



**University of
Nottingham**

UK | CHINA | MALAYSIA

Vehicle-to-Vehicle Communication: Design, Performance, and Disruption Mitigation in Real-World Environment

Xinao Wang

PhD Engineering Surveying and Space Geodesy

Abstract

In recent years, the field of vehicular automation has undergone rapid development thanks to advancements in technology, extensive marketing efforts, and regulatory pressure for electrification. Numerous manufacturers have introduced models with varying degrees of autonomy. Among the developments, the technology known as Vehicle-to-Vehicle (V2V) communication holds great promise for further enhancing the capabilities of autonomous vehicles, such as improved traffic safety, efficiency, and environmental benefits.

This thesis investigates the performance of 802.11p-based V2V communication in real-life scenarios, and explores potential practical applications such as GNSS correction data broadcasting to improve the positioning accuracy of nearby vehicles, and enhancing communication robustness by preemptively predicting potential disruptions with the assistance of Machine Learning (ML) models. A custom V2V On-board Unit (OBU) hardware platform was developed, and real-world multi-vehicle outdoor experiments were planned and carried out. The collected data was examined and used to train a number of ML models, and their performance was compared.

The experiments revealed that the custom OBU was fully functional, and signal quality and communication range were observed to be affected by real-world imperfections. The GNSS correction data broadcasting was shown to notably increase the positioning accuracy of nearby vehicles, and the ML models trained from Key Performance Indicators (KPIs) demonstrated excellent prediction accuracy, allowing pre-emptive actions to be taken to reduce the downtime from communication disruption.

Acknowledgements

I wish to extend my heartfelt gratitude to my supervisors, Professor Xiaolin Meng and Professor Rong Qu, for their unwavering support and astute guidance throughout the course of my PhD journey. Professor Meng's enthusiasm and insightful discussions were instrumental in the progression of my research. Professor Rong Qu has been an immeasurable source of wisdom, especially in guiding me through the machine learning aspects of my research. I am particularly thankful for the invaluable opportunities my supervisors afforded me to engage with professionals from various organisations through seminars and international conferences, which significantly enriched my academic experience.

Furthermore, I would like to express my sincere thanks to the esteemed members and colleagues of the Nottingham Geospatial Institute, including Dr. Joseph Walters, Dr. Qiyi He, Dr. Yijian Cui, and Dr. Simon Roberts. Their generosity in sharing their knowledge, as well as their constructive academic discussions and support, has been invaluable.

Additionally, I am grateful for the assistance and camaraderie of my friends and colleagues in the UK, including Mr. George Ye, Mrs. Jun Ye, and Mr. Keith Geary, whose unwavering support has been a great asset during this project.

Above all, my profound gratitude is extended to my parents, Dr Qing Wang and Mrs Jing Liu. Their unconditional love and support over the years have been guiding me through this challenging but fulfilling journey, without which this achievement would not have been possible.

List of Publications

- Wang, Xinao, Meng, Xiaolin, Ye, George, "Design of a Versatile Vehicle-to-Everything (V2X) Communication Platform for Autonomous Vehicular Environments," *Proceedings of the 35th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2022)*, Denver, Colorado, September 2022, pp. 397-416.
- Xinao Wang, Joseph Walters, "LiDAR-based Terrain Recognition in Off-road Mobile Robot," *Proceedings of the 33rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2020)*, September 2020, pp. 2590-2604.
- Roberts, S.; Meng, X.; Xu, C.; Wang, X.; Cui, Y.; Ye, G.; Bonenberg, L. CoDRIVE – Delivering High Accuracy, Ubiquitous Positioning Through Combined Radio Navigation and Inertial Sensing Technologies. 2020, 2020090037.
- Meng, Xiaolin, Xu, Chang, Wang, Xinao, Roberts, Simon, Cui, Yijian, He, Qiyi, Chen, Qusen, "First Testing Results in using V2X Technology to Enhance N-RTK Availability in the UK," *Proceedings of the 30th International Technical Meeting of the Satellite Division of The Institute of Navigation*, Portland, 2019, Oregon, pp. 700-717.

List of abbreviations

V2V	Vehicle-To-Vehicle Communication
V2X	Vehicle-To-Everything Communication
ITS	Intelligent Transportation Systems
DSRC	Dedicated Short-Range Communications
C-V2X	Cellular V2X
OBU	On-Board Unit
GNSS	Global Navigation Satellite System
SBC	Single-Board Computer
RTK	Real-Time Kinematic Positioning
CAN	Controller Area Network
IMU	Inertial Measurement Unit
MEMS	Microelectromechanical Systems
HDMI	High-Definition Multimedia Interface
SD	Secure Digital
PCB	Printed Circuit Board
EDA	Electronic Design Automation
EMI	Electromagnetic Interference
DC	Direct-Current
AC	Alternating Current
IC	Integrated Circuit
KPI	Key Performance Indicator
FOSS	Free and Open-Source Software
GIS	Geographic Information System
RSSI	Received Signal Strength Indicator
dBm	Decibels Referenced to One Milliwatt
SNR	Signal-To-Noise Ratio

PER	Packet Error Rate
PRR	Packet Reception Rate
IPG	Interpacket Gap
CORS	Continuous Operating Reference Station
CSV	Comma Separated Values
FSPL	Free-Space Path Loss
LOS	Line-Of-Sight
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
k-NN	K-Nearest Neighbour
REPTree	Reduced Error Pruning Tree

Table of Contents

Chapter 1: Introduction	1
1.1. Research Motivations	1
1.2. Research Objectives	3
1.3. Research Methods and Procedures	4
1.4. Contributions of the Research	5
1.5. Thesis Structure	7
Chapter 2: Literature Review	9
2.1. Vehicle-to-Everything Communications	9
2.1.1. Background	9
2.1.2. Working Principle and Benefits	10
2.1.3. Early History	11
2.1.4. Dedicated Short-Range Communications (DSRC).....	13
2.1.5. Cellular V2X (C-V2X)	16
2.1.6. Comparison and Integration of DSRC and C-V2X	17
2.1.7. On-Board Unit	20
2.1.8. Deployment Status	21
2.1.9. Potential Issues.....	23
2.1.10. Summary	25
2.2. Machine Learning Applications.....	27
2.2.1. ZeroR and OneR.....	27
2.2.2. k-Nearest Neighbour	27
2.2.3. Decision Tree	29
2.2.4. Random Forest	31
2.3. Machine Learning in Intelligent Transportation Systems (ITS).....	32
Chapter 3: V2V OBU Design	35
3.1. Overview.....	35
3.2. Design Goals and Requirements	36
3.3. Component Selection	37
3.3.1. Processor.....	38
3.3.2. V2X Transceiver	42
3.3.3. GNSS Receiver	43

3.3.4.	Cellular Modem.....	45
3.3.5.	Wi-Fi / Bluetooth.....	46
3.3.6.	CAN Bus Transceiver	47
3.3.7.	Inertial Measurement Unit	48
3.3.8.	Physical Interfaces	49
3.4.	Printed Circuit Board	50
3.4.1.	Overview	51
3.4.2.	PCB Elements	51
3.4.3.	Layer Stack-up	53
3.4.4.	Impedance Control	55
3.4.5.	Differential Signals.....	56
3.5.	All-in-One OBU.....	57
3.6.	Modular OBU	59
3.7.	Initialisation Procedures	63
Chapter 4:	Essential Tools and Metrics	65
4.1.	Software Packages and Services	65
4.1.1.	Python.....	65
4.1.2.	QGIS.....	66
4.1.3.	OpenStreetMap.....	67
4.1.4.	RTKLIB	68
4.1.5.	Weka	69
4.2.	Key Performance Indicators.....	69
4.2.1.	Received Signal Strength Indicator (RSSI)	70
4.2.2.	Signal-to-Noise Ratio	71
4.2.3.	Packet Error Rate (PER).....	71
4.2.4.	Packet Reception Rate (PRR) / Communication Range.....	72
4.2.5.	Interpacket Gap (IPG)	73
Chapter 5:	Preliminary Test at Wollaton Park	75
5.1.	Test Objectives	75
5.2.	Test Plan and Procedures.....	75
5.3.	Dataset Visualisation and Analysis.....	77
5.4.	Summary.....	82

Chapter 6:	Multi-Vehicle Road Test	84
6.1.	Test Objectives	84
6.2.	Test Plan and Procedures.....	85
6.2.1.	Route Planning	85
6.2.2.	Vehicles and Equipment.....	87
6.2.3.	Logistics and Safety.....	89
6.2.4.	Payload Structure and Data Path.....	90
6.2.5.	Cellular Environment	93
6.2.6.	Data Processing and Visualisation	93
6.3.	Test Results	94
6.3.1.	Driving Environments	95
6.3.2.	Dataset Overview	95
6.3.3.	Signal Quality	96
6.3.4.	Communication Range.....	105
6.3.5.	GNSS Positioning Accuracy.....	109
6.3.6.	Cellular Environment	117
6.4.	Summary.....	122
Chapter 7:	Interruption Mitigation with Machine Learning	124
7.1.	Overview.....	124
7.2.	Dataset Preparation	125
7.2.1.	Class Variable Creation	125
7.2.2.	Parameter Selection	126
7.2.3.	Class Balancing	129
7.3.	Evaluation Methods	133
7.3.1.	Prediction Outcomes	133
7.3.2.	Validation Methods	134
7.3.3.	Confusion Matrix.....	135
7.3.4.	Accuracy, Recall, and Precision.....	136
7.3.5.	F1 Score.....	137
7.3.6.	Training Time.....	137
7.4.	Model Training and Evaluation	138
7.4.1.	ZeroR.....	139

7.4.2.	OneR	141
7.4.3.	K-Nearest Neighbour	143
7.4.4.	J48 Decision Tree	145
7.4.5.	REPTree	148
7.5.	Summary.....	152
Chapter 8:	Conclusions	154
8.1.	Main Contributions	154
8.2.	Limitations and Areas for Improvement.....	158
8.3.	Future Work	159
8.4.	Summary.....	160
Chapter 9:	Bibliography	162

List of Tables

Table 1: Sample decision tree dataset	30
Table 2: OBU Specification.....	63
Table 3: Outward Trip Statistics	78
Table 4: Return Trip Statistics.....	81
Table 5: Parameters of road trial dataset	96
Table 6: Statistics of vehicle following distance (meters).....	97
Table 7: Statistics of vehicle Speed (MPH)	97
Table 8: Road trial signal quality statistics: Overall	101
Table 9: Road trial signal quality statistics: Built-up Areas	101
Table 10: Road trial signal quality statistics: Inter-Urban.....	101
Table 11: Road trial signal quality statistics: High-Speed	102
Table 12: OBU GNSS Solution Types.....	112
Table 13: OBU GNSS performance statistics, overall.	113
Table 14: OBU GNSS performance statistics, Built-up Area.....	113
Table 15: OBU GNSS performance statistics, Inter-Urban.....	114
Table 16: OBU GNSS performance statistics, High-Speed.	114
Table 17: Cellular Service Type Statistics.....	119
Table 18: Cellular Signal Quality Statistics	119
Table 19: Correlation coefficient between signal quality parameters.	127
Table 20: Updated parameters of the training dataset.....	129
Table 21: Statistics of class labels.....	131
Table 22: A sample confusion matrix	135
Table 23: ZeroR Classification Results	140
Table 24: OneR Classification Results.....	141
Table 25: Influence of K value on KNN model accuracy.....	144
Table 26: KNN Classification Results with K = 45	144
Table 27: J48 Decision Tree Classification Results	147
Table 28: REPTree with depth limit of 4 classification results	150

Table 29: REPTree with depth limit of 2 classification results151

List of Figures

- Figure 1: Example of k-NN classification28
- Figure 2: A sample decision tree.....30
- Figure 3: UDOO x86 Computer39
- Figure 4: Odroid-H2 Computer39
- Figure 5: Raspberry Pi 3 Model B+42
- Figure 6: Raspberry Pi Compute Module 3+42
- Figure 7: A Typical Two-layer PCB Construction51
- Figure 8: Production Process of a Simple PCB53
- Figure 9: Four-layer stack-up used in this project.....55
- Figure 10: Illustration of Differential Signalling57
- Figure 11: All-in-one OBU Circuit Design.....59
- Figure 12: All-in-one OBU assembled PCB.59
- Figure 13: The PC/104 OBU62
- Figure 14: Wollaton Park test environment76
- Figure 15 : Visualisation of Test Route77
- Figure 16: Received Power Level, Outward Trip.77
- Figure 17: Received Noise Level, Outward Trip.78
- Figure 18: Signal-to-Noise Ratio, Outward Trip.78
- Figure 19: Received Power Level, Return Trip.80
- Figure 20: Received Noise level, Return Trip.80
- Figure 21: Signal-to-Noise Ratio, Return Trip.80
- Figure 22: Proposed Test Route87
- Figure 23: Road trial equipment setup89
- Figure 24: Data Path of DSRC Transmitter92
- Figure 25: Data Path of DSRC Receiver.....93
- Figure 26: Road trial received power level99
- Figure 27: Road trial received noise level.....99
- Figure 28: Road trial signal-to-noise ratio99

Figure 29 : Road trial IPG and PER	107
Figure 30: OBU GNSS Solution Type.....	111
Figure 31: OBU GNSS Performance Comparison.	112
Figure 32: Accuracy improvements in Built-up Areas	115
Figure 33: Road Trial Cellular Service Type.....	119
Figure 34: Road Trial Cellular Signal Quality	119
Figure 35: Visualisation of correlation between signal quality parameters.....	128
Figure 36: Class labels visualisation	130
Figure 37: Histogram of class labels	131
Figure 38: Instances in the training dataset	139
Figure 39: OneR model rules	141
Figure 40: Constructed J48 Decision Tree	146
Figure 41: REPTree with depth limit of 4.....	149
Figure 42: REPTree with depth limit of 2.....	151

Chapter 1: Introduction

1.1. Research Motivations

The rapid growth in vehicular traffic worldwide has led to an increasing demand for Intelligent Transportation Systems (ITS) to address the problems of congestion, road safety, and environmental sustainability. In this context, Connected and Autonomous Vehicles (CAVs) have emerged as a promising technology for revolutionizing the future of transportation. One of the key enabling technologies for CAV is Vehicle-to-Vehicle (V2V) communication, which allows vehicles to exchange information with each other via short-range radio or cellular networks [1].

By allowing vehicles to share information about their surroundings, V2V communication can be expected to significantly improve road safety by reducing the likelihood of collisions, as well as enhance traffic efficiency by enabling better coordination among vehicles [2]. Additionally, V2V communication can contribute to minimising environmental impact by optimising driving patterns and reducing fuel consumption and emissions [3].

Numerous studies over the past 40 years have envisioned the benefits of widely deployed V2V networks, however, it was only in the recent decade that technological advancement has made V2V communication feasible in the real world. Numerous pilot projects and field trials have been carried out across the globe to assess the feasibility and effectiveness of V2V networks, such as the European Union's Cooperative Intelligent Transport Systems (C-ITS) initiatives [4], the United States' Connected Vehicle Pilot Deployment Program [5], and China's Intelligent and Connected Vehicles (ICV) demonstrations [6]. These projects have provided valuable insights into the technical, regulatory, and social aspects of V2V communication deployment, helping to pave the way for broader implementation.

To facilitate the deployment of V2V communication systems, several standardisation bodies, such as the Institute of Electrical and Electronics Engineers (IEEE) and the European Telecommunications Standards Institute (ETSI), have developed standards and specifications for V2V communication technologies. For instance, the IEEE 802.11p standard provides the technical foundation for DSRC-based V2V communication, while the ETSI ITS-G5 standard defines the requirements for C-ITS applications in Europe. Additionally, regulatory authorities worldwide have been working on establishing legal and regulatory frameworks to govern the deployment of V2V communication systems, addressing issues such as spectrum allocation, security, and privacy.

The automotive industry has also been looking into gradually integrating V2V communication capabilities into their vehicles, forging partnerships between technology providers, infrastructure developers, and governing bodies. Several automotive manufacturers have already released vehicles with V2V capability in certain parts of the world such as China and Japan [7], and more manufacturers are planning to equip their future models with V2V communication capabilities [8].

While there has been numerous existing research on V2V communication, a notable portion of these studies have primarily been conducted within simulated environments. These investigations, though instrumental in providing insights into V2V systems performance, are inherently constrained by the assumptions and simplifications built into the simulation models, which might fail to accurately represent the unique complexities and challenges associated with deploying V2V systems in real-world situations.

This thesis seeks to bridge this gap in existing research by conducting a comprehensive evaluation of V2V communication systems under real-world conditions. It focuses not only on assessing the performance of these systems but also the process of their development, setup, and deployment in practical

scenarios. Many aspects of this technology are investigated, including cost, specification, availability, procurement, and the feasibility of developing custom experimental hardware platform with off-the-shelf communication modules. Detailed outdoor experiments are carried out with road-legal vehicles on public roads, and performance data is collected, analysed, and evaluated. Additional applications of V2V are also explored, such as broadcasting GNSS correction data to nearby vehicles, and using performance data to train machine learning models for pre-emptive mitigation of communication disruptions.

By focusing on these aspects, the research aims to provide a comprehensive understanding of the practical challenges and performance characteristics of V2V communication systems in realistic settings. This will help bridge the gap between simulation-based studies and real-world deployments and contribute to the development and implementation of more effective V2V communication solutions for connected and autonomous vehicles.

1.2. Research Objectives

The primary aim of this PhD study is to investigate the development, deployment, and the performance of V2V communication using Dedicated Short-Range Communications (DSRC) technology under real-world conditions, and to explore the potential use of machine learning models to classify and predict possible upcoming communication interruptions by monitoring Key Performance Indicators (KPIs) of V2V systems.

Objectives of this research are outlined below:

- Conduct a review of V2V communication, including history, method of operation, benefits, challenges, competing standards, deployment status, performance parameters, as well as current literatures, in order to gain an understanding of the current state of development and identify gaps and limitations of existing research.

- Investigate the functionality and key components of V2V On-Board Units (OBU), and to design, produce, and assemble a custom OBU for the real-world experiments.
- Plan and conduct comprehensive planning of real-world V2V experiments involving multiple vehicles on public roads, including route planning, logistics, data path structure, and data collection.
- Execute the outdoor experiments according to plan, perform data processing, visualisation, and analysis of collected data, identify and discuss the trends and observations in the results.
- Use the KPI from the experiment dataset to train a number of Machine Learning models in order to classify and predict communication interruption scenarios. Evaluate different models and identify the one with optimal performance.
- Summarise and discuss the outcome of the project, identify areas for improvements and topics for further research.

1.3. Research Methods and Procedures

The initial phase of this research will involve a comprehensive literature review, focusing on understanding the underlying principles of V2V communication, existing standards, and relevant research in the field. This will include a thorough examination of academic publications, technical reports, and standardisation documents, as well as a critical evaluation of the current state of knowledge and the identification of research gaps and opportunities.

Following the literature review, the research will proceed with the development of custom V2V hardware tailored to the specific requirements of this project. Including research and comparison to identify suitable hardware components, such as communication modules, antennas, processing units, peripheral devices, and the form factor of the device. Based on the selected components, Printed

Circuit Board (PCB) design will be developed, fabricated, assembled, and debugged to create a functional V2V communication hardware platform.

The next phase of the research will involve conducting real-world outdoor experiments to evaluate the performance of the custom V2V hardware. This will encompass detailed test planning, including the selection of test locations, definition of test routes, and planning of logistics and safety protocols, contents of payloads, and the performance parameters to collect for the dataset.

After the conclusion of the experiment, the research will focus on data processing, visualisation, and analysis. The collected data will be processed and analysed using appropriate statistical techniques and data visualisation tools to facilitate the identification of trends, patterns, and anomalies. The results will be discussed and compared to existing documents and literatures to assess the performance and reliability of the custom V2V hardware and to identify potential areas for improvement.

The final phase of the research will involve exploring the feasibility of using the machine learning models to identify and predict imminent communication interruption scenarios, so that action can be taken to switch to alternative means of communication preemptively. Suitable models will be suggested, parameters from the dataset will undergo preprocessing before being used to train the models, and their performance is evaluated and compared.

1.4. Contributions of the Research

This research project offers several unique contributions to the fields of V2V communication and CAVs.

A significant contribution of this research is the design of a novel experimental hardware platform that combines a wide range of wireless connectivity options, high-accuracy positioning, sensors, and physical interfaces in a compact and

modular form factor. The device is able to communicate wirelessly via DSRC, cellular network, RFID, Wi-Fi, and Bluetooth.

The device also features Real-Time Kinematics (RTK) capable GNSS receiver for high-accuracy positioning, inertial measurement units (IMU) to provide dead-reckoning, Controller Area Network (CAN) for communicating with vehicle internal networks, along with additional physical interface options.

This versatile hardware platform, custom designed for this project, is proven to be functional, reliable, and stable, and is suitable for a wide range of research activities and experiments involving V2V communication, as well as CAV positioning and control. The potential for commercialising the product further highlights its value to both academic and industrial applications.

Another contribution is the collection and analysis of data from real-world experiments, which provides valuable insights into the performance of V2V communication systems under real-world situations. The analysis of Key Performance Indicators (KPIs) such as communication range, effects of vehicle speed, following distance, presence of obstacles, data throughput, and latency offers a more comprehensive understanding of V2V performance, addressing the limitations of simulation-based studies.

This research also demonstrates that broadcasting GNSS correction data over V2V channels can notably improve the positioning accuracy of nearby vehicles, even if they do not have internet connections of their own. This finding illustrates one of the many practical benefits of V2V communication, enabling cost savings in data connection fees and highlighting the potential for further applications in real-world scenarios.

Additionally, by identifying the characteristics of KPIs during communication dropouts resulting from signal degradation due to obstacles or excessive following distances, a machine learning model has been developed to predict and detect communication disruptions before they occur. This innovative

approach reduces communication disruptions by enabling the system to switch to alternative communication methods, such as cellular networks or satellite communication, thereby enhancing the overall reliability of V2V communication systems.

In summary, this research project provides numerous unique ideas and findings, demonstrating the versatility and potential of V2V communication systems in real-world applications. These contributions serve as a solid foundation for further research, advancing the knowledge and understanding of V2V communication and its role in the development of more effective and reliable connected and autonomous vehicle systems.

1.5. Thesis Structure

The rest of this thesis is arranged as follows.

Chapter 2 provides a comprehensive review of the existing literature on V2V communication, including its background, operational principles, advantages, competing standards, infrastructure requirements, and future challenges. Furthermore, the chapter evaluates the strengths and weaknesses of several popular machine learning classification algorithms, laying the theoretical groundwork for the chapters that follow.

Chapter 3 details the design and development process of a custom on-board unit (OBU) hardware platform. It discusses design goals, hardware requirements, justification for the development of custom hardware, and key design considerations. It also presents the innovative incorporation of the PC/104 form factor to enhance modularity and upgradeability.

Chapter 4 introduces various software packages and digital tools used for data processing, visualisation, and analysis. It also defines the Key Performance Indicators (KPIs) for assessing system performance, setting the stage for the upcoming experimental chapters.

Chapter 5 reports on the preliminary outdoor test utilising the OBU hardware. This includes a detailed description of the test plans and procedures, and a discussion of the observed KPIs. The chapter validates the OBU hardware's functionality and stability, highlighting a maximum communication range of nearly 900 meters in open-air, line-of-sight conditions.

Chapter 6 details a more comprehensive outdoor experiment involving multiple road-legal vehicles navigating public roads under diverse traffic conditions and speeds. It presents and discusses the results, and investigates various parameters to find a practical communication range. This chapter also explores V2V's capability to transmit and broadcast GNSS correction data, and examines the cellular environment along the test route.

Chapter 7 delves into the use of machine learning classification models to predict potential disruptions in V2V communication. It outlines the dataset preprocessing and the training of five models, and discusses the feasibility of the idea, as well as the potential for improvements.

Finally, Chapter 8 presents the conclusion of the thesis, including a summary of the work completed, the contributions to the field, the identification of research limitations, and proposals for areas of further investigation.

Chapter 2: Literature Review

2.1. Vehicle-to-Everything Communications

2.1.1. Background

As we move towards an increasingly interconnected society thanks to the advance of the Internet and smart devices, much attention have been turned to the relatively untapped field of automobiles in recent years. It is believed that allowing vehicles on the road to exchange information with each other about their dynamics, such as position, speed, and heading, can lead to beneficial outcomes in terms of increased safety and efficiency.

It was pointed out that in the past decades there has been significant progress in the “passive safety” of automobile design, where new materials and structures were utilised to strengthen the passenger cell in the event of a collision [9]. In evolution, the next big step up in automotive safety was envisioned to rely on active, in-vehicle sensors to assist drivers in potentially dangerous situations. Real-world examples include the Electronic Stability Program, which can reduce the risk of losing traction, and radar sensors that provide automated emergency braking without user intervention [10].

Despite the capabilities of such sensors and the reduction of casualties as the result, most of the existing safety technologies focus more on “crash-imminent situations” rather than warning drivers ahead of time, and the majority of currently fitted safety sensors have a similar “range of vision” as a human driver. It was envisioned that it would be of great advantage if vehicles were able to exchange information with each other over longer distances [11]. This ability of communication can relay information about events beyond drivers “line of sight” and allow cooperation between vehicles in order to assist their decision making. This Vehicle-to-Vehicle communication, or V2V, would be ideal to further increase the safety of automobiles [9].

2.1.2. Working Principle and Benefits

Vehicle-to-Vehicle Communication can be described as “[using] on-board dedicated short-range radio communication devices to transmit messages about a vehicle’s speed, heading, brake status, and other information to other vehicles, and receive the same information”. The communication range can be up to 300 meters, allowing vehicles to “see” around corners or “through” other vehicles to locate potential hazards sooner than using human eyes or traditional sensors like cameras and radars. By allowing vehicles to communicate with each other, V2V can help prevent accidents by providing drivers with real-time information about potential hazards on the road ahead. Additionally, V2V system tends to be unaffected by weather and lighting conditions [9].

V2V communication can be also extended to include other entities such as roadside infrastructures (V2I), pedestrians (V2P) and networks (V2N). With the additional capabilities, the system can then be referred to as Vehicle-to-Everything (V2X) communication [2].

There are several notable scenarios where V2X communication can greatly reduce the risk of a traffic accident, such as where a driver is approaching an unseen disabled vehicle. With V2V, the offending vehicle can broadcast its information and warn other drivers hundreds of meters away, allowing them to take action much earlier. This would be especially beneficial in high-speed situations where radar and camera-based systems might not be able to provide a timely warning.

Another example is Blind Intersection Warning, where two cars approach an intersection but cannot see each other. Existing vision-based safety systems might not be able to address this scenario, while V2X-equipped vehicles can easily warn each other by communicating beyond “line of sight”. All in all, it was estimated that if deployed widely in United States, V2X technology has potential to prevent up to half a million crashes, and save tens of thousands of lives [10].

Another advantage of V2V is improved efficiency. It was proposed that with proper coordination, CAVs can travel in closed groups on high-speed motorways, reducing air resistance and improving energy efficiency. By communicating with the infrastructure, CAVs can also theoretically travel at an optimal speed to avoid decelerating for traffic lights, again improving comfort, and reducing energy consumption [12].

V2V also has the potential to enhance mobility by providing new transportation options for individuals who are unable or unwilling to drive themselves. For example, V2V-capable vehicles could provide safe and reliable transportation for individuals who are unable to drive due to age or disability [13].

2.1.3. Early History

The vision of enhanced driving autonomy using V2V communication can be dated back to as early as 1980s, with the concept being explored in notable research projects both in Europe and in North America.

The EUREKA Prometheus Project (Programme for a European Traffic of Highest Efficiency and Unprecedented Safety), from 1987 to 1995, was an early venture in the realm of autonomous vehicles, holding the record as the largest R&D project in its field, and saw the collaboration of numerous universities and car manufacturers. The project was organized into seven sub-projects, including driver assistance by computer systems (PRO-CAR), vehicle-to-vehicle communication (PRO-NET), and vehicle-to-environment communication (PRO-ROAD), artificial intelligence methods and systems (PRO-ART), custom hardware for intelligent processing in vehicles (PRO-CHIP), communication methods and standards (PRO-COM), and developing new traffic scenarios for the assessment and introduction of new systems (PRO-GEN).

A number of demonstrations was carried out in 1994 to showcase the result of the research project, including innovations such as vision enhancement, lane-keeping support, collision avoidance, cooperative driving, autonomous intelligent

cruise control, and several others. A highlight was the vehicles VaMP and VITA-2, developed by Ernst Dickmanns and his team, which autonomously navigated more than 1,000 kilometres on a Paris highway in heavy traffic at speeds up to 130 km/h. Later in 1995, a heavily automated Mercedes S-Class successfully completed a 1,758 km journey from Munich to Copenhagen and back, managing significant portions of the drive without human intervention, reaching speeds over 175 km/h, and performing overtaking manoeuvres in real-world traffic under human supervision [14].

As a result, this project was seen as one of the foundational moments in the development of connected and autonomous vehicles, setting the stage of the rapid developments of various Advanced Driver-Assistance Systems (ADAS) features, including Intelligent Cruise Control (ICC), Lane Departure Warning (LDW), and Automatic Emergency Braking (AEB) systems that have become commonplace on passenger vehicles decades later [15].

While the EUREKA Prometheus Project was a pioneering effort in autonomous vehicle technology, there were a number of limitations and research gaps reflecting the technological and theoretical constraints of its time, such as the lack of satellite-based navigation and mapping services, limited sensor options and computation power, lack of machine learning and AI integrations, and the ambiguous legal and ethical frameworks at the time. These gaps have guided subsequent research, and has led to the more sophisticated systems and robust frameworks in use today.

Similarly in United States, the Automated Highway System (AHS) project was initiated in 1992 as part of the Federal Highway Administration's Intelligent Vehicle-Highway Systems (IVHS) initiative, aiming to address the increasing demand and inefficiency of the U.S. highway system. The AHS focuses on developing a prototype for a future fully automated, intelligent vehicle-highway system, and involves leveraging sensors and communication devices to enhance

driving performance and reduce driver error, leading to increased highway efficiency in order to reduce fuel consumption and accommodate more vehicles on highways, ensure safer driving irrespective of weather conditions, and provide enhanced mobility for all drivers [16].

A crucial demonstration in San Diego in 1997 showcased the technical feasibility of the concept. Vehicles operated on a 7.6-mile stretch of Interstate 15 under automated control, suggesting the potential for an operational system by 2002. Despite these efforts, the program faced significant obstacles. The legislation lacked clear direction for research and did not precisely define what a "fully automated highway system" entailed. This ambiguity, coupled with the project's technical and logistical complexities, led to its premature termination in 1998 [17].

However, despite the promising early results, technical limitations at the time such as the lack of low-cost radios and computing power prevented practical implementations of V2V communication in passenger vehicles, and many of those early experiments implemented their own wireless communication protocols. However, for a successful mass-market deployment of V2X systems, standardisation and interoperability would be of crucial importance. As a result, there have been efforts to standardise the V2V communications protocol, and several V2X standards have been proposed in recent years [12].

2.1.4. Dedicated Short-Range Communications (DSRC)

Dedicated Short-Range Communications (DSRC) is a wireless communication technology specifically designed for automotive use. In short, "DSRC-equipped vehicle broadcasts its basic state information [...] several times per second over a range of a few hundred meters" [12]. The vehicle also receives other vehicles' messages and uses them to determine possible collision risk, and take actions accordingly by warning the driver or provide assistance in controlling the vehicle.

The Federal Communications Commission (FCC) in United States has allocated a 75MHz spectrum at the 5.9GHz wireless frequency band for DSRC applications. IEEE 802.11p Wireless Access for Vehicular Environments (WAVE), modified from the popular IEEE 802.11 standard widely used for Wi-Fi, was used as the low-level communication stack for DSRC [18]. The network stack supports WAVE Short Message Protocol, a special messaging protocol optimised for single-hop transmissions with reduced overhead, as well as traditional protocols such as IPv6, TCP and UDP. The SAE J2735 standard at the top of the stack specifies a set of common message formats to be used for communication between different vehicles. Since the chance of communication might be fleeting in high-speed vehicular environments, several traditional 802.11 functions such as access point beaconing, clock synchronisation and authentication were removed to allow a greater chance of data exchange.

The most important message is the Basic Safety Message, which contains information such as position, speed, heading, steering angle, acceleration, brake status, and vehicle size. It is broadcast frequently and used for safety critical calculations. A consensus has been reached to send all safety critical messages on Channel 175 [19]. For increased security, certificates are used to authenticate each user, and both symmetric and asymmetric cryptographies are used to encrypt messages [12].

An updated next-generation DSRC standard was proposed by IEEE Task Group 802.11bd in 2019. Improvements include doubling the throughput at MAC and enabling longer communications ranges by reducing the noise sensitivity level. Some key changes in the physical layer of the next-gen involve the inclusion of Orthogonal frequency-division multiplexing (OFDM) carrier modulation system of IEEE 802.11ac for better efficiency, adoption of Low-Density Parity Check forward error correction codes to enable higher coding density, utilisation of midambles for improved channel estimation, adoption of higher modulation and coding schemes such as 256-QAM with reduced overhead [20]. In optimal conditions,

DSRC is claimed to support communication data rate up to 27Mbps, while in real world tests, the average data rate was found to be around 6Mbps [21].

A European counterpart of DSRC named ITS-G5 also exists. Published by European Telecommunications Standards Institute (ETSI) with contributions from German vehicle manufacturers, the ETSI ITS-G5 standard is similar to DSRC as it also operates around 5.9GHz using IEEE 802.11p for communication stack, although the messaging stack is different, thus making them incompatible [4]. ITS-G5 contains two message types for vehicle safety: Cooperative Awareness Message (CAM) and Decentralized Environmental Notification Message (DENM). CAM broadcasts real-time vehicle data, including position and sensor information, at a rate adjustable based on vehicle dynamics, constrained by Decentralized Congestion Control (DCC) to reduce channel congestion. It also includes a security system with certificates to ensure trusted message sources. On the other hand, DENM, event-driven and not regularly broadcast, communicates urgent road events like accidents, with a mechanism for message updating and cancellation. In the meantime, the U.S., the Society of Automotive Engineers (SAE) defines the Basic Safety Message (BSM) for vehicle-to-vehicle safety, broadcasting vehicle operational data at a set frequency including general vehicle information and application-specific data, with a linear message rate control system. Both regions also contain additional message classes for specific functions [21].

It was observed that while V2X technology in United States mainly focuses on reducing crashes, European Union takes a slightly different approach as it focuses on driver safety advisories, driver support messages, and commercial applications such as toll and insurance. The EU standard also covers a broader set of applications and has a more market-driven approach [12] [4].

2.1.5. Cellular V2X (C-V2X)

Cellular Vehicle-to-Everything (C-V2X) is a similar technology that leverages standardised 3rd Generation Partnership Project (3GPP) 4G LTE or 5G mobile cellular network to provide communication capabilities in vehicular systems. It is promoted by multiple industry organisations, such as 5G Automotive Association, Qualcomm, and Autotalks [22] [23] [24] [25].

The operation of C-V2X primarily involves two modes of communication. First is the direct communication mode, or Device-to-Device mode, which functions independently of any cellular network infrastructure. The User Equipment (UE) directly communicates with another UE over a wireless link known as PC5 interface, enabling vehicles to communicate directly with each other, pedestrians, and traffic infrastructure [26]. The PC5 interface was originally designed to address safety-critical situations where infrastructure is unavailable. However, its functionality has been expanded in later releases to cater to general-purpose V2X applications [27].

In addition to direct communication between UEs over PC5, C-V2X also allows devices to use the cellular data connection to transfer information over the internet, also known as Vehicle-to-Network communication [28]. This mode relies on traditional cellular networks, facilitating the sending and receiving of information from cloud-based services, other vehicles not within direct range, and any other entities connected to the cellular network [29], with performance goals of sub 100ms latency and 10Hz message frequency [30].

Although 3GPP defines the data transport features for cellular V2X, it does not cover the semantic content of V2X. Instead, it suggests employing ITS-G5 standards such as CAM, DENM, BSM, with C-V2X transport instead [31].

C-V2X is claimed to have several advantages over 802.11p-based V2X solutions. By utilising the cellular data network, C-V2X can provide longer-range communications, better non-line-of-sight performance, as well as provide more

functionality with cloud-based services [32]. C-V2X can also take advantage of existing cellular infrastructures, and is claimed to be scalable, evolvable, and less costly to deploy. C-V2X might also have better performance under congestion conditions [33] [34].

Despite the perceived advantages of C-V2X, this communication standard is still in its early stages of deployment. As a result, many of its performance claims and assessments are based on simulation models instead of real-world tests [22] [33]. While a number of manufacturers have expressed interest in the C-V2X technology and have announced plans of producing compatible equipment, many C-V2X devices are not yet commercially available as of April 2020 [23] [24] [35]. Additionally, some automotive manufacturers have expressed doubt about the reliability of the C-V2X system [36], and some potential disadvantages have been pointed out.

While C-V2X can operate without a cellular network for direct communication, many of its benefits come from network communication, which requires a reliable cellular network. In areas with poor cellular coverage, its performance may be degraded. Also, although 5G networks promise single-millisecond latency, in existing 4G networks the latency may be noticeably higher, which could be problematic for safety-critical applications. C-V2X may also require more expensive hardware to support both Device-to-Device and Device-to-Network modes, and potentially recurring network fees, depending on the usage of cellular communication.

2.1.6. Comparison and Integration of DSRC and C-V2X

A number of publications examined the ongoing debate over the competing standards. They based their analyses on metrics such as Packet Delivery Ratio (PDR), End-to-End Delay (EED), throughput, range, interference tolerance, scalability, security, and cost-effectiveness, as well as the performance in different traffic environments such as urban areas, highways, tunnels, and

intersections. In [1], it was discovered that both technologies demonstrate unique strengths and weaknesses depending on the specific application scenarios.

For example, DSRC presented similar-to-higher PDR than LTE-based C-V2X in low-to-medium vehicle densities. It also demonstrated good scalability, as it can support up to thousands of vehicles in a single cell via ad-hoc networking. However, DSRC suffers from interference issues in high-density scenarios due to its use of the 5.9 GHz band which is very close to the frequency used by certain Wi-Fi network communication bands.

On the other hand, LTE-V2X showed better range than DSRC due to its use of licensed cellular bands. It also supports higher data rates of up to 1 Gbps and has better interference tolerance due to its use of advanced interference mitigation techniques such as beamforming and power control. However, LTE-V2X suffers from higher latency than DSRC due to its reliance on cellular network protocols such as IP-based communication. LTE-V2X also has limited scalability compared to DSRC, as it can only support up to hundreds of vehicles in a single cell. Most of all, cellular infrastructure is required for C-V2X operation, which may not always be available in rural areas.

Cost-effectiveness is another important factor when comparing these two technologies. DSRC requires dedicated roadside infrastructure for operation, which can be costly to deploy. C-V2X can leverage existing cellular networks for operation, which can simplify and reduce deployment costs. However, C-V2X also requires ongoing subscription to a cellular network provider, which can add ongoing operational costs, especially if the communication involves frequent high data rate exchange.

In terms of security, both technologies provide similar levels of protection against eavesdropping and message tampering through the use of encryption and digital

signatures. However, DSRC has been criticised for its lack of privacy protection since it uses static pseudonyms that can be easily tracked by third parties.

Due to the unique nature of both technologies, researchers have looked into combining those two types of V2V communication in a hybrid approach to overcome the individual limitations, and achieve more consistent connectivity, lower latency, higher throughput, and improved reliability and redundancy. DSRC is well-suited for short-range, high-priority safety messages, while C-V2X can provide longer-range communications for infotainment and other non-safety-critical applications. Furthermore, the use of both technologies can help mitigate potential interference issues that may arise from the coexistence of multiple wireless communication systems in the same frequency band.

In [37], the author proposed a novel architecture and protocol stack for Radio Access Technology (RAT) selection and Vertical Handover (VHO) by taking into account various network parameters such as signal strength, interference, and traffic load. A dynamic channel model (DCM) was constructed to predict future network conditions based on historical data. This scheme enables proactive RAT selection and VHO decisions that can improve network performance and user experience. Simulation results show that the proposed approach outperforms existing approaches in terms of network throughput, delay, packet loss rate, and handover success rate.

Another study proposed a Traffic Differentiated Clustering Routing (TDCR) mechanism for vehicular data collection in a hybrid DSRC and C-V2X network [38]. The TDCR mechanism is designed to address the challenge of low data rate and long end-to-end delay in Vehicular Ad-hoc Networks (VANETs), while also reducing cellular bandwidth costs. The proposed TDCR algorithm consists of two main components: a centralised one-hop clustering approach and a data delivery optimisation method.

The clustering approach involves dividing the network into clusters, with each cluster having a Cluster Head (CH) responsible for collecting and aggregating data from the vehicles within the cluster. The CHs are selected based on their proximity to other vehicles and their ability to communicate with Road-Side-Units (RSUs). The CHs then transmit their aggregated data to RSUs either through multi-hop V2V transmissions or through cellular networks.

The data delivery optimisation method is used to determine the optimal delivery method for each CH, based on a trade-off between cellular bandwidth cost and end-to-end delay. The optimisation problem is formulated as a joint optimisation problem of delivery method selection and routing, which is solved using CPLEX software and a proposed heuristic algorithm.

The TDCR algorithm also includes mechanisms for handling different types of traffic, such as delay-sensitive and non-delay-sensitive traffic. Delay-sensitive traffic is given priority in the routing process, while non-delay-sensitive traffic is routed through less congested paths.

Simulation results show that the TDCR algorithm provides an efficient mechanism for collecting and delivering vehicular data in a hybrid SDN-enabled DSRC and C-V2X vehicular network, while ensuring Quality of Service (QoS) for delay-sensitive services and reducing cellular bandwidth costs.

2.1.7. On-Board Unit

An essential component of V2V communication systems is the On-board Unit (OBU). OBUs are specialised devices installed within vehicles that play a central role in V2V communication by receiving, processing, and transmitting information such as position, speed, direction, and other sensor data to nearby vehicles, enhancing the situational awareness of the driver or the vehicle control system [39]. The OBU can also facilitate communication with infrastructure (Vehicle-to-Infrastructure, V2I) and pedestrians (Vehicle-to-Pedestrian, V2P), although those applications are beyond the scope of this project.

OBUs are designed to support one or more V2V communication protocols, such as DSRC, C-V2X, or even satellite-based messaging. OBUs are equipped with embedded microprocessors that can process and fuse data from multiple sources. This data fusion enables vehicles to have a comprehensive understanding of their surroundings, thereby supporting advanced driver assistance systems (ADAS) and autonomous driving capabilities [39].

To ensure secure and trustworthy V2V communication, OBUs may employ cryptographic mechanisms for authentication, encryption, and non-repudiation. Moreover, to protect sensitive user information while still allowing vehicles to share necessary data, OBUs may implement privacy-preserving techniques [35].

OBUs can also integrate with Global Navigation Satellite System (GNSS) receivers, such as GPS, GLONASS, and Galileo, to provide accurate positioning information. This information is essential for cooperative localisation, lane-level navigation, and other location-based services.

Some OBUs can also communicate relevant information and warnings to drivers or passengers through various Human-Machine Interface (HMI) elements, including auditory, visual, and haptic feedback [40]. This interface ensures that users can understand and react to time-critical safety warnings promptly.

As intelligent transportation systems continue to evolve, OBUs are expected to become increasingly sophisticated, incorporating advancements in fields such as machine learning, sensor technology, and cyber-physical systems, thereby enhancing the safety, efficiency, and sustainability of road transport.

2.1.8. Deployment Status

With the established V2V standards, a number of real-world trials were planned and conducted in Europe and the United States. One such example was “Sichere Intelligente Mobilität – Testfeld Deutschland” (simTD), the first large-scale V2X communication field test in Europe, from 2008 to 2013. Unique aspects of this

test involve the large number of V2X equipped vehicles, as well as the fact that they were driven by members of the public who were unaware of the technological details, in order to simulate the situation of real-life deployment. Different traffic environments were considered, including high-speed motorway sections, rural roads, and inner-city streets [10]. The SimTD project provided successful testing and validation of V2X systems, as well as gaining insight into user acceptance, feasibility of large-scale implementation, guidance for future policy and regulatory frameworks.

More recently, with the advance of technology and push from governmental bodies, a number of commercial products for V2X communication have been made available in the form of ready-to-use devices [41] as well as bare transceiver modules [25] [42]. Additionally, a number of vehicle manufacturers have already introduced new car models with V2X capability, such as the 2016 Mercedes-Benz E-Class, 2017 Cadillac CTS, and 8th generation Volkswagen Golf [43] [44] [8]. The United States Department of Transportation has also proposed to mandate DSRC capability on all new cars and light trucks [6].

However, despite the recent developments, the rollout of V2X technology has not been as brisk as many hoped. As of 2019 only three car manufacturers have V2X capable models in the United States and Europe [17] [18] [19], and in November 2017 it was reported that the push for mandating DSRC capability in United States has been delayed by the new administration [45]. In early 2020, Federal Communications Commission (FCC) proposed to reallocate the lower 45MHz portion of the 5.9GHz band for unlicensed operations such as Wi-Fi [5]. In response, lawmakers expressed “concern” and “alarm” over the proposal, commenting that the changes could undercut the potential to prevent a claimed 37,000 traffic fatalities each year by “impeding the development and deployment of safety-critical technologies” [46]. In April 2020, FCC unanimously voted to open up the 6GHz band for unlicensed Wi-Fi use, greenlighting the next generation of Wi-Fi specification of IEEE 802.11ax with improvements in data

transfer speed and throughput [47] [24]. Some have expressed concern that this might erode the effectiveness of the neighbouring 5.9GHz band reserved for V2X communications [48].

Despite the setbacks, several car manufacturers have started to self-mandate the V2V capability in new vehicle models. Toyota has been selling DSRC-equipped vehicle in Japan since 2015, and is planning a large deployment in United States in the early 2020s [26]. Volkswagen also has a similar plan, aiming to start fitting its new vehicles with V2V technology in 2019 [8] [49].

In the meantime, the upcoming C-V2X technology is starting to gain traction among automobile manufacturers, with Ford announcing a “definitive timeline for introduction of C-V2X” starting in 2022 with the backing of additional technology partners including Qualcomm and Intel [50] [51]. BMW is also reported to be pushing the adoption of C-V2X in Europe with telecommunication partners [52].

2.1.9. Potential Issues

Despite the promising safety and commercial outlook of V2X applications, a number of issues need to be addressed to ensure a successful deployment in real-world situations.

One such issue is the wireless spectrum of the communication band. Currently both 802.11p-based and cellular-based V2X system use the 5.9 GHz frequency band. With the potential large-scale deployment, in addition to the continued expansion of Wi-Fi networks that operates at a similar frequency, there is a concern of wireless channel congestion in dense urban vehicular environments, which could adversely affect the safety intent of V2X communications [12].

Privacy is also a major concern of such technology. Since all V2X-equipped vehicles broadcast their dynamic information wirelessly within a large range, there are obvious risks of individuals or government entities unlawfully tracking

or logging such sensitive and private information. Additionally, it might be possible for malicious entities to interfere with the normal operation of vehicles by broadcasting spoofed safety messages. Therefore, cybersecurity measures should be of great importance to prevent abuse of V2X systems, such as unauthorized data collection, eavesdropping on V2V communications, and manipulating traffic data. [9].

Liability is another concern expressed by automotive manufacturers. Some argue that V2X might burden them with more legal risks since the safety system depends on the information from other vehicles outside their control and of unknown quality. On the contrary, it can be argued that V2X should be viewed more as safety warning system instead of autonomous driving aid, and its liability should be similar to existing safety warning systems today [1].

Since the V2X can only benefit cars equipped with such systems, there are reservations that its effectiveness might be limited at the initial stage of deployment, where most vehicles on the road do not yet have V2X capability. Some people have also expressed concerns that DSRC technology might become outdated before the completion of mass deployment [53].

Another area of concern is the fragmentation of standards. Two competing V2X communication methods exist, 802.11p-based (DSRC, ITS-G5) and cellular-based (C-V2X). DSRC enjoys a more established presence with existing products and infrastructures already in place in certain parts of the world, as well as support from numerous automotive manufacturers. By contrast, C-V2X claims a number of advantages over 802.11p-based communication, and is favoured in certain countries such as China, with increasing interest in the United States [12] [51] [6]. However, some argue that the cellular infrastructure is not yet sufficiently mature in United States for mass deployment of C-V2X [8]. With manufacturers and researchers taking side in the competing standards, there are concerns that valuable development time and resources might be wasted over the format war

[54]. Attempts and researches have also been conducted over integrating the two methods of V2V communication together, although the experiments are still mostly inside simulations.

2.1.10. Summary

V2V communication is poised to revolutionise the automotive industry by significantly enhancing safety and efficiency. By enabling vehicles to communicate with each other and their surroundings, V2V systems can react more rapidly to potential hazards, even in situations where those threats are beyond the line of sight of human drivers and conventional vision-based sensors.

Over the years, there have been numerous research efforts and collaborations aimed at realising the potential of V2V communication. However, it was only in recent years that large-scale testing and the development of proper communication standards have accelerated the adoption of V2V technologies. The introduction of Dedicated Short-Range Communications (DSRC), Intelligent Transportation Systems-G5 (ITS-G5), and Cellular Vehicle-to-Everything (C-V2X) has led to the emergence of V2V-equipped vehicles, as well as various aftermarket products. Moreover, an increasing number of automotive manufacturers are planning to incorporate V2V technology into their upcoming models.

Historically, V2V communication has been primarily focused on safety applications, such as collision avoidance, intersection safety, and vulnerable road user detection. However, the potential benefits of V2V extend far beyond safety, as the technology can also improve traffic flow, reduce fuel consumption, and contribute to more efficient use of existing transportation infrastructure. Furthermore, V2V communication is expected to play a crucial role in the adoption of CAVs, as it can significantly enhance their perception and decision-making capabilities.

Despite the promise of V2V communication, the deployment of this technology is still in its nascent stages, and several challenges need to be addressed. One of the concerns is wireless channel congestion, as an increasing number of connected vehicles and devices compete for limited bandwidth. This issue necessitates the development of efficient communication protocols and spectrum management strategies to ensure reliable and low-latency V2V communication.

Privacy and liability concerns are another challenge facing V2V deployment. As vehicles continuously transmit and receive data about their location, speed, and other parameters, ensuring the privacy of this information is of utmost importance. Moreover, the legal and regulatory framework surrounding V2V communication needs to be developed to address questions of liability in the event of accidents or system failures.

Another challenge involves the uncertain outlook of competing communication standards, such as DSRC/ITS-G5 and C-V2X. The coexistence of multiple standards may impede the interoperability of V2V systems, and limit the technology's adoption rate. This issue highlights the need for ongoing collaboration between industry stakeholders, standardisation bodies, and policymakers to establish a unified V2V communication framework.

Lastly, it is important to note that the full safety benefits of V2V communication may only become apparent after mass deployment. As such, a notable period of time may be needed before the technology's real-life impact can be accurately evaluated. In the meantime, continued research, development, and large-scale pilot projects are necessary to refine V2V communication systems, address the challenges mentioned above, and pave the way for the widespread adoption of this transformative technology.

2.2. Machine Learning Applications

Machine learning is a subfield of computer science that focuses on the study and construction of algorithms that are able to learn from data, and make predictions or decisions as a result, all without being explicitly programmed [55]. A multitude of machine learning algorithms have been developed over the years, each with strengths and weaknesses for particular use cases. The algorithms of interest in this study are the ones that perform classification, where categorical class labels of new instances are predicted based on past observations [56].

A selection of popular classification algorithms will be considered for this study, and a brief introduction and their strengths and weaknesses are discussed in the following sections.

2.2.1. ZeroR and OneR

ZeroR is one of the simplest classification methods, as it labels all unknown input data as the most common class in the training dataset. While it is apparent that this method is oversimplified and will likely have poor accuracy, it serves as a useful baseline to compare the performance of more complex algorithms.

Building on the ZeroR algorithm, OneR makes use of only a single parameter in the training dataset to make predictions. The algorithm examines each feature in the dataset independently and constructs rules to predict the target variable. It then evaluates the effectiveness of these rules based on their error rate. The rule with the lowest error rate is selected as the model's rule. While it still runs the risk of oversimplification in complex datasets, OneR is often effective in a surprising number of real-life situations due to its simplicity, interpretability, and fast training speed [57]. As such, it is often used during the exploratory phase of data analysis or as a benchmark for other models.

2.2.2. k-Nearest Neighbour

The k-Nearest Neighbour (k-NN) is an instance-based learning algorithm that classifies an unknown data instance based on the majority vote of its 'k' nearest

neighbours in the feature space [58]. The primary parameter 'k' determines the number of neighbouring data points to consider during classification.

An intuitive example of k-NN is shown in Figure 1, where two classes, star and triangle, are present in the existing dataset, and the model is tasked with classifying an unknown point at the yellow rectangle.

The k-NN algorithm calculates the distance between the unknown point to each existing data point, finds its k-nearest neighbours, and classifies the unknown point as the class with the most frequent occurrence.

In the example, with $k = 1$, the nearest neighbour is of class star, therefore the unknown point will be classified as star. With $k = 3$, its three nearest neighbours comprise two triangles and one star, leading to the classification of the unknown point as a triangle. Similarly, the classification result is star when $k = 6$.

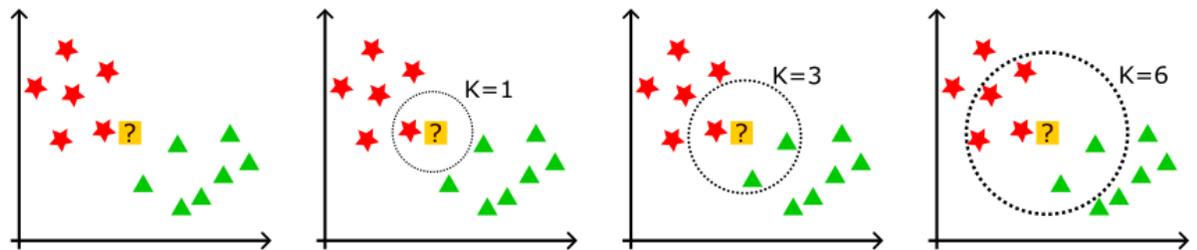


Figure 1: Example of k-NN classification

It can be observed that the classification result may change depending on the value of k . The optimal choice depends on the specifics of the dataset, although larger 'k' value tends to yield smoother classification boundaries, reducing noise and potential overfitting [59].

k-NN is a straightforward algorithm that is intuitive to understand and implement. It is particularly effective given sufficiently large training data. The algorithm does not make strong assumptions about the distribution of the data. Instead, it operates under the premise that similar instances are close to each other in the feature space, even when the data patterns are non-linear [56].

However, k-NN also has several limitations. Determining the 'k' value often requires experimental tuning. Being a lazy learning algorithm, k-NN does not perform generalisation or abstraction during the training phase. This makes k-NN computationally inexpensive to train but can be resource-intensive during prediction, as the entire training dataset is stored in the model, which can become problematic for large datasets. In high-dimensional datasets, k-NN can also be computationally expensive due to the increased distance calculations. Furthermore, an unbalanced training set might bias the output towards the more popular class. [59]. All in all, while k-NN is a simple and effective classification algorithm, it comes with potential pitfalls that require consideration.

2.2.3. Decision Tree

Decision Tree algorithm has been described as “a classifier expressed as a recursive partition of the instance space” [60]. It is a popular model used in many domains for knowledge discovery and pattern recognition.

A decision tree can be viewed as a directional and acyclic graph, featuring a finite and non-zero number of nodes and edges. Exactly one node is referred to as root node and has no incoming edges. A node with outgoing edges is referred to as internal node, while others are called leaf nodes [61]. Root and internal nodes represent a test on a certain attribute. The connection edges represent the possible outcomes of the test, and leaf nodes commonly hold the classification information.

A decision tree can be constructed by splitting the learning data into subsets while achieving the maximum information gain. This process is repeated recursively until all instances at a node belong to the same class [62]. A sample dataset for a decision tree is shown in Table 1, and the constructed tree can be seen in Figure 2.

In order to predict the class of an unknown instance, one can start from the root node and travel along the edges according to the results of the test over

attributes at each node. When a leaf node is reached, the information at that leaf node corresponds to the prediction outcome [63].

Table 1: Sample decision tree dataset

Colour	Height (m)	Class
Grey	3	Elephant
Yellow	5	Giraffe
Brown	1	Monkey
Yellow	1	Tiger

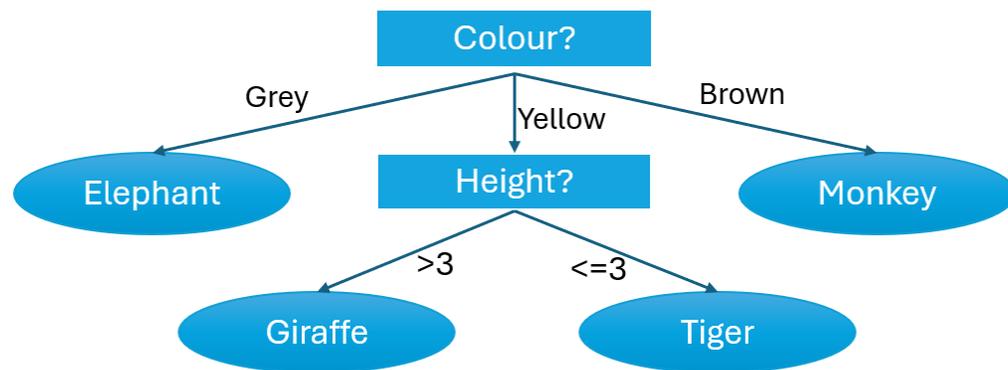


Figure 2: A sample decision tree

Decision Tree algorithm is a white box model, where the construction of the tree and the reason for splitting at each node is directly observable. This also makes the model simple to understand and interpret even for non-experts, where one only needs to start from the root node, follow the directions and arrive at a prediction on a leaf node. Decision tree can handle both numerical and categorical data, performs well with large datasets, and requires little data preprocessing [64].

Despite the numerous advantages, a decision tree can also be very rigid, where a small change in the training data might lead to a large change in the tree structure and the quality of final predictions. There is also a possibility for Decision Tree algorithm to create an overly complex tree that does not generalise

well when presented with unknown data, although mechanisms such as pruning can be used to avoid this problem [64].

2.2.4. Random Forest

Building upon the Decision Tree algorithm, Random Forest provides classification by constructing a large number of different decision trees and letting them vote for the most popular classification. By merging the result from multiple decision trees, random forest tends to offer a more accurate and stable prediction [65].

Random Forest is created by bootstrapping data samples from the original dataset. Various samples from the original dataset are selected with replacement, and the new subset of data is used to create a decision tree. This process is repeated several times, resulting in an ensemble of decision trees, each created with knowledge of a random set of training data [66] [67].

Another distinct feature of the Random Forest algorithm is evident during the construction of the individual decision trees. Instead of searching for the most important feature while splitting a node, the algorithm randomly selects a subset of features and find the best split point based on those. It is believed that a random forest performs best when its internal trees have low error rates, and are uncorrelated with each other [65].

By introducing randomness during the construction of the model, the Random Forest algorithm is reported to be more resilient to overfitting [68]. And the increased diversity of the internal decision trees often leads to a more robust predictive performance.

This algorithm is also relatively simple to use, with few hyperparameters that is straightforward to understand. This model is fast to train compared to more complex algorithms, generally has good predictive performance, and can handle different feature types [46]. Although the large number of trees might slow down the algorithm and render it ineffective for real-time predictions.

2.3. Machine Learning in Intelligent Transportation Systems (ITS)

A number of existing literatures in the field of ITS utilise Machine Learning in their research efforts. In [69], an in-depth exploration of the application of Artificial Intelligence and Machine Learning in ITS was carried out, particularly focusing on their role in the Internet of Vehicles and V2X systems, where they can optimise traditional data-driven approaches, enhance edge computing and caching, facilitate efficient multimedia communication and location-based services, and implement robust security mechanisms for routing protection against various threats and attacks, all while effectively addressing challenges related to traffic congestion, dynamic communication topology, resource allocation management, and vehicular communication in the evolving field of wireless communication technology.

In [70], the article provides an overview of the scientific and technological advances that have the potential to shape future 6G V2X communications. The author envisions that machine learning will play an instrumental role in advanced vehicular communication and networking by enabling more effective resource provisioning and improved network operation, handling situations where traditional communication systems might fail, and aiding in decision-making processes for autonomous driving using multiple data streams from various sensors. The article emphasizes the physical layer, radio resource allocation, and system security, and discusses the implementation of promising technologies such as federated learning.

In [71], Darlan et al. explored the use of machine learning to predict Quality of Service (QoS) in cellular vehicular-to-everything (C-V2X) communication. The article discusses various ML techniques, particularly supervised learning, along with the autoregressive integrated moving average filter, for predicting packet delivery within specific latency windows. Using simulated dataset, the prediction of QoS levels two seconds ahead of time was achieved with 85% reliability. The

work also addresses the challenges of handling class imbalance in ML models for more accurate predictions in C-V2X scenarios.

Article [72] presents a novel approach for reliable routing in vehicular ad-hoc networks (VANETs). It addresses challenges in VANETs such as dynamic mobility and limited connectivity by leveraging decision tree-based routing, focusing on predicting the most reliable routing paths. Machine learning techniques including decision tree and random forest, are evaluated, with decision tree found to be most efficient in terms of accuracy and time consumption. The proposed model significantly improves packet delivery ratios and connectivity, achieving up to 16% PDR with a 99% accuracy rate, outperforming existing solutions and demonstrating the effectiveness in high mobility scenarios.

A more hands-on application of ML in ITS is explored in [73], where the article acknowledges the increasing importance and potential benefits of CAVs, but also highlights the risks of cyber security threats. The article focuses on devising a robust framework for improving CAV cyber security through machine learning-based anomaly detection, with the objectives of defining and assessing potential attacks on CAVs, collecting new CAV cyber security datasets, developing and evaluating machine learning models for CAVs, and improving these models' performance via feature selection methods.

The article examines a wide set of potential attacks within the CAV Cyber Security (CAVCS) framework, assessing their severity based on target assets, risks, and consequences. Notably, it identifies Denial of Service (DoS) and Fuzzy attacks as the most severe cyber threats.

The study gathers four datasets of CAVCS for simulating and assessing potential attacks, covering different attack scenarios in both simulated and real-world environments. Machine learning models, specifically Decision Tree and Naive Bayes classifiers, are applied to each of the four datasets. These models' performance is evaluated based on accuracy, false-positive rate, and runtime.

While Decision Tree outperforms Naive Bayes, the results show a need for enhanced performance improvements, especially in terms of runtime.

Feature selection methods, including Info Gain, Gain Ratio, Correlation-based Feature Selection, and Pearson method, are employed to improve the machine learning models' performance. These methods reduce runtime while preserving accuracy. Although their impact is more pronounced on datasets with a large number of features, the Decision Tree model consistently offers the best performance. This study contributes to CAVCS by providing an effective method for assessing and detecting CAV cyber-attacks and setting a baseline for future CAVCS research.

In summary, the integration of Machine Learning in ITS across various research studies highlights its pivotal role in transportation technologies. ML's data processing capabilities can enable significant advancements in vehicular communications, network security, and predictive analytics in transportation.

Chapter 3: V2V OBU Design

3.1. Overview

The work undertaken in this thesis contributed to the CoDRIVE (Cooperative Development of a Roadmap for Initial V2X Implementation in Europe) project, a European Space Agency (ESA) funded demonstration project focusing on infrastructure and smart cities [74], with the goals of:

- Assessing smart infrastructure and network requirements for future Connected Autonomous Vehicles.
- Implementing seamless V2X communication to support high frequency and low latency data sharing.
- Evaluating EU radio navigation and national digital data services.
- Developing a reliable, robust, and resilient absolute positioning system for intelligent mobility services, enabled by localisation and V2X communications based on GNSS and communication capabilities.
- Building a cloud-based platform for the initiatives of transitioning towards the Mobility-as-a-Service (MaaS) in the near future.

CoDRIVE is a collaborative project between UbiPOS UK Ltd, University of Nottingham Geospatial Institute (NGI), Low Level Earth Observation Ltd (LLEO), and Chang'an Motors UK Ltd.

Commencing in late 2019, the CoDRIVE project saw steady and significant progress from all participating parties over two years and concluded successfully in early 2022 with outstanding ratings. The project was demonstrated at the European Centre for Space Applications and Telecommunications (ECSAT) in Oxfordshire, United Kingdom to the ESA Director General Mr Josef Aschbacher in July 2022.

3.2. Design Goals and Requirements

As a major part of the CoDRIVE project is dedicated to examining various aspects of V2V communication in real-world scenarios, with a particular focus on its integration into smart infrastructure and MaaS applications, a dependable hardware OBU platform would be a crucial element in carrying out the investigations of this project. Several requirements, which the OBU should satisfy, were identified at the early stage of its development:

- Supports current V2V standards such as DSRC and ITS-G5.
- High-accuracy GNSS receiver for positioning and localisation applications.
- Cellular connectivity for data link and potential C-V2X investigations.
- Additional wireless interfaces such as RFID, Wi-Fi, and Bluetooth for novel positioning augmentation experiments in GNSS-denied environments.
- CAN bus interface for communication with vehicle internal networks.
- Reliable and dependable performance.
- Ease of use with clear documentations and streamlined development workflow.

A number of commercially available OBU options were examined [41] [75] [76] [77]. However, most of them were newly announced at the time of the investigation, with low availability and long lead time as well as significant cost, often reaching thousands of US dollars per unit [78]. They also tended to have closed and proprietary development stacks, limited documentation, and did not have the additional sensors and interfaces needed for this project [79]. Due to the availability, cost, and capability constraints of the existing commercial options, the decision was made to develop a custom experimental V2V OBU hardware platform for the CoDRIVE project, which offers numerous advantages that is beneficial to the execution of this project.

Among the most apparent benefits is the level of flexibility and customisation. Custom designing an OBU enables the team to build a solution tailored precisely

to the project's requirements, as well as integrate seamlessly with the specific use-cases and investigation topics of the CoDRIVE project.

This approach also provides valuable insights into the design and implementation process of V2V communication systems. It offers a unique opportunity to gain first-hand experience in constructing and troubleshooting these systems. This not only enhances the team's technical understanding but also empowers them to devise innovative solutions to potential challenges.

Moreover, a custom developed OBU contributes to the creation of unique intellectual properties for this project. This enriches the project's value proposition, allows the team to maintain control over the technology's evolution, and can serve as a basis for further development, upgrades, and commercialisation. Custom-built hardware can also facilitate future vertical integration, as the team has the full knowledge and authority to modify or expand the system as necessary.

Lastly, cost-effectiveness is another important advantage of the custom development approach. As every element is developed and assembled internally, there is no need to account for the markups typically associated with commercial products.

In conclusion, the OBU plays an important role in achieving the goals of the CoDRIVE project, and the development of a custom OBU offers significant advantages in terms of customisation, educational value, intellectual property generation, future development possibilities, and cost management.

3.3. Component Selection

With the decision to develop a custom OBU, this section describes the major components used in this project, including the background information, the roles they play in the product, design requirements, available options, and the reasoning for the final choice.

3.3.1. Processor

The main processor plays a crucial role in determining the rest of the system design, as each processor architecture exhibits their own software and hardware strengths and limits. Many processor options are available for embedded applications, with two major architectures considered for this project.

The first category is x86-based development boards, such as the UDOO x86, Rock Pi X, or Odroid H2. The advantage of x86 processors includes their relative ease of setup and operation, due to the close architectural resemblance to conventional desktop computers. These processors can also deliver commendable performance and offer compatibility with popular operating systems such as Windows 10 or Ubuntu. In addition, manufacturers of various sensor modules often provide support and drivers for x86 systems, simplifying integration into the existing setup and facilitates the validation of their functionality, thereby making them an appealing option for this project. Some example x86 single-board computers are shown in Figure 3 and Figure 4.

On the other hand, x86 architecture does present certain drawbacks for embedded computing applications. Notably, x86 development boards tend to be more expensive and larger in size compared to other alternatives, which could pose challenges in terms of hardware integration and sourcing, especially for high-volume production. Moreover, these boards are often associated with increased power consumption and heat output, which can present additional challenges in terms of thermal management. Furthermore, depending on the manufacturer, these boards may not have comprehensive documentation or strong community support, which can impact troubleshooting and development efforts. There are also concerns for the longevity of supply, as the assurance of continued availability of low-power x86 processors and development boards might not be as strong as for some other architectures.



Figure 3: UDOO x86 Computer



Figure 4: Odroid-H2 Computer

ARM-based development boards represent another major choice in the field of embedded systems. Examples encompass a broad range from high-performance devices such as Nvidia Jetson to more affordable options such as Raspberry Pi and its numerous clones. It is even possible to design a custom processor board from scratch, using the bare chip coupled with a minimal amount of supporting circuitry. ARM-based platforms predominantly run on variations of the Linux operating system, and the diversity of Linux distributions offer a multitude of customisation options catered to certain design priorities, such as computational performance, security, real-time responsiveness, energy efficiency, and many others.

ARM-based architecture plays a prevalent role in the embedded computing market today, and underpins a wide range of devices from virtually all smartphones and tablets [80], to deeply embedded applications found in appliances, network routers, and microcontroller chips.

The ubiquity of ARM-based processors offers several compelling advantages. The highly competitive landscape and the open-source ARM GNU toolchain without licensing cost makes this architecture a very cost-effective proposition, and the resulting popularity has fostered a vibrant community producing a wealth of detailed documentation and resources that can significantly simplify the development process. The competitive landscape also offers a variety of vendors

with a wide range of features and price points, allowing a robust selection of options when it comes to integrating custom designs into the hardware.

However, the ARM architecture does come with its challenges. There can be a somewhat steep learning curve, particularly when it comes to setting up the development environment, especially when working with barebone Linux. Nevertheless, the flexibility and cost-effectiveness of the ARM-based architecture can still make it worthwhile for many projects.

The Nvidia Jetson TX2 module was initially considered for this project due to its impressive specifications. This embedded system-on-module is equipped with a quad-core ARM Cortex-A57 processor, 8GB 128-bit LPDDR4 memory, and an integrated 256-core Pascal GPU [81]. These components lend themselves to computationally intensive tasks such as real-time video processing and machine learning applications. While the system specification is impressive, a few drawbacks were identified. This device was announced only a few months prior to the start of this project, as a result, the availability of documentation and online resources at the time was relatively limited, potentially complicating the development process. The cost of the development board was also fairly high, at over £300 per module. As the primary function of the OBU involves mostly data processing, logging, and telemetry, the advanced specifications might be excessive for this project. As a result, the decision was made to explore other potentially more suitable and cost-effective alternatives.

The Raspberry Pi (RPi), seen in Figure 5, is another highly popular single-board computer (SBC) often used in embedded applications. Although its specifications are relatively modest compared to the alternatives discussed previously, its popularity and widespread use since its introduction in 2012 give it some distinct advantages. It is available from a wide range of distributors, and a large online community has been built up with detailed instructions available regarding hardware integration, software troubleshooting, as well as interfacing with

external peripherals. The popularity also results in a comprehensive collection of open-source software libraries and pre-compiled package repositories. The RPi's popularity coupled with its robust support system contributes significantly to its appeal, making it an effective and user-friendly starting point for embedded applications, and it has been utilised in a number of research projects in the field of V2V communication. Examples include implementing a secure wireless message propagation protocol for Internet of Vehicles (IoVs) [82], acting as the processor of a low-cost OBU unit using off-the-shelf components along with Wi-Fi, ZigBee, and nRF24L01 wireless communication [83], as well as for implementing a lightweight mutual authentication protocol for IoV communication [84].

One potential drawback of the Raspberry Pi board is its form factor. Originally conceived as a low-cost single-board computer, its physical dimensions are relatively large, comparable to the size of a credit card. In addition, the board provides limited General-Purpose Input/Output (GPIO) connections for additional sensors and peripherals, which could pose integration challenges when pairing with other components in the OBU.

Fortunately, Raspberry Pi Foundation offers an alternative product in the form of the Raspberry Pi Compute Module, shown in Figure 6. Retaining the same system architecture as its larger counterpart, the Compute Module presents a significantly more compact footprint, comprising of only the processor itself and a minimal amount of supporting circuitry in the form factor of a 200-pin SO-DIMM laptop memory module, while having a lower retail cost compared to the regular Raspberry Pi. This format is explicitly designed to be embedded into custom-designed commercial products, while still preserving the user-friendly development environment characteristic of the standard Raspberry Pi. This makes it an excellent choice for projects requiring a more streamlined design without compromising processing capability, and the official documentation offers detailed instructions on hardware integration [85]. Given these factors, the

Raspberry Pi Compute Module was selected as the processor base for this project due to its balance of size, cost, functionality, and ease of integration.



Figure 5: Raspberry Pi 3 Model B+



Figure 6: Raspberry Pi Compute Module 3+

3.3.2. V2X Transceiver

The V2X transceiver is a critical component of the OBU design, as it is responsible for sending and receiving wireless V2X signals. As the main component of this project, the V2X transceiver needs to meet several requirements such as stable and reliable operation, support for major V2X bands such as ITS-G5 and DSRC, availability during the project timeframe, reasonable pricing, and comprehensive documentation and support. Moreover, it is crucial for the transceiver to provide a high degree of compatibility and interoperability with other V2X-enabled devices, ensuring seamless communication and data exchange in various traffic scenarios.

At the onset of this project, V2X transceiver modules were still not widely available, and research and inquiries were carried out to investigate the available options. The u-blox THEO-P173 was chosen due to the manufacturer's reputation and immediate availability, accompanied by an evaluation kit, documentation, and sample codes [86]. The THEO-P173 module was also used by Cohda Wireless in their OBU and Road-side Unit (RSU) commercial products [87] [88], as well as a major component in a recent study on implementing a Vehicle Approaching Reminder Device for CAVs [89]. The THEO-P173 module supports both ITS-G5 and DSRC standards, thereby covering European and US markets. It comprises a

dual-channel radio operating in the 5.9 GHz band with up to +23 dBm output power, ensuring efficient and robust communication.

The open-air communication range of the THEO-P173 is claimed to exceed 1000 meters, with data rates of up to 54 Mbps. The module interfaces with the host computer through a USB 2.0 or Serial Peripheral Interface (SPI) link, facilitating efficient data transfer and integration [90]. An evaluation kit containing two modules was procured to verify the specifications and evaluate the quality of documentation and ease of development.

3.3.3. GNSS Receiver

A Global Navigation Satellite System (GNSS) receiver is another essential component of the OBU, as it provides geographical positioning information as part of the payload to be transmitted over V2V communication, as well as for trajectory telemetry and logging purposes.

However, a standard GNSS receiver without using any augmentation techniques typically can only achieve meter-level precision [91]. While this might suffice for generic navigation applications, it falls short of the precision requirements of advanced applications associated with CAVs, where much higher accuracy is needed for critical functions like lane-level navigation, collision avoidance, and cooperative manoeuvring. Therefore, a high-accuracy GNSS receiver compatible with augmentation methods is needed for the OBU.

Real-time Kinematic Positioning (RTK) is a technique often employed to significantly improve the level of GNSS positioning accuracy, often attaining centimetre or even millimetre precision, depending on factors such as equipment setup, atmospheric conditions, and receiver movement patterns. RTK positioning utilises a network of fixed reference stations with precisely known coordinates. These reference stations continuously monitor the phase of the GNSS signal's carrier wave as well as the code and pseudorange data. By comparing the measured position with the known position, the reference station can generate

correction data, which is then transmitted to the RTK-enabled GNSS receiver, either via radio or cellular communication, allowing mitigation of a number of sources of inaccuracies such as clock drift, atmospheric delays, and multipath effects [92].

Some GNSS receivers can perform RTK calculations internally, necessitating only an antenna and a correction data stream input. In contrast, other receivers are designed to output raw measurement data and delegate the RTK processing to the host computer.

Opting for external RTK calculation on the host computer was considered to be a more flexible approach in this project, as it enables the fine-tuning and testing of different processing configurations. Additionally, the GNSS receiver itself can be more cost-effective without the requirement for on-board RTK processing.

Numerous GNSS receiver modules are available from various vendors. In this project, the u-blox NEO-M8T was chosen due to its reputable manufacturer, reasonable pricing, comprehensive documentation, extensive online resources, robust software and hardware support, as well as prior experience with the device. As the name implies, the NEO-M8T is primarily designed to deliver high-accuracy timing and frequency reference for high-sensitivity equipment. It supports all major GNSS constellations, including GPS, GLONASS, BeiDou, Galileo, as well as satellite-based augmentation systems (SBAS), and boasts an unaided horizontal position accuracy of 2.5 meters [91].

The NEO-M8T module has been assessed and evaluated in a number of studies, and has been observed to achieve up to 1 meter of horizontal accuracy in single positioning mode [93], and under "Where in Lane" accuracy level (0.5m) for CAV applications [94]. As a result, it has been selected as a low-cost high-accuracy GNSS receiver in many research projects, such as building a cost-effective multi-purpose GNSS platform [95], developing novel algorithms to improve its

performance in urban environments [96], and to provide localisation for a V2V OBU in a real-world trial in Korea [97].

Although the NEO-M8T does not offer built-in RTK processing support, it can output raw GNSS measurement data through various interfaces for utilisation on the host computer. This functionality is particularly useful when employing open-source libraries such as RTKLIB, which enables the experimentation with different configurations and the logging of raw data for post-processing if required.

3.3.4. Cellular Modem

The inclusion of a cellular modem in the OBU design allows internet connectivity in areas with cellular network coverage. This connectivity enables several crucial use cases, and enhances the overall functionality and performance of the OBU.

One key application of the cellular modem is the streaming of GNSS RTK correction data, which can significantly improve positioning accuracy. By leveraging precise GNSS corrections provided by reference stations, the OBU may achieve centimetre-level accuracy, which is vital for advanced driver assistance systems, autonomous driving applications, and other location-based services.

Another important use case is the provision of real-time telemetry updates to a cloud-based database. This capability allows the OBU to transmit critical vehicle data, such as speed, position, and sensor information, to a remote server for analysis, monitoring, and decision-making purposes. Consequently, this feature supports fleet management operations, traffic management systems, and other intelligent transportation applications.

Additionally, the cellular modem enables over-the-air updates and bug fixes for the OBU's system software, ensuring that the device remains up-to-date and secure. This remote update capability is crucial for maintaining system stability, addressing vulnerabilities, and adding new features as required.

For this project, the Quectel EC20 module was selected due to previous experience and its popularity in several similar studies requiring cellular connectivity, being able to maintain stable cellular data connection for real time data streaming [98] [99]. The EC20 offers support for cellular voice and data communication, with up to 100 Mbps downlink and 50 Mbps uplink speeds [100]. The module's comprehensive English documentation ensures ease of integration, while the existing experience with the module allows for the reuse of proven, reliable code, minimising the need for developing new code from scratch. The EC20 module connects to the host computer via a USB interface, simplifying the integration process and offering flexibility in the overall OBU design.

3.3.5. Wi-Fi / Bluetooth

Wi-Fi and Bluetooth are wireless communication technologies that have become essential components in modern electronic devices, providing various advantages in terms of connectivity and data exchange. The integration of both of those technologies offers numerous benefits and enhances the overall functionality of the system.

Wi-Fi, a set of wireless network protocols based on the IEEE 802.11 family of standards, enables local area networking and internet access for compatible devices. The incorporation of Wi-Fi in an OBU design not only allows for seamless software updates and data transmission to and from the internet but also facilitates remote management and diagnostics of the device during development and deployment.

Bluetooth, another short-range wireless technology, operates on the 2.4 GHz frequency band and is primarily used for data exchange between devices in close proximity. Integrating Bluetooth into an OBU design offers several advantages, such as enabling real-time data exchange with other in-vehicle devices (e.g., smartphones, tablets, or head units), supporting wireless peripherals (e.g.,

keyboards, mice, or sensors), and facilitating remote diagnostics and firmware updates.

A miniature USB Wi-Fi/Bluetooth adapter was used to add those capabilities. Benefits of doing so include allowing the efficient utilisation of space on the printed circuit board for the inclusion of other essential components or more compact designs. Also, as these adapters are mass-produced and can be acquired at a relatively low cost, using them is a cost-effective alternative to embedding individual Wi-Fi and Bluetooth chipsets.

Moreover, the modular design of the USB adapter facilitates straightforward upgrades and replacements as new Wi-Fi and Bluetooth standards emerge, or as the performance requirements of the OBU evolve over time. Lastly, the operating system's native support for the adapter eliminates the need for developing specialised drivers for custom hardware, further streamlining the design process and reducing overall development time and effort.

3.3.6. CAN Bus Transceiver

The Controller Area Network (CAN) is a communication bus standard predominantly employed in vehicular environments. This system facilitates seamless communication between in-car microcontrollers and other electronic devices using a message-based protocol. The CAN bus architecture employs twisted pair wiring and differential signalling to enable high-speed data transmission while maintaining robustness and reliability, even in the challenging conditions typical of vehicular environments. Additionally, it reduces the number of conductors in the wiring harness, saving material costs and weight.

A crucial component of the OBU is the CAN bus transceiver, which serves as a gateway to the vehicle's internal control network. This device allows the OBU to access a wide range of information, such as vehicle speed, pedal position, steering input angle, and diagnostic data. The acquired information can be employed for a variety of purposes, including for data logging and subsequent

analysis, telemetry, and broadcasting through Vehicle-to-Vehicle (V2V) communication systems.

Among the numerous CAN bus transceivers available on the market, the MCP2561 chip from Microchip Technology is a popular choice for projects involving CAN bus and embedded processors, such as error code readout for vehicle maintenance [101] and real-time vehicle status monitoring and management [102]. This preference is primarily attributed to the substantial amount of existing software libraries and sample codes available for the MCP2561 chip [103]. These resources enable developers to set up a functioning prototype with relative ease and efficiency, greatly reducing the time and effort typically required for such endeavours, ensuring efficient setup and implementation.

3.3.7. Inertial Measurement Unit

The Inertial Measurement Unit (IMU) is an electronic device designed to measure the specific force, angular rate, and orientation of an object using a combination of sensors such as accelerometers, gyroscopes, and magnetometers. IMUs have wide-ranging applications, from mundane tasks such as providing orientation information in smartphones, digital photo and video stabilisation, to mission critical roles such as acting as control references for aircraft and spacecrafts. In the context of this project, the IMU is expected to complement the GNSS receiver by providing dead-reckoning positioning results when GNSS signals are unavailable, as might be the case in tunnels, dense urban environments, or under electronic interference.

IMUs are available in a diverse array of designs, ranging from purely mechanical devices to solid-state microelectromechanical systems (MEMS) packages. MEMS-based IMUs are generally preferred for their compact size, energy efficiency, and reliability. The market offers IMUs at various price points, with options catering to different size, accuracy, and interface requirements, from less than 1 GBP to

tens thousands of pounds for tactical-grade components. Key performance parameters that define IMUs include Bias (the IMU reading when stationary), Bias repeatability (consistency between measurements), Bias stability (change in Bias over time), Random Walk (noise in the output), and Vibration sensitivity.

Two IMUs are selected for this project, ADIS16460 and ADIS16488, both from Analog Devices. ADIS16460 is a Six Degrees of Freedom MEMS IMU with built-in gyroscope and accelerometer. It features excellent Bias stability and Random Walk figures, and is capable of high-speed sampling from SPI communication bus [104]. The ADIS16460 was evaluated and utilized in a number of studies such as implementing velocity-based optimization-based alignment (VBOBA) algorithm to reduce the heading angle alignment error [105], and the development of a tightly coupled inter-system RTK model, where the inclusion of this IMU brought about notable improvements to the positioning performance both in simulation and real-life environments [106], demonstrating excellent performance for vehicular applications.

ADIS16488 is an even more feature-rich IMU with slightly superior specifications, and ten degrees of freedom with the addition of a magnetometer and air pressure sensor [107], and was used in several Unmanned Aerial Vehicle (UAV) positioning studies [108] [109]. Further improvements in accuracy can also be achieved by compensating for temperature and acceleration effects [110].

Both IMUs use similar connectors and communication interfaces, therefore it was decided to include both to explore their suitability for this application.

3.3.8. Physical Interfaces

Additional physical interfaces are also included on the OBU for development activities, data logging, connecting with external devices, and expanding the overall functionality of the OBU.

Two USB Type-A connectors are included, allowing the attachment of external storage, Wi-Fi and Bluetooth adapters, mice and keyboards, and any compatible devices, allowing straightforward and flexible expansion of the OBU's capabilities.

The OBU can be powered by either a USB Type-C connector or a 5V barrel jack power adapter.

The OBU also features a High-Definition Multimedia Interface (HDMI) connector. This digital interface supports high-resolution video output and can be utilised to connect an external monitor, beneficial during development phases, testing, and for demonstrations.

For reliable and high-speed data transmission, an Ethernet port is included. This provides a wired internet connection, offering greater stability and data throughput compared to wireless alternatives, especially during the development process.

Furthermore, two Secure Digital (SD) card slots are integrated into the OBU. One hosts the SD card containing the Linux operating system from which the OBU boots. The second slot can hold an extra SD card for additional removable data storage.

A micro-SIM card slot is included for users to insert their own SIM card to be used with the on-board cellular module.

Finally, six SMA antenna connectors are available for attaching external antennas. This includes one connector for the GNSS receiver, one for the RFID module, two for the cellular module, and two for the V2V module.

3.4. Printed Circuit Board

Printed Circuit Board (PCB) serves as the foundation of the OBU hardware, hosting the electrical connections of all the aforementioned components. This section gives a detailed look into a few key aspects of PCB design and production, including the working principle, construction and manufacturing, and areas in

need of special attention for the performance requirements of the OBU. This discussion allows a better understanding of the design and routing considerations needed to achieve optimal performance of the hardware.

3.4.1. Overview

A printed circuit board is a flat electronic assembly that uses metal conductors (usually copper) on top of an insulated substrate to create electrical connections between components in a controlled manner [111]. Electronic components are typically installed on a PCB through a process known as soldering. With the circuit board itself providing mechanical support, the assembly process can be dramatically simplified compared to earlier methods of circuit construction such as point-to-point wiring or wire wrapping. This allows for a highly automated production process and large-scale, cost-effective manufacturing of electronic devices today.

3.4.2. PCB Elements

A typical PCB consists of a flat sheet of insulating material, known as the substrate, and a layer of thin copper foil laminated to its surface [111]. A diagram of a simple two-layer PCB is shown in Figure 7.

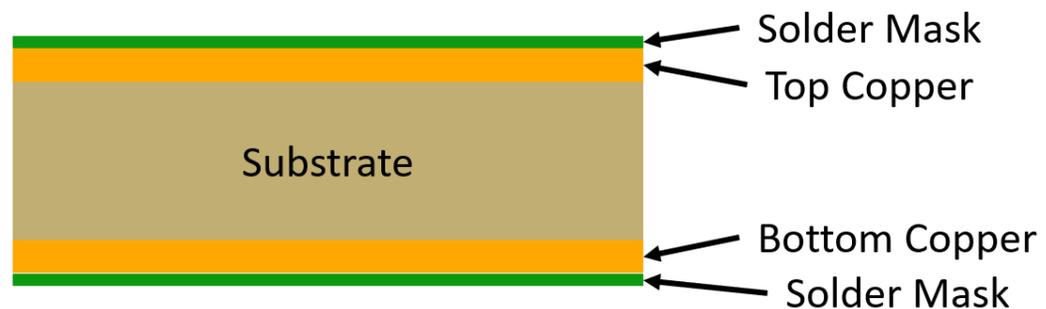


Figure 7: A Typical Two-layer PCB Construction

The copper is chemically etched away into separate conducting lines, known as traces, which function as connecting wires between components. Additional features might be added to the board, such as vias to connect traces between

different layers of copper, pads for component connections, and holes for mechanical mounting.

To protect the copper from corrosion and reduce the chance of a short circuit, a non-conductive coating known as the solder mask is often added to the surface of the PCB. Green is a very popular colour for the solder mask and is often associated with the distinct appearance of circuit boards, but other colours are often used as well, mostly for aesthetic reasons.

Finally, text might be printed on the PCB, showing the component types, value, and descriptive and identifying information. This is known as the silkscreen.

Figure 8 shows the production steps of a simple PCB, the schematics and board design are created using an Electronic Design Automation (EDA) software. The finalised files are sent to the manufacturer, who etch away the copper according to the design. After additional mechanical processing, the circuit board is coated with solder mask, coloured white in this case, to protect from oxidation and short circuit. Text markings are printed via silkscreen, and the PCB is populated with components to form a functional product.

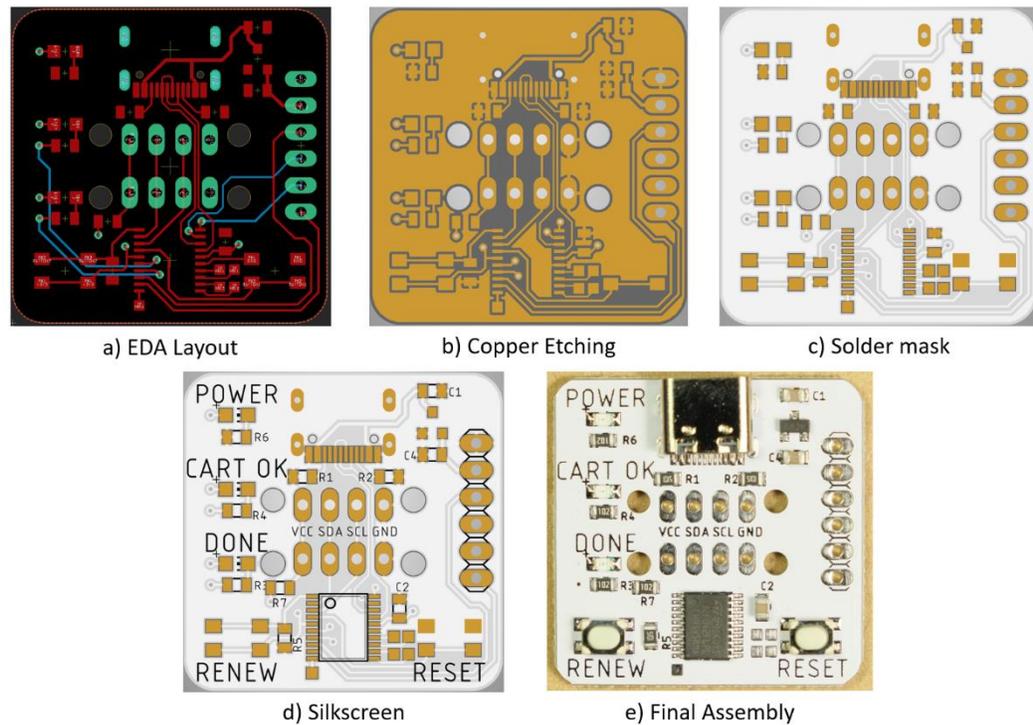


Figure 8: Production Process of a Simple PCB

While the basic principles and construction of a simple PCB may appear straightforward, there are design considerations that are critical in ensuring optimal performance and functionality of a device, especially when dealing with sensitive high-speed or low-noise signals. The following sections will cover a number of those topics that contribute to high-quality PCB design and the overall effectiveness of the device.

3.4.3. Layer Stack-up

PCB layer stack-up refers to the arrangement of the different layers that make up a PCB. A well-designed stack-up can simplify signal routing and impedance control, minimise unwanted electromagnetic radiation, reduce vulnerability of external electromagnetic interference (EMI), and potentially reduce the cost of production [112].

Many low-cost PCBs have two conducting copper layers on each side of the substrate. Typically, the top copper layer is employed for the majority of signal and power trace routing, while the bottom copper layer is dedicated to serving as a ground plane, providing a low-resistance current return path and shielding from external interference. Signal traces that cannot be optimally routed on the top layer can also travel on the bottom layer through vias, but the continuity of the ground plane should be maintained as much as possible to optimise current carrying capacity and maximise EMI shielding effects.

However, for more complicated circuits, two copper layers might be insufficient due to the complexity of trace routing or EMI requirements. In such cases, additional copper layers can be incorporated at the expense of extra cost of fabrication. For the custom OBU in this project, 4-layer PCBs were used.

The four copper conducting layers allow additional functionality of each layer, although the standard practice remains largely the same. Typically, one layer will remain as ground plane, providing EMI shielding and current return path. Another layer can be allocated as a power plane to distribute required voltages to all necessary components. With these needs catered for, the remaining two layers can be exclusively dedicated to signal traces, leading to much more freedom and flexibility in signal routing.

With four copper layers stacked on top of each other, there are 24 permutations to consider, and it is important to choose the most optimal configuration. As with the 2-layer stack-up, signal and ground (or power) planes should be closely coupled in order to minimise EMI emission in high-speed signals, reduce input signal noise, and contribute to effective impedance control. This criterion reduces the viable combinations to mainly two, with the two signal layers on the outside and the power and ground plane on the inside, or having this arrangement in reverse. The first arrangement was chosen due to the advantages of easier routing with fewer vias, and the ease of inspection and repair due to

the exposed traces. Of course, compared to the other option, the exposed outward-facing traces may be more susceptible to EMI with limited shielding and may be prone to physical damage. However, this stack-up remains a very popular choice for 4-layer PCBs, and the trade-offs were deemed appropriate for an experimental prototype device like this. The stack-up is visualised in Figure 9, and additional copper layers can be observed on the inside of the circuit board. Both Core and Prepreg (resin pre-impregnated glass fibre) are insulating substrates, although prepreg is typically significantly thinner to allow easier impedance control, discussed in the next section.

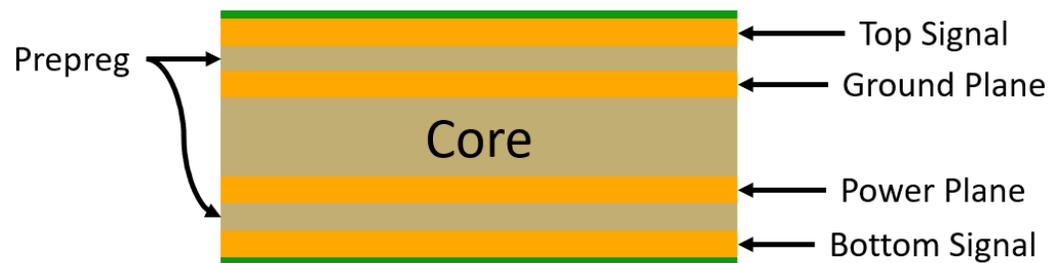


Figure 9: Four-layer stack-up used in this project.

3.4.4. Impedance Control

Another important aspect of PCB design, especially when high-speed signal is involved, is the careful optimisation of trace impedance. In the context of digital circuit design, the impedance is the sum of all the resistance and reactance components of an electrical signal path. Copper, being an excellent conductor, generally offers negligible resistance to direct-current (DC) signals. However, the copper trace on a PCB also possesses a small amount of inductance and capacitance due to the imperfect nature of the material as well as the coupling effect of the adjacent layers and traces. The combined effect of the inductance and capacitance, known as reactance, can appear as additional resistance when a high-speed alternating current (AC) signal is applied [113]. Therefore, it is very important to match the impedance of the signal source to the impedance of the PCB trace, in order to maximise power transfer and minimise signal loss or

distortion. An unmatched impedance can cause a portion of the signal to be reflected back towards the source and cause destructive interference, leading to signal degradation and increased electromagnetic emissions.

Careful PCB design practices are necessary to maintain a controlled impedance environment, and many factors can influence the impedance of a PCB transmission line, such as trace width, trace thickness, substrate thickness, and the dielectric constant of the substrate material. Numerous models have been developed to approximate the impedance value given those factors. Although in recent years, this process has been automated in many EDA software [114], and PCB manufacturers have also started providing in-house online calculators [115], allowing designers to accurately estimate impedance tailored to their specific production processes.

3.4.5. Differential Signals

Many high-speed signals use a technique known as differential signalling, in which data is transmitted over two complementary signals sent on two separate conductors as a differential pair. One conductor carries the regular signal, and the other carries an inverted version of the same signal. An example can be seen in Figure 10. This arrangement allows double the signal voltage swing between the differential pair, and improves noise immunity as the interference affecting both conductors can be cancelled out at the receiver. The equal and opposite current in the two conductors can also significantly reduce EMI, as their magnetic fields are closely coupled with much lower radiation leakage [116]. The advantages allow notably higher signal frequency, and in turn, data throughput, to be achieved.

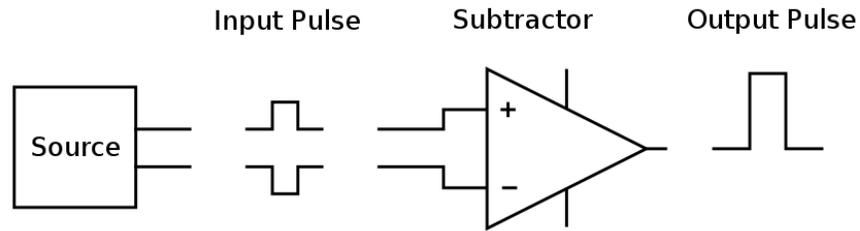


Figure 10: Illustration of Differential Signalling

Of course, there are additional elements to consider for valid and effective routing of differential signal traces on a PCB. The two traces should be routed in parallel at a consistent spacing in close proximity at all times, in order to maintain the required differential impedance. The differential pair should also be kept well away from other differential pairs to prevent crosstalk. Both conductors should be of equal length to ensure their signals reach the receiver inputs at the same time, otherwise issues such as signal glitching or unwanted EMI emission might occur [117].

Discontinuities in the signal path, where trace impedance deviates from the specified value, should be reduced as much as possible to minimise signal reflections and attenuation at those locations. This can be achieved by avoiding sharp bends in signal routing, avoiding using unnecessary vias and in-line components, and maintaining consistent trace width and spacing [113].

Connectors and input/output pins on an integrated circuit (IC) chip are also sources of discontinuities. Although they cannot be entirely avoided, thoughtful routing should still be performed to ensure the signal integrity across the whole signal path.

3.5. All-in-One OBU

With the understanding of important PCB design practices covered in the previous section, the first iteration of OBU was designed on a single PCB containing all the required components using the Autodesk Eagle EDA software. The circuit design is shown in Figure 11. The sections coloured red represent the

copper traces and pads on the top signal layer, while sections coloured blue refer to those on the bottom signal layer. The internal ground and power planes are not shown. The Raspberry Pi compute module is mounted on the centre of the PCB to facilitate connection of other components. The GNSS receiver can be found on top left of the PCB, shared with a CR1210 battery holder, allowing preservation of configuration and satellite orbit information when the system is powered off. The battery backup is also connected to the Real-time Clock (RTC) to maintain system time and date. The V2X communication module and cellular modem can be found on the left-middle and left-bottom of the PCB, respectively.

The antennas from the GNSS receiver, V2X module, and cellular module are located on the left edge of the device using SMA male connectors, with traces matched to 50 Ohm impedance as specified in product documentation.

The lower edge of the PCB hosts the micro-SIM card slot for the cellular module, the two SD card slots for operating system and additional storage, and the DB9 CAN bus connector for interfacing with vehicle internal networks. The right edge contains two USB-A connectors, Ethernet port, and a micro-USB port for powering the OBU. Finally, the HDMI connector can be found on the top edge for digital video output.

Due to the amount of USB devices and ports required, a LAN9514 USB Hub and Ethernet Controller chip was used to extend the single USB port on the Raspberry Pi Compute Module to four downstream USB ports. Two are used internally for the V2X and Cellular module, and two are made available for external devices. This chip also supports Ethernet and makes it available as another USB device to the system [118].

A pair of prototype circuit boards were assembled by hand, shown in Figure 12, using stencils and reflow oven to achieve a reliable solder connection. The official Debian Linux-based operating system was loaded to the SD card, and the system was able to boot without complications, although additional drivers and software

packages were needed to enable the functionality of the multitude of on-board peripherals. Each major component was then individually tested to ensure their functionality, which all appeared to be operational.

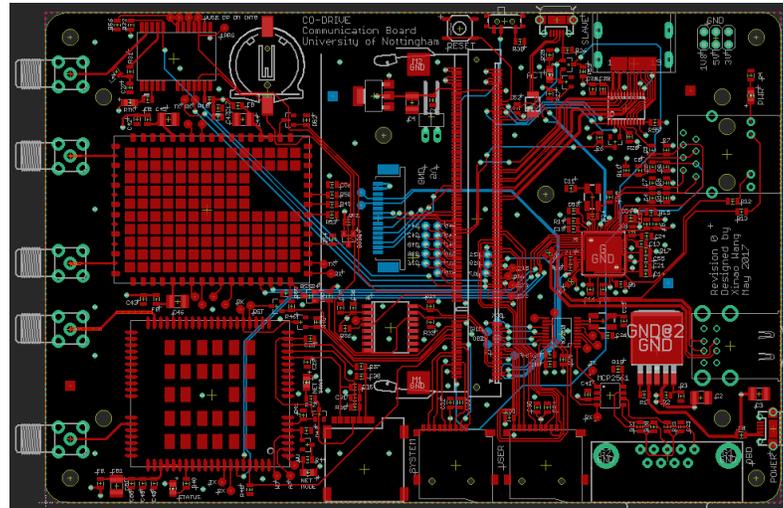


Figure 11: All-in-one OBU Circuit Design



Figure 12: All-in-one OBU assembled PCB.

3.6. Modular OBU

While the all-in-one design proved to be fully functional and was used in the real-world preliminary outdoor experiment which will be discussed in the next chapter, the progress of the CoDRIVE project soon dictated additional sensors and components to be included, such as the IMU modules and RFID transceiver.

A new design iteration was decided upon, taking into consideration aspects of modularity and upgradability instead of simply adding the new components to the existing circuit board, in order to future-proof the device and facilitate more straightforward updates and modifications. An analysis of existing standards for compact and modular computer form factors were conducted. Miniature computer motherboards standards exist such as Mini, Nano, and Pico-ITX form factors, with dimensions as small as 75mm x 45mm. However, their expansion options are limited, often providing only a single slot, if at all. Another miniature PC form factor known as PC/104 was investigated, which appears to better suit the modularity and upgradability requirements.

PC/104 is a family of standards, devised by the PC/104 Consortium, that defines both form factors and connection pinouts for embedded computers. The circuit boards measure 90 x 96mm in size, and are designed to be stacked on top of each other. Layers are interconnected via a 104-pin header on one side of the circuit board and secured by screws and standoffs on each corner [119]. This design offers advantages such as modularity and upgradability, compact footprint, and more rugged construction with better resistance to shock and vibration due to the additional fasteners and reduced circuit board flex, making it ideal for vehicular environments. With this form factor, the new OBU hardware can be easily upgraded via additional circuit board layers, allowing it to remain relevant and adaptable to emerging technologies and use-cases.

With the aforementioned advantages in mind, the second revision of the OBU is designed in accordance with the electrical and physical dimensions of PC/104 format. The new revision consists of three stackable layers, each with a particular function in mind.

The bottom layer handles computation, and contains the hardware of the embedded computer itself, including the Raspberry Pi Compute Module, Real-

time Clock, power supply circuits, SD card slots, Ethernet port, USB port, and HDMI video output.

The middle layer handles communication, and contains the V2X transceiver, cellular module, their respective antenna connectors, micro-SIM card slot, CAN Bus transceiver and controller, as well as an additional USB port.

The top layer is responsible for positioning, and contains the newly added IMU modules and RFID transceiver, the GNSS receiver, and the respective antennas. Numerous quality-of-life upgrades were also introduced, such as removal of unused parts, USB-C power connector instead of Micro-USB, and a new barrel-jack power connector for a sturdier power connection.

By splitting the system into three functionally distinct layers, the footprint of the OBU is approximately reduced by half compared to the previous iteration, and the platform is now easily upgradable by adding or replacing circuit boards on the stack. The individual layers and the complete stack are shown in Figure 13. A summary of hardware specifications is listed in Table 2.



a) Compute Layer



b) Communication Layer



c) Positioning Layer



d) Fully Assembled

Figure 13: The PC/104 OBU

Table 2: OBU Specification

Component	Model	Specification
Processor	ARM Cortex A53	Quad-core 1.2GHz
DSRC Transceiver	u-blox THEO-P173	IEEE 802.11p -97dBm RX sensitivity +25 dBm TX power Up to 54 Mbps data rate
GNSS Receiver	u-blox NEO-M8T	2.5cm CEP with RTK 5Hz update rate All major constellations
Cellular modem	Quectel EC20	LTE/CDMA/EDGE/GSM Up to 100Mbps download 50Mbps upload
Wi-Fi / Bluetooth	Realtek RTL8723BS	IEEE 802.11 b/g/n Bluetooth 2.1/3.0/4.0 + EDR
CAN Bus Transceiver	MCP2515	CAN V2.0B at 1Mb/s High Speed SPI Interface
IMU	ADIS16488 ADIS16460	Tactical Grade 10 DoF
RFID	ThingMagic M6E	EPC Global Gen 2 Tags Read/Write Up to 27dBm output power Up to 200 tags/sec reading

3.7. Initialisation Procedures

Upon power-up, the On-Board Unit (OBU) executes an initialisation script to set up the various modules, making them ready for use. First, the system clock is synchronised to the date and time from the GNSS receiver, ensuring that the timestamps of the data and logs collected during the experiments are accurate. This is followed by the execution of a shell script to clean up data and temporary files from previous experiments, ensuring a clean environment for the new setup. Subsequently, the GNSS receiver is initialised to provide raw measurement data while observing all major constellations, including GPS, GLONASS, BeiDou, and Galileo.

To establish a data connection, the user employs a Point-to-Point Protocol (PPP) through the ttyUSB3 interface. Following this, power is supplied to the EC20 cellular module by executing the `ec20_poweron` script. The system's routing table is updated to allow data to flow through the PPP interface, and connectivity is verified by pinging Google's DNS server at 8.8.8.8. A keep-alive script periodically checks to ensure the connection remains active and attempts to reconnect if the cellular data connection is interrupted.

Subsequently, RTKLIB is set up using a predefined configuration file, establishing the RTK correction data stream and the required data path for the experiments. The data logging for the Inertial Measurement Unit (IMU) is then initiated to provide a source of data for dead reckoning applications. Lastly, the DSRC transceiver is activated, and data transmission commences.

This detailed initialisation process ensures that all the necessary modules and components of the OBU are properly configured and ready for data collection and transmission during the experiments, which is critical for the success and integrity of the experimental results.

Chapter 4: Essential Tools and Metrics

With the completion of the OBU hardware platform design, multiple outdoor experiments were planned in order to verify its functionality and to assess the performance of V2V communication in real-world conditions. This chapter introduces the software packages and digital tools that were utilised for data processing, visualisation, and analysis, as well as outlining a selection of Key Performance Indicators (KPI) which serve as the metrics to gauge the performance of the system in a quantifiable manner. By doing so, this chapter provides an overview of the concepts which will be essential for the discussions in later chapters of this thesis.

4.1. Software Packages and Services

This project makes extensive use of Free and Open-Source Software (FOSS) packages and services during the development of the OBU hardware and the subsequent experiments, including data gathering, processing, visualisation, and creation and training of machine learning models.

This section provides an overview of the key software components and services used in this project. The benefits of these FOSS selections include not just cost-effectiveness, but also the flexibility and customisability to adapt to project-specific needs, as well as comprehensive community support during development and troubleshooting. These elements greatly helped in optimising the project's resources and the realisation of its objectives in an efficient manner.

4.1.1. Python

Python is a popular general-purpose programming language, known for its comprehensive standard library, uncluttered syntax, and a design philosophy that emphasises readability and simplicity.

Python is often used for general-purpose scripting, network programming, data analysis, machine learning modelling, scientific computing, and more due to its versatility and gentle learning curve [120]. Its popularity has resulted in an expansive community and the development of a comprehensive collection of high-quality third-party libraries.

Python is used inside the OBU for configuring various sensors and communication modules on start-up, as well as reading and logging their output to onboard storage. The devices include the DSRC Transceiver, GNSS receiver, cellular modem, and IMU.

NumPy and Matplotlib are also used for data processing and visualisation after the experiments. They are two important libraries in the Python ecosystem.

NumPy, short for Numerical Python, provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions, serving as the backbone for numerical computations in Python [121]. Matplotlib is a plotting library that generates a static, animated, and interactive plots in a highly customisable manner [122].

When used together, NumPy's numerical capabilities and Matplotlib's visualisation tools form a powerful basis to manipulate, analyse, and visualise large datasets efficiently, and they play a significant role in machine learning and scientific computing using Python.

4.1.2. QGIS

QGIS is a free and open-source geographic information system (GIS) application that facilitates viewing, editing, and analysis of geospatial data. QGIS supports a wide variety of vector, raster, and database formats as well as map composition and georeferencing, making it a comprehensive tool for managing and analysing geographical data [123].

Compared to paid and proprietary software such as ArcGIS, QGIS allows free and open access to a comprehensive suite of GIS tools without the financial burden of licenses, in addition to the robust support and regular updates from the community. As a result, QGIS is used in a wide range of geospatial applications and studies, from ecosystems modelling and traffic simulation [124] [125], to usage in governmental institutions such as US National Security Agency (NSA) [126] and Land Information New Zealand [127].

QGIS is used in this project to provide visualisation of GNSS trajectories, categorical data, and magnitude of variables of the data collected from the outdoor experiments.

4.1.3. OpenStreetMap

OpenStreetMap is a free and open-source project committed to generating a comprehensive, freely accessible, and editable map of the world, involving a diverse community of contributors who gather data through means such as manual surveys, GPS devices, aerial imagery, and other freely available sources. It operates on a premise similar to Wikipedia, allowing users globally to contribute and amend the map database.

Unlike its commercial counterparts, OpenStreetMap operates under the Open Database License, eliminating the usual restrictions associated with proprietary mapping services. This allows individuals and organisations to use its data in innovative ways without the constraints of licensing fees or usage limitations, and ensures geospatial information is not only limited to those who can afford it, making it a valuable resource for non-profit organisations and researchers alike. The community can also continuously update and refine the data, ensuring its timeliness and accuracy, especially in rapidly changing environments such as during natural disasters [128]. The data from OpenStreetMap has been utilised in commercial mapping and navigation services such as Facebook Map and Apple

Maps [129] [130], and many corporate sources and government agencies have also made notable contributions back to the map database [131].

OpenStreetMap also forms the cornerstone of a large portion of geospatial research studies, where the open-source mapping data is used in many different topics, such as humanitarian mapping [132], Machine Learning training [133], socio-economic mapping [134], and autonomous vehicle navigation [135].

In this project, OpenStreetMap is used extensively during route planning, distance measurement, as well as the overlay for visualisation of GNSS trajectories and related data inside QGIS.

4.1.4. RTKLIB

RTKLIB is an open-source software package dedicated to GNSS positioning processing. It provides a powerful suite of applications for single-point and differential positioning (DGPS/DGNSS), Precise Point Positioning (PPP), and Real-time kinematic positioning (RTK), offering accuracy up to centimetres or millimetres with suitable equipment under appropriate conditions [136]. Tools for data conversion, visualisation, and post-processing analysis are also included, both in command-line and graphical user interfaces.

RTKLib has been extensively used in studies to provide high-accuracy positioning output using its RTK or PPP capabilities. Examples include achieving centimetre-accuracy using low-cost GNSS receiver in UAV applications [137], precise positioning in real-time embedded systems [138], and improving smartphone localization accuracy [139].

In this project, RTKLIB is used to perform real-time RTK processing during the outdoor experiments by combining the raw measurement data from the GNSS receiver with the correction data stream received from either cellular network or DSRC communication to improve the positioning accuracy. The highly

customisable nature of the software allows the configuration to be fine-tuned to the specific setup for the equipment and environments during the tests.

4.1.5. Weka

Weka is a free and open-source software suite developed at the University of Waikato, New Zealand. It contains a comprehensive collection of tools suitable for tasks involving machine learning, data analysis, and data mining, such as preprocessing, clustering, classification, regression, visualisation, and feature selection [140].

The tools are presented in intuitive graphical user interfaces, and are designed to work seamlessly with one another, allowing users to conduct complex analytical tasks on a cohesive and integrated platform from start to finish. As a result, Weka is used extensively in teaching and research, and enjoys a broad user base from beginners to seasoned academic researchers, and is used in a wide range of applications such as data mining [141], medical classification [142], and ML model training.

In this project, Weka is used to visualise collected data from the V2V communication module during the outdoor experiments to identify trends and correlations between related parameters. It is also used to pre-process and perform machine learning experiments with the relational data, with different algorithms, training parameters, selected features, and verification methods, in an attempt to explore the possibilities of using machine learning models to predict and prevent communication interruptions in V2V communication.

4.2. Key Performance Indicators

Key Performance Indicators (KPIs) serve as quantifiable measurements that assess the performance of a system during its testing phase, allowing ascertainment of whether the system is meeting the predetermined objectives and requirements, as well as to identify areas of improvement or optimisation

[97]. In the scope of this project, a selection of specific KPIs has been identified to evaluate the OBU's performance.

4.2.1. Received Signal Strength Indicator (RSSI)

In the context of telecommunications, the Received Signal Strength Indicator (RSSI) serves as a vital metric for measuring the power level received by a wireless radio. A higher RSSI value typically corresponds to a stronger received signal, and thus a higher likelihood of successful communication between two radio units [143].

Although the relationship between the value of RSSI and specific physical parameters is not universally standardised, most manufacturers opt to represent the RSSI value in terms of power level measured in milliwatts (mW), or decibels referenced to one milliwatt (dBm), as is the case for the V2V transceiver utilised in this project.

The decibel-milliwatt (dBm or dBmW) is a unit of measurement that expresses power levels in decibels (dB) relative to a single milliwatt (mW). Given its logarithmic properties, dBm efficiently represents values spanning a wide range in a compact format, and therefore is the preferred unit for quantifying absolute power in various communication networks, including radio, microwave, and fiber-optic systems. The logarithmic nature of the dBm scale means that a power level of 0 dBm corresponds to a power of 1 mW, and a change of 3 dBm approximately doubles or halves the power level.

The uBlox P173 V2V communication module has maximum transmit power of 23 dBm, which is approximately 200 mW, and a minimum receiving sensitivity of -97 dBm, as it is the weakest signal it can successfully decode [42]. RSSI can also be affected by external factors such as quality of antenna, quality and length of cabling, antenna trace on the printed circuit board, and interference with other signals in the environment. To improve RSSI, a high-quality antenna, short and low-loss antenna cables, impedance-matched PCB trace, and ground shielding,

and a more sensitive receiver itself can be utilised. These measures can contribute to improved communication reliability, reduced transmission errors, and an overall more robust communication system.

4.2.2. Signal-to-Noise Ratio

The signal-to-noise ratio (SNR) measures the relationship between the power of a desired signal and the power of background noise. SNR is defined as the ratio of signal power to noise power. A ratio higher than 1 indicates that there is more signal than noise.

A high SNR indicates that the desired signal is clear, distinct, and readily detectable or interpretable. Conversely, a low SNR suggests that the signal is corrupted or obscured by noise, making it difficult to differentiate or recover. SNR can be improved by increasing transmission power, reducing the noise level, filtering out unwanted noise, or implementing error correction techniques. In this project, the V2V transceiver reports both the received signal power and received noise power in dBm. Consequently, the SNR can be easily computed by subtracting the noise power from the signal power.

In [144], Wang et al. examined the impact of communication system and environmental variables on spectrum efficiency and data rate within a DSRC and Wi-Fi shared spectrum environment. Their analysis highlighted that SNR and DSRC node density are critical factors affecting the communication performances.

4.2.3. Packet Error Rate (PER)

Packet Error Rate (PER) is another important parameter in evaluating the performance and reliability of wireless communication systems. It represents the ratio between the number of transmitted packets that are either missed or contain errors, and the total number of transmitted packets.

Each DSRC message packet contains a unique sequence number that increments linearly with each successive transmission. By examining the received sequence numbers at the receiving end, it is possible to identify missing or out-of-order packets, which can then be used to calculate the PER. In [145], M. Shi et al. found that a higher PER indicates a less reliable and effective wireless link, potentially affecting the quality and reliability of the transmitted information.

A variety of factors can contribute to packet errors in wireless communication systems, including signal attenuation due to distance, interference from other radio signals, multipath propagation interference, and fading effects caused by changes in the environment or the relative positions of the transmitting and receiving devices. In V2V communication, these factors can be exacerbated by the dynamic nature of vehicular environments, where vehicles are constantly in motion and potentially experiencing rapidly changing communication conditions.

The impact of PER on V2V communication performance can be significant. High PERs may lead to increased latency, reduced throughput, and a greater likelihood of communication failures, which can undermine the safety and efficiency benefits provided by V2V communication systems.

4.2.4. Packet Reception Rate (PRR) / Communication Range

Similar to PER, Packet Reception Rate (PRR) is defined as the ratio of successfully received packets to the total number of transmitted packets, which can be expressed as $PRR = 1 - PER$.

PRR is a valuable tool for determining the effective communication range between two transceivers in a V2V communication system, as it defines the maximum distance over which reliable communication can be maintained. As the distance between the transmitter and receiver increases, the signal strength diminishes, resulting in a higher likelihood of packet errors and a decrease in PRR.

The maximum communication range is typically considered to be reached when the PRR drops below a specified threshold, as a lower PRR could increase the likelihood of communication failures, which would undermine the safety and efficiency benefits provided by V2V systems. The choice of an appropriate threshold may depend on the specific application requirements and the desired balance between communication range, available alternative channels, cost of operation, and reliability.

In practical scenarios, the PRR threshold is often set between 90% to 99%, depending on the type of the payload, where safety critical messages have a higher threshold [97] [146].

4.2.5. Interpacket Gap (IPG)

IPG is defined as the time interval, typically measured in milliseconds, between the reception of two consecutive packets at the receiving end of the communication system. It provides valuable insights into the performance and reliability of the wireless communication link.

In DSRC communication, the transmitter typically broadcasts messages at a regular interval, maintaining a constant flow of information between vehicles. When the communication link is functioning optimally, the IPG remains relatively stable. However, when a packet is lost or delayed due to interference or lost signal, the IPG at the receiver increases, indicating deteriorating communication conditions, which could eventually lead to communication failures or reduced performance.

In [97], Jeong et al. pointed out that IPG is correlated to the PRR indicator, and in their real-world test the IPG remained around 100ms until the space between the test vehicles exceeds the maximum communication distance, where the IPG increases rapidly to over 1000ms. Therefore, the authors deemed IPG an important metric in evaluating the performance of a wireless communication system.

By monitoring IPG performance, the transceivers can identify potential issues and take appropriate measures to optimise the communication link or mitigate the effects of interference and other adverse factors.

Chapter 5: Preliminary Test at Wollaton Park

Following the completion of the first OBU hardware prototype, a preliminary outdoor experiment was planned and carried out at a local park (Wollaton Park) with wide open outdoor spaces, to validate the functionality and observe the performance of the device. This chapter covers the test objectives, plan, procedures, data visualisation and analysis, and a discussion of the outcome of the experiment.

5.1. Test Objectives

The primary objectives of this real-world experiment included assessing the functionality of the key components, such as the processor, the DSRC communication module, and the GNSS receiver module, as well as determining the communication range, stability, and reliability of the entire system under near-ideal conditions. By comparing the performance parameters obtained during the test with the values claimed in the respective datasheets, the experiment aims to establish a baseline for the system's performance in an environment characterised by line-of-sight communication and minimal obstructions. This initial assessment will not only provide valuable insights into the effectiveness of the hardware prototype but also lay the foundation for further experiments and refinements in the development of a robust V2V communication system.

5.2. Test Plan and Procedures

A test plan and set of procedures were devised to evaluate the performance of the newly assembled OBU. The experiment utilised two identical prototype devices, one configured as a transmitter and the other as a receiver. The transmitter was mounted at a fixed location, remaining stationary throughout the test. The receiver was mobile, allowing for free movement relative to the transmitter. Each OBU was equipped with two manufacturer-recommended omnidirectional 5.9GHz DSRC rod antennas.

Operating at its maximum output power of 23dBm, the transmitter broadcast DSRC messages with both antennas set to transmit mode. The receiver, in turn, listened for the DSRC broadcasts, with both its antennas set to receive mode. This configuration aimed to maximise the potential for optimal signal reception and transmission.

The test starts with the transmitter and the receiver next to each other in close proximity. Once the DSRC link is confirmed, the receiver is gradually moved away from the transmitter in a straight line. The GNSS receiver onboard the receiver, operating in an unassisted standard mode without the aid of RTK or satellite-based augmentation systems, recorded its coordinates. Concurrently, the KPIs were logged for subsequent processing and analysis. Both GNSS and KPI data were logged at a rate of 1Hz. The test continued until the receiver could no longer detect any signal from the transmitter. Once the DSRC wireless link is lost, the receiver begins the return trip to the transmitter along the same route, until it reaches the starting point. The gathered data is processed, visualised, and discussed in the next section.

Photos of the test environment are shown in Figure 14.



Figure 14: Wollaton Park test environment

5.3. Dataset Visualisation and Analysis

This section presents a detailed analysis and visualisation of the data gathered from both outwards and return trip of the outdoor experiment.

For the outwards portion of the experiment, Figure 15 provides a map overlay of the GNSS positioning data, with colours representing the received signal power levels in dBm. Figure 16 to Figure 18 show scatter plots of the received power level, noise level, and signal-to-noise ratio for both Antennas A and B in dBm, against the distance between the transmitter and receiver in meters. These figures also include a red line that shows the running average of the latest 10 consecutive data points, providing a visual aid for understanding the average trends in the data. The statistics of the collected data are shown in Table 3, including mean, median, max, min, range and standard deviation of all attributes.



Figure 15 : Visualisation of Test Route

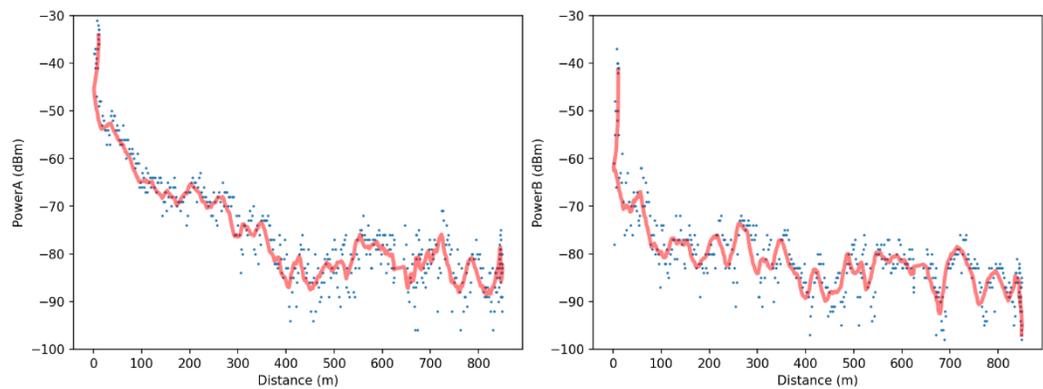


Figure 16: Received Power Level, Outward Trip.

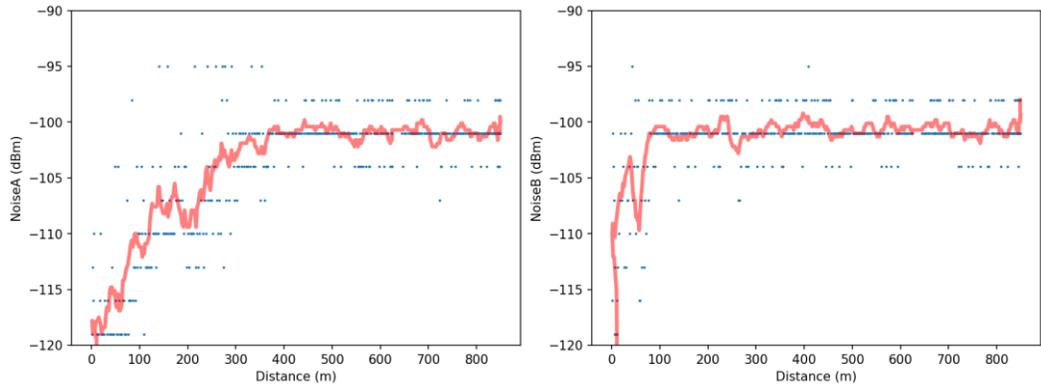


Figure 17: Received Noise Level, Outward Trip.

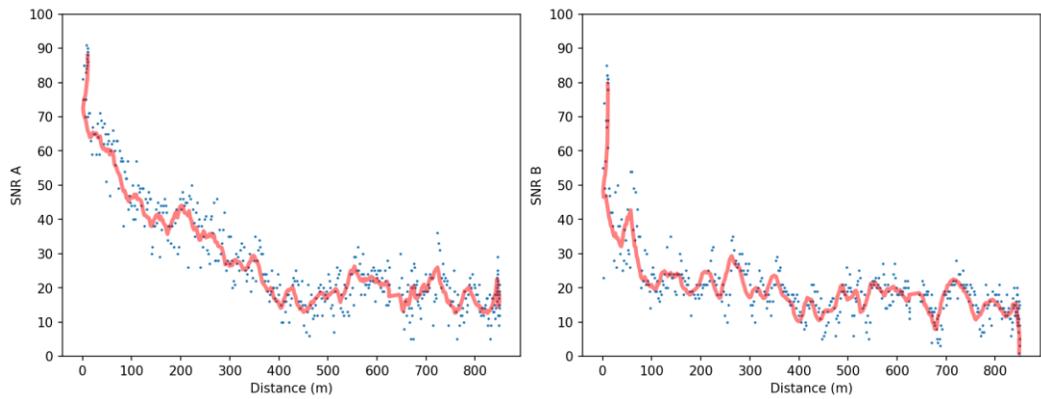


Figure 18: Signal-to-Noise Ratio, Outward Trip.

Table 3: Outward Trip Statistics

	Power (dBm)		Noise (dBm)		SNR	
	ANT A	ANT B	ANT A	ANT B	ANT A	ANT B
Mean	-75	-81	-104	-101	30	21
Median	-78	-82	-101	-101	23	19
Max	-31	-37	-95	-95	91	85
Min	-96	-98	-125	-122	5	-2
Range	65	61	30	27	86	87
SD	12.51	9.51	6.48	4.34	18.27	13.10

It can be observed that as the distance between the transmitter and receiver increases, the received power decreases proportionally, with the communication

interruption occurring at around -95 dBm. The maximum communication distance in this experiment appears to be around 890 meters.

The received noise power level appears to increase with distance, demonstrating a trend similar to logarithmic growth. This trend plateaus at around -101 dBm for the majority of the distance, with a sharp increase to around -95 dBm just before communication disruption. At this point, the received power level and the noise level are almost the same, leading to an SNR close to 1, indicating that the useful signal is being overwhelmed by noise.

Evaluating the performance of the two antennas, it appears that Antenna A presents a higher average received power level and a lower noise level compared to antenna B, resulting in a higher average signal-to-noise ratio. However, the range between the maximum and minimum values appears to be similar for both antennas.

Antenna B's statistics demonstrate less variance and standard deviation compared to Antenna A, indicating a tighter clustering around the mean value. This can be observed in the plots where the received power level, noise level, and SNR for Antenna B exhibit a more rapid decline and stabilisation compared to Antenna A with increasing distance.

Similar analyses are performed on the data collected during the return trip of this experiment. The received power levels, noise levels, and the signal-to-noise ratio for both Antenna A and B in dBm are charted against distance between the transmitter and receiver in meters in Figure 19 to Figure 21, with the red line indicating a 10-element running window average. The plots use the same x and y axis scales as the outward trip. The statistics are presented in Table 4.

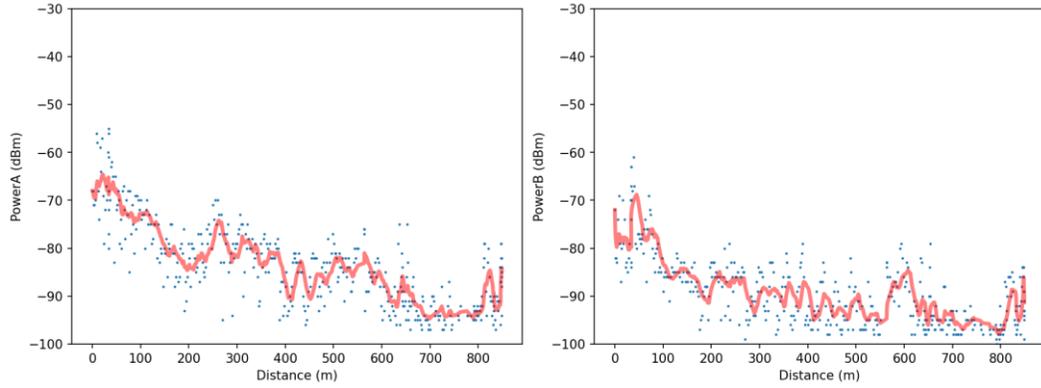


Figure 19: Received Power Level, Return Trip.

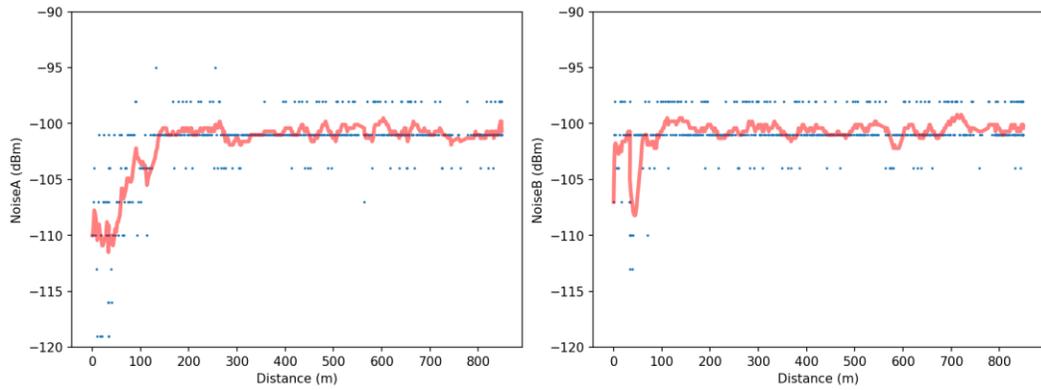


Figure 20: Received Noise level, Return Trip.

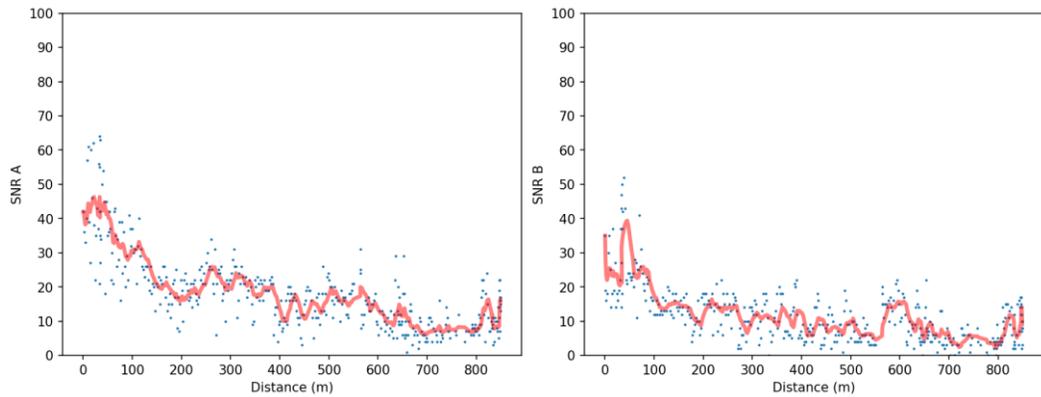


Figure 21: Signal-to-Noise Ratio, Return Trip.

Table 4: Return Trip Statistics

	Power (dBm)		Noise (dBm)		SNR	
	ANT A	ANT B	ANT A	ANT B	ANT A	ANT B
Mean	-83	-89	-101	-101	18	12
Median	-84	-90	-101	-101	17	10
Max	-55	-61	-95	-98	64	52
Min	-97	-100	-119	-113	1	-2
Range	42	39	24	15	63	54
SD	8.65	7.12	3.41	2.15	11.08	8.20

The patterns observed during the return trip largely echo those seen during the outward trip. The received power level decreases as the distance between the transmitter and receiver increases. Similarly, the received noise level on both antennas increases with distance, resembling a logarithmic growth, and eventually stabilises around -101 dBm. Antenna A continues to outperform Antenna B in terms of higher average received power levels and lower noise levels, resulting in a superior average signal-to-noise ratio overall. Data from Antenna B still demonstrates less variance and standard deviation compared to Antenna A, and clusters more closely around its mean value, although this difference is less distinct than it was on the outward trip.

However, some differences can be observed when compared to the outward trip. Most notably, the overall average received power level is lower compared to the outward trip, at about 8 dBm less. Additionally, the noise level seems to be higher when the transmitter and receiver are relatively close, approximately less than 200 meters apart. As a consequence, the overall signal-to-noise ratio is approximately 8 units lower.

Several potential factors could be responsible for the discrepancy observed in the average received power between the outward and return trip. One possible factor could be the orientation of the antenna. The employed omnidirectional antenna is designed to distribute radio power equally in all directions orthogonal to the antenna axis. However, any deviation from the optimal position, such as

tilting or off-axis movements, could impact the amount of signal the antenna can receive [147]. If the antenna orientation was not maintained consistently during the return trip, it could have contributed to a lower average received power.

Another plausible explanation may be the obstruction caused by the human body, given that the unit was handheld during the test. Human bodies, composed predominantly of water, can absorb radio frequency waves, thus acting as a barrier that weakens the signal strength reaching the antenna.

The influence of the surrounding environment and terrain is another factor to consider. The physical landscape can affect the propagation of radio waves, leading to reflection, refraction, or diffraction off various obstacles. The interplay of these multiple signal paths can lead to either constructive or destructive interference, thereby affecting the strength of the received signal.

5.4. Summary

This preliminary outdoor evaluation signifies a milestone in the assessment of the newly developed OBU hardware prototype. The primary objective of this experiment was to confirm the functionality of the hardware and evaluate the performance of components including the embedded processor, which is responsible for controlling the overall system, the GNSS receiver, responsible for receiving and processing satellite signals to determine precise location information, and the V2V module, which handles the transmission and reception of DSRC communication.

The experiment demonstrates that the hardware design is not only functional, but also capable of operating in a stable and reliable manner. The data collected during the experiment indicates that a maximum communication range close to 900 meters can be achieved under close-to-ideal conditions. Furthermore, it was observed that the two antennas exhibited slight differences in their performance, and the received signal power level are likely to be sensitive to antenna placement, orientation, and environmental obstructions.

The information and insights derived from this experiment offer valuable understanding of how DSRC technology performs in real-life environments, and it will be instrumental in guiding the planning and execution of future testing phases.

Chapter 6: Multi-Vehicle Road Test

Upon the satisfactory conclusion of the preliminary test outlined in the previous chapter, a new and more extensive outdoor experiment was soon set in motion, with the objectives closely aligning with the primary goal of this thesis of evaluating the real-world performance of V2V communications.

This upcoming experiment was designed to be considerably more comprehensive than its predecessor, featuring real-world driving scenarios with multiple road-legal vehicles navigating public roadways under a diverse range of traffic conditions, road environments, and vehicular speeds.

This chapter covers the details of the objectives of this experiment, the planning process, the implemented procedures, and the parameters of the data collected. It further explores the methods employed for data visualisation and analysis, and discusses observations made during the experiment.

6.1. Test Objectives

The primary aim of this experiment is to thoroughly assess the performance of the OBU under realistic conditions, specifically, on public roads with real vehicles. To achieve a comprehensive evaluation, several objectives have been identified and are outlined as follows:

- Careful planning of a test route with longer distance and diverse environments, using road-legal vehicles under various traffic situations and road conditions at different speeds.
- Organising the logistics of the experiment, including the selection of suitable vehicles for the test, coordinating with team members to determine an appropriate date for the experiment, and the development of safety protocols to ensure that the experiment is conducted in a safe manner.

- Collecting KPI data from the V2V communication module. This data will provide valuable insights into the functionality and reliability of the OBU hardware in operation.
- Transmit and receive DSRC data at a higher frequency to mirror real-world application conditions more closely.
- Transmission of GNSS correction data stream as payload via DSRC. This is intended to enhance the positioning accuracy of the vehicle that is receiving the DSRC data. This is particularly advantageous for vehicles that are not equipped to receive the GNSS correction data stream directly, presenting a practical and innovative use of DSRC.
- Gathering data on the cellular environment along the test route. The data will help in understanding the potential C-V2X communication as an alternative to DSRC.
- Processing, visualisation, and analysis of the collected data, a detailed discussion of the results, and a summary of the outcomes of the experiment.

The objectives outlined here are designed to ensure the smooth and successful execution of the new outdoor experiment. By providing clear guidance and defining the target outcomes, a robust assessment of OBU's performance under real-world driving conditions can be expected to be achieved.

6.2. Test Plan and Procedures

A detailed test plan was formulated in order to cover each objective in a thorough and comprehensive manner. This section provides a description of different aspects of the plan and procedures involved in the execution of the experiment.

6.2.1. Route Planning

Given the primary objective of evaluating OBU performance under real-world circumstances, it is crucial to select an appropriate test route featuring a variety

of road types and driving conditions representative of typical road environments in UK.

Considering the points above, as well as time and personnel requirements, a test route approximately 30 miles in length was proposed. The route forms a counterclockwise loop, beginning and ending at the same location, and should take around 2 hours to complete in moderate traffic.

The test commences from the village of Plumtree in Nottinghamshire, UK. The vehicles start by travelling northwest following the A606, then continues on the A52 dual carriageway, leading into the city of Nottingham. The route traverses westward through the residential area of Wollaton, then heads north via the A6002 to the town of Bulwell. This is the halfway point of the test route, and a short break can be taken if needed. The return journey takes the vehicles back along the A6002, although this time through additional residential areas of Wollaton Vale. The final stretch of the route travels through the town of Beeston before merging back onto the A52 and returning to Plumtree. A visualisation of the proposed route is shown on Figure 22, coloured according to the speed limit.

This test route features a diverse blend of typical road types and driving environments in UK, including dual carriageways with a speed limit of 70MPH, single carriageways with speed limits from 40 to 60 MPH, city streets with 30MPH limit, and residential streets with 20MPH speed limit, optimising the opportunity of obtaining insightful and valid data on the performance of the OBU in a real-world context.

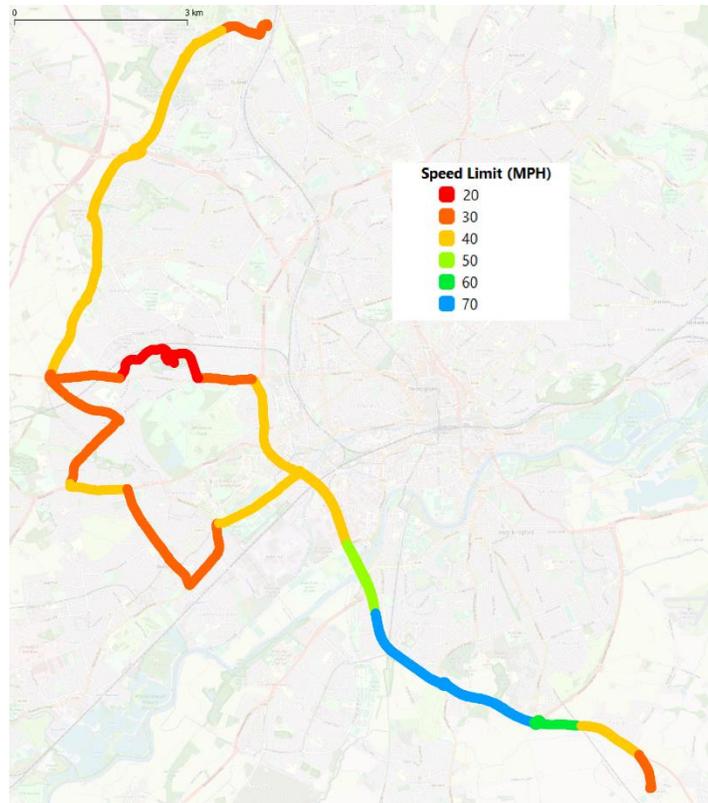


Figure 22: Proposed Test Route

6.2.2. Vehicles and Equipment

Two road-legal vehicles will take part in this experiment by traveling along the predetermined route in a normal manner, obeying the speed limit as well as all other traffic rules, while carrying the necessary equipment. One vehicle will always stay in front, hereby referred to as the "leading vehicle", and the other vehicle will always be following behind, hereby referred to as "following vehicle".

The leading vehicle's equipment includes an OBU configured as a transmitter for V2V communication, a Leica GS10 RTK GNSS receiver serving as the reference for GNSS positioning, and a laptop containing software for the purposes of device initialisation and data collection. A number of accessories were also included, such as a power inverter, backup battery pack, and a dash-mounted front-facing camera for a video record of the experiment's duration.

The OBU and Leica receiver's GNSS antennas are installed at the centre of the leading vehicle's roof, maintaining an approximate separation of 40 centimetres from one another. Moreover, the OBU is situated on the parcel shelf at the rear of the vehicle, employing the manufacturer-provided 5.9GHz DSRC antenna. The DSRC module is configured to operate at its maximum transmission power of 23dBm, using both antennas for broadcasting.

Similarly, the following vehicle is also equipped with an OBU, functioning as the receiver, a Leica GS10 GNSS reference receiver, and a laptop for device initialisation and data collection, as well as similar supporting accessories. The GNSS antennas are also mounted in a similar fashion on the roof of the following vehicle. However, the OBU and its DSRC antennas are positioned on the dashboard of the vehicle, and the two antennas on the DSRC module are both configured for signal reception to ensure optimal signal quality.

Beyond V2V communication, each OBU is also equipped with an active cellular data connection with the same service provider, EE. This allows real-time GNSS correction data stream to be fetched from the internet, as well as logging a range of cellular environment parameters for later examination of the feasