMSE values per geographic region grouped by longitude.

Latitude Bin	Longitude Bin	RMSE (dB)	Std Dev (dB)
[51.8, 52.3]	$[-9.5, -8.8]$	2.68	0.17
[52.3, 52.7]	$[-8.8, -8.1]$	2.73	0.14
[52.7, 53.3]	$[-8.1, -7.4]$	2.88	0.16

Table 6.7: Expected RMSE values per spatial region grouped by latitude and longitude.

The spatial partitioning was implemented by discretising the latitude and longitude coordinates into three equally sized bins using the **pd.cut**[13] function. This ensured a balanced distribution of samples across regions, avoiding imbalanced subsets. After each training fold, the validation set was grouped according to these bins, and the RMSE was computed separately for each group by comparing the predicted and measured RSRP values. Finally, results were aggregated across folds to obtain the regional averages and standard deviations. The choice of three bins represented a trade-off between spatial granularity and sufficient sample size within each region, avoiding instability in the error estimation.

The results in both tables demonstrate that the model generalises well across different regions:

- In the **1D partitioning**, RMSE values range from 2.73 dB to 3.10 dB with low standard deviation ($\leq 0.19 dB$), indicating minimal sensitivity to longitude variation.
- In the more fine-grained **2D spatial division**, RMSE values remain within 2.68 dB to 2.88 dB, with even smaller inter-regional variation.

These consistent RMSE values, all constrained within 0.2 to 0.4 dB across regions, strongly support the effectiveness of the proposed hybrid model. By combining deterministic path loss features, semantic image-based context, and tabular radio KPIs, the model achieves robust spatial performance. This ensures its reliability in diverse and previously unseen urban environments.

6.9 Robustness Analysis of Tabular Features

To evaluate the robustness of the hybrid model to variations in the tabular input features, a series of controlled experiments was conducted. These tests aim to address the following question:

How sensitive is the model to noise, missing features, or incomplete input vectors?

This is an essential consideration for real-world deployments, where sensor readings or network logs may be noisy, incomplete, or partially unavailable. Understanding the model's behaviour under such conditions helps assess its reliability and generalisation capability in the presence of uncertainty.

Three variations of the tabular data were evaluated in this analysis:

[13] `https://pandas.pydata.org/docs/reference/api/pandas.cut.html`

- **Normal:** The standardised tabular feature set, including RSRQ, SNR, CQI, RSSI, serving cell distance, and throughput. This configuration serves as the baseline for comparison.
- **Noisy:** To simulate measurement errors and variability in signal quality indicators, Gaussian noise with a standard deviation of $\sigma = 0.1$ was injected into all tabular features. This setup evaluates the model's tolerance to data corruption.
- **Normal Without CQI:** The CQI feature was removed from the input vector. This condition examines how the model compensates when a relevant input is missing, simulating incomplete data logs or unavailable sensor readings.

In the noisy condition, Gaussian noise was injected after feature standardisation, ensuring that the perturbations were defined relative to the normalised scale rather than the raw measurement units (e.g., dB). As a result, the deviations introduced are proportional to the statistical variability of each feature, maintaining consistency across inputs. The noise level was set to $\sigma = 0.1$, which was chosen to represent a moderate degree of measurement error. This value introduces sufficient variability to challenge the model's predictive capacity without producing unrealistic distortions in the input data.

All configurations used the same model architecture and training hyperparameters, which are fully detailed in Table 5.2. This ensures fair comparison across the experimental conditions. The evaluation metric used was the RMSE. The results are summarised in Table 6.8.

Experiment Condition	**Average RMSE (dB)**	**Std Dev (dB)**
Normal Features	2.77	0.07
Noisy Features	3.21	0.68
Normal Features Without CQI	2.85	0.06

Table 6.8: Expected RMSE results under different tabular input conditions.

The results clearly demonstrate that both noise and missing features have a negative impact on the model's prediction accuracy. The *Noisy Features* configuration yielded the highest RMSE and standard deviation, indicating that the model is sensitive to measurement variability and less robust to corrupted inputs.

Interestingly, the removal of the CQI feature resulted in a smaller increase in RMSE compared to the noisy condition. This suggests that random corruption across all input features is more detrimental to model accuracy than the absence of a single feature, since the network can partially compensate for the missing CQI by relying on the remaining indicators (e.g., RSRQ, SNR, RSSI). Moreover, the noisy condition produced a considerably higher standard deviation of RMSE across folds. This increased variability arises from the stochastic nature of the injected noise, reflecting the uncertainty and instability introduced by corrupted measurements.

The *Without CQI* condition also led to a noticeable increase in RMSE, albeit less pronounced. This confirms the importance of the CQI feature as a relevant input for predicting RSRP, given its role in characterising link quality.

Nonetheless, despite these degradations, the model maintained reasonable performance across all configurations, suggesting that the hybrid architecture exhibits a degree of resilience in its learned representations. This robustness is crucial for ensuring reliable operation in real-world scenarios where input completeness and precision cannot always be guaranteed.

From a practical standpoint, the noisy configuration is highly representative of real-world deployment scenarios, where network measurements are rarely error-free. Sensor imprecision, interference, reporting errors, and corrupted signal logs can all introduce distortions similar to those simulated in this setup. By explicitly evaluating robustness under noisy conditions, the analysis provides evidence of the hybrid model's reliability and its ability to operate effectively when exposed to imperfect data inputs.

Chapter 7

Discussion

This chapter presents a comprehensive discussion of the experimental results, highlighting the insights gained and their implications for the signal strength prediction task. The discussion is structured around model performance, spatial generalisation, and the added value of incorporating physical propagation knowledge into the deep learning model.

7.1 Performance and Accuracy

The experimental results demonstrate that the proposed hybrid model consistently achieved prediction errors below 3 dB at all folds of cross-validation. This represents a significant improvement over the baseline approach proposed by Thrane et al. [1], who reported root mean squared errors of approximately 6 dB. The reduction in error by more than half highlights the effectiveness of the proposed architecture in capturing the complex propagation dynamics present in urban vehicular communication scenarios.

A key factor behind this improvement lies in the multimodal fusion of heterogeneous data sources. By combining semantic map-based features extracted from height-map images, tabular key performance indicators (RSRQ, SNR, CQI, RSSI, throughput, and serving cell distance), and physical priors derived from path-loss models, the hybrid network was able to exploit complementary information that is otherwise ignored in single-modality approaches. This integration enabled the model to capture both the deterministic trends of radio propagation and the stochastic variations caused by obstacles, shadowing, and multipath effects, leading to superior accuracy and robustness.

7.2 Robustness and Generalisation

Another key observation is the robustness of the proposed model when evaluated in different distance quartiles. The RMSE remained consistently low, with only minor variations of approximately 0.3 dB between short-range and long-range intervals. Even in the most distant range (greater than 1200 m), the prediction error stayed below 3 dB, demonstrating that the method scales effectively with increasing distance from the serving base station. This stability is particularly relevant for vehicular communication systems, where users frequently move between near-field and far-field conditions.

In addition to distance-based stability, the model also demonstrated strong spatial consistency. When evaluated in different city regions, the performance remained uniform, indicating that the network did not fit too tightly to specific geographical areas. This suggests that the hybrid architecture is capable of generalising well to unseen environments, preserving its predictive accuracy across diverse propagation

scenarios. Such resilience to geographical variability is a crucial property for real-world deployment in intelligent transportation systems, where scalability and robustness are fundamental requirements.

The achieved stability across distance quartiles has direct implications for real-world vehicular communication systems. In particular, maintaining a consistently low RMSE below 3 dB ensures predictable QoS, which is a key requirement for safety-critical operations in autonomous driving. This level of accuracy can support more reliable handover decisions in dense urban deployments, where frequent transitions between cells are inevitable. Furthermore, the robustness of the model across both near-field and far-field conditions enhances its applicability to V2X scenarios, where stable connectivity is essential for cooperative perception, collision avoidance, and route optimisation in heterogeneous urban environments.

7.3 Spatial Performance Discussion

To assess the model's ability to generalise across different geographical regions, we conducted a spatial evaluation using two strategies: (1) one-dimensional binning based on longitude and (2) two-dimensional binning using both latitude and longitude. The motivation behind this test is to validate whether the model performs consistently across urban areas or if its accuracy deteriorates in specific locations, which would indicate a lack of spatial robustness.

The results of this analysis are summarised in Tables 6.6 and 6.7. These tables report the RMSE across different spatial bins for each fold of the cross-validation. In general, the average RMSE per longitude region remained within a stable range, demonstrating that the model maintained performance consistency along the longitudinal axis. However, the two-dimensional spatial analysis revealed greater variance, indicating that the model struggles more with generalisation when both latitude and longitude change simultaneously.

This suggests that, while the hybrid model captures spatial and semantic dependencies effectively, its predictive performance is still influenced by the spatial distribution of the training data. Certain regions with fewer data samples or more complex urban structures may introduce greater prediction errors. These findings emphasise the importance of spatially diverse training data when applying learning-based models to real-world radio propagation scenarios.

Moreover, these spatial evaluations provide insights into where the model might need refinement. For instance, incorporating auxiliary geospatial features such as building density, road networks, or terrain elevation could improve spatial generalisation. Future work could also explore spatially-aware loss functions or domain adaptation techniques to mitigate regional performance discrepancies.

7.4 Comparison with Empirical and Physical Models

A critical aspect of the evaluation is the comparison between the proposed hybrid approach and traditional path-loss models derived from empirical or physical formulations. Classical models, such as those based on 3GPP TR 38.901 or COST-Hata formulations, tend to degrade significantly with distance. As the propagation path increases, these models fail to capture local effects such as shadowing, multipath interference, and the impact of urban morphology, resulting in errors that often exceed 10 dB.

In contrast, the proposed hybrid model maintained stable accuracy across all distances, consistently achieving RMSE values below 3 dB. This robustness can be attributed to the multimodal fusion of information: while the physical priors provide general propagation behaviour, the incorporation of semantic

map data introduces awareness of the urban environment, including the presence of streets, open areas, and building structures. These features enable the model to adjust its predictions to various urban contexts, capturing effects that purely theoretical models cannot account for.

The results clearly highlight the benefits of this integration: by combining theoretical propagation principles with data-driven corrections informed by real-world features, the hybrid model effectively bridges the gap between physical accuracy and contextual adaptability, outperforming classical baselines in both short-range and long-range scenarios.

7.5 Limitations

Despite the promising results obtained, several limitations of the proposed hybrid approach should be acknowledged:

- **Dataset size and geographical scope:** The experiments were conducted using measurements collected in Cork, Ireland, covering a single urban deployment. While this setting is representative of a dense metropolitan area, the restricted geographical scope may limit the generalizability of the findings to other environments, such as rural, suburban, or distinct cities with different morphologies.
- **Computational cost of multimodal training:** The integration of semantic maps, tabular features, and physical priors increases both memory consumption and training time. Although the model achieved strong performance, the computational overhead of multimodal deep learning could limit scalability in resource-constrained settings.
- **Dependency on map quality:** The approach relies heavily on semantic features extracted from OSM. Inaccuracies or missing data in the map sources can introduce errors into the height map representations, potentially affecting prediction performance negatively.
- **Temporal variability and map freshness:** The model assumes static map conditions, which may not reflect recent urban changes such as roadworks. This temporal mismatch may introduce discrepancies between the semantic map inputs and real-world signal propagation conditions.
- **Sensitivity to noise:** The model's reliance on tabular features like signal quality indicators (RSRQ, RSSI) makes it susceptible to measurement noise or sensor errors. Without robust preprocessing or noise-handling strategies, these inconsistencies may affect prediction reliability.

These limitations underline the importance of further validation across broader datasets, as well as optimisations in model efficiency and robustness against noisy or incomplete map data.

Chapter 8

Conclusion

This dissertation investigated the problem of reliably predicting signal strength in vehicular environments, with a particular focus on the RSRP. The overarching objective was to design and validate a hybrid deep learning framework that integrates semantic geographic features, radio indicators, and physical priors to achieve robust and generalisable predictions suitable for autonomous vehicle communications.

The main contribution of this dissertation is the hybrid predictive framework resulting from the development of a model that combines semantic OpenStreetMap imagery, tabular radio features (RSRQ, SNR, CQI, RSSI, throughput), and deterministic path loss predictions into a unified learning-based architecture.

The proposed framework was comprehensively validated, considering an extensive set of experiments, using the publicly available dataset [13], with analysis across distance intervals, quartile-based evaluations, and comparative benchmarking against the state-of-the-art approach of Thrane et al. [1].

The obtained results highlight several findings. First, the proposed hybrid model achieved an average RMSE of approximately 2.7 dB, confirming its ability to capture both spatial and signal-domain patterns. Second, the error remained stable across varying distance quartiles, with only marginal increases beyond 1200 m, demonstrating robustness even in long-range propagation scenarios. Third, in direct comparison with the benchmark of Thrane et al. [1], our approach delivered improved predictive accuracy, confirming the added value of integrating semantic and physical knowledge within the same predictive pipeline.

Despite these promising outcomes, some limitations should be acknowledged. The reliance on OpenStreetMap data, while lightweight and scalable, introduces dependency on the completeness and consistency of volunteered geographic information. Additionally, although the model was validated on a realistic dataset, further testing across diverse cities, frequencies, and network deployments is necessary to establish its generalizability fully. These limitations motivate several future research directions, including the incorporation of temporal dynamics, digital twin integration, and uncertainty-aware modelling.

Ultimately, this dissertation not only makes a methodological contribution but also takes a practical step towards communication-aware autonomous vehicles. By integrating empirical modelling, semantic geospatial data, and data-driven learning, this work lays the foundation for more resilient, interpretable, and scalable vehicular networks in the 5G and beyond era.

8.1 Future Work

Building on the identified limitations, several directions for future research are suggested:

- **Cross-City Generalisation:** Evaluate the model on datasets collected from different cities or

countries to test the transferability of the learned features across diverse urban morphologies.

- **Real-World Deployment Scenarios:** Explore the model's integration into vehicular platforms for dynamic signal strength prediction in V2X environments.
- **Model Optimisation:** Investigate lightweight versions of the model for deployment in resource-constrained environments, including pruning, quantisation, or knowledge distillation techniques.
- **Data Augmentation and Fusion:** Integrate temporal data, 3D urban models, and sensor inputs such as accelerometer or gyroscope readings to further improve prediction accuracy.
- **Explainability and Interpretability:** Develop explainable AI tools to better understand how the model combines image and tabular features and to assist in trustworthy decision-making.

Overall, these research directions aim to enhance the generalizability and robustness of the proposed hybrid model, thereby promoting its adoption in large-scale and diverse vehicular communication scenarios.

Bibliography

[1] Jakob Thrane et al. “Deep learning-based signal strength prediction using geographical images and expert knowledge”. In: *GLOBECOM 2020 - 2020 IEEE Global Communications Conference* (2021). DOI: 10.1109/GLOBECOM42002.2020.9322089. URL: https://ieeexplore.ieee.org/document/9410462.

[2] Siva Priya Thiagarajah R. Mardeni. “Optimised COST-231 Hata Models for WiMAX Path Loss in Urban and Suburban Environments”. In: *Proceedings of the International Conference on Wireless Broadband and Ultra Wideband Communications (AusWireless)*. IEEE, 2009, pp. 1–6. DOI: 10.1109/AUSWIRELESS.2009.5348302. URL: https://ieeexplore.ieee.org/document/5348302.

[3] Ruslan V. Akhpashev and Andrey V. Andreev. “COST 231 Hata Adaptation Model for Urban Conditions in LTE Networks”. In: *2016 17th International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices (EDM)*. IEEE, 2016. DOI: 10.1109/EDM.2016.7538693. URL: https://ieeexplore.ieee.org/abstract/document/7538693.

[4] Henrik Lehrmann Christiansen Jakob Thrane Darko Zibar. “Model-Aided Deep Learning Method for Path Loss Prediction in Mobile Communication Systems at 2.6 GHz”. In: *IEEE Access* 8 (2021). DOI: 10.1109/ACCESS.2020.2964103. URL: https://ieeexplore.ieee.org/abstract/document/8950164.

[5] Raied Caromi Thao T. Nguyen Nadia Yoza-Mitsuishi. “Deep Learning for Path Loss Prediction at 7 GHz in Urban Environment”. In: *IEEE Access* 11 (2023), pp. 33498–33508. DOI: 10.1109/ACCESS.2023.3264230. URL: https://ieeexplore.ieee.org/document/10091527.

[6] Marlon Jeske et al. “Enhancing Reference Signal Received Power Prediction Accuracy in Wireless Outdoor Settings: A Comprehensive Feature Importance Study”. In: *IEEE Transactions on Antennas and Propagation* (2025). DOI: 10.1109/TAP.2025.3576492. URL: https://ieeexplore.ieee.org/document/11030236.

[7] Qiuming Zhu et al. “3GPP TR 38.901 Channel Model”. In: *Wiley 5G Ref: The Essential 5G Reference Online*. Wiley, 2021. DOI: 10.1002/9781119471509.w5gref048. URL: https://www.researchgate.net/publication/348790466_3GPP_TR_38901_Channel_Model.

[8] Claudine Badue, Rânik Guidolini, and Raphael Vivacqua Carneiro. “Self-driving cars: A survey”. In: *Expert Systems with Applications* 165 (2021), pp. 1–8. DOI: https://doi.org/10.1016/j.eswa.2020.113816.

[9] Laurène Claussmann et al. “A review of motion planning for highway autonomous driving”. In: *IEEE Transactions on Intelligent Transportation Systems* 21.5 (2020), pp. 1826–1848. DOI: 10.1109/TITS.2019.2913998. URL: https://ieeexplore.ieee.org/abstract/document/8715479.

[10] Mate Boban et al. “Connected Roads of the Future: Use Cases, Requirements, and Design Considerations for Vehicle-to-Everything Communications”. In: *IEEE Vehicular Technology Magazine* 13.3 (2018), pp. 110–123. DOI: 10.1109/MVT.2017.2777259. URL: https://ieeexplore.ieee.org/abstract/document/8410403.

[11] Tibor Petrov et al. “A Performance Benchmark for Dedicated Short-Range Communications and LTE-Based Cellular-V2X in the Context of Vehicle-to-Infrastructure Communication and Urban Scenarios”. In: *Sensors* 21 (2021), p. 5095. DOI: https://doi.org/10.3390/s21155095. URL: https://www.mdpi.com/1424-8220/21/15/5095.

[12] Mongi Ben Ameur et al. “Machine learning for improved path loss prediction in urban vehicle-to-infrastructure communication systems”. In: *Frontiers in Artificial Intelligence* (2025). DOI: 10.3389/frai.2025.1597981.

[13] Darijo Raca et al. “Beyond throughput: A 4G LTE dataset with channel and context metrics”. In: *MMSys ’18: Proceedings of the 9th ACM Multimedia Systems Conference*. 2018, pp. 460–465. DOI: https://doi.org/10.1145/3204949.32081. URL: https://dl.acm.org/doi/abs/10.1145/3204949.3208123.

[14] Omar Ahmadien et al. “Predicting Path Loss Distribution of an Area From Satellite Images Using Deep Learning”. In: *IEEE Access* 8 (2020), pp. 64982–64991. DOI: 10.1109/ACCESS.2020.2985929. URL: https://doi.org/10.1109/ACCESS.2020.2985929.

[15] Haiyao Yu et al. “Distributed Signal Strength Prediction using Satellite Map empowered by Deep Vision Transformer”. In: *2021 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2021. DOI: 10.1109/GCWkshps52748.2021.9682021. URL: https://ieeexplore.ieee.org/document/9682021.

[16] Enes Krijestorac; Samer Hanna; Danijela Cabric. “Spatial Signal Strength Prediction using 3D Maps and Deep Learning”. In: *ICC 2021 - IEEE International Conference on Communications*. 2021. DOI: 10.1109/ICC42927.2021.9500970. URL: https://ieeexplore.ieee.org/abstract/document/9500970.

[17] Enes Krijestorac; Samer Hanna; Danijela Cabric. “Spatial Signal Strength Prediction using 3D Maps and Deep Learning”. In: *ICC 2021 - IEEE International Conference on Communications*. 2021. DOI: 10.1109/ICC42927.2021.9500970. URL: https://ieeexplore.ieee.org/abstract/document/9500970.

[18] Saeed Hamood Alsamhi et al. “Predictive Estimation of Optimal Signal Strength From Drones Over IoT Frameworks in Smart Cities”. In: *IEEE Transactions on Mobile Computing* 22.1 (2021), pp. 402–416. DOI: 10.1109/TMC.2021.3074442. URL: https://ieeexplore.ieee.org/document/9410462.

[19] Jie Zhang Kehai Qiu Stefanos Bakirtzis. “Deep Learning-Based Path Loss Prediction for Outdoor Wireless Communication Systems”. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2023. DOI: 10.1109/ICASSP49357.2023.10095501. URL: https://www.researchgate.net/publication/370323291.

[20] Irma Zakia Muhammad Brata. “Path Loss Estimation at Sub-6 GHz and Millimeter Wave Frequencies Using Fine-Tuning”. In: *IEEE Access* 12 (2024), pp. 138142–138154. DOI: 10.1109/ACCESS.2024.3465653. URL: https://ieeexplore.ieee.org/document/10685351.

[21] Nihat İnanç Majd Alkorabi Alireza Souri. "Deep Learning Algorithms for Autonomous Vehicle Communications: Technical Insights and Open Challenges". In: *Concurrency and Computation: Practice and Experience* (July 2025). DOI: 10.1002/cpe.70218. URL: https://doi.org/10.1002/cpe.70218.

[22] Dimitar Minovski et al. "Throughput Prediction Using Machine Learning in LTE and 5G Networks". In: *IEEE Transactions on Mobile Computing* 22.3 (2021), pp. 1825–1840. DOI: 10.1109/TMC.2021.3099397. URL: https://ieeexplore.ieee.org/abstract/document/9495144.

[23] Jose M Peña Jamie Mersh Erik Larsson. "Sensitivity of Coverage Prediction to Propagation Model Parameters". In: *24th International Conference on Systems Engineering (ICSEng)*. Digitala Vetenskapliga Arkivet, 2025. URL: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1967380&dswid=-53.

[24] Aliza Bt Sarlan Umar D. Maiwada* Kamaluddeen U. Danyar. "An Improved 5G Mobility Handover Efficient by Creating a Digital Twin Network: A Review". In: *International Journal of Software Engineering and Computer Systems* 10.2 (2025), p. 1494. DOI: https://doi.org/10.15282/ijsecs.10.2.2024.11.0129. URL: https://journal.ump.edu.my/ijsecs/article/view/10466.

[25] Anastasios Foliadis et al. "Deep Learning-Based Positioning With Multi-Task Learning and Uncertainty-Based Fusion". In: *IEEE Transactions on Machine Learning in Communications and Networking* 2 (2024), pp. 1127–1141. DOI: 10.1109/TMLCN.2024.3441521. URL: https://ieeexplore.ieee.org/document/10632202.

[26] Mohammad Java Taheri et al. "Predicting Drive Test Results in Mobile Networks Using Optimization Techniques". In: *arXiv preprint arXiv:2502.09305* (2025). URL: https://arxiv.org/abs/2502.09305.

[27] Ella Peltonen Vishaka Basnayake Prabash Rathnayake. "Vehicle-to-Everything Services in 3GPP 5G Networks: An Empirical Analysis". In: *Proceedings of EdgeSys '25*. ACM, 2025. DOI: 10.1145/3721888.3722094. URL: https://doi.org/10.1145/3721888.3722094.

[28] Alexey Dosovitskiy et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: *arXiv preprint arXiv:2010.11929*. arXiv, 2021. URL: https://arxiv.org/abs/2010.11929.

[29] Raphaël Couturier Stéphane Cuenat. "Convolutional Neural Network (CNN) vs Vision Transformer (ViT) for Digital Holography". In: *2022 2nd International Conference on Computer, Control and Robotics (ICCCR)*. Shanghai, China: IEEE, Mar. 2022, pp. 645–649. DOI: 10.1109/ICCCR54399.2022.9790134. URL: https://doi.org/10.1109/ICCCR54399.2022.9790134.

[30] Jorge Bernardino José Maurício Inês Domingues. "Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review". In: *Applied Sciences* 13.11 (2023). DOI: 10.3390/app13095521. URL: https://www.mdpi.com/2076-3417/13/11/5521.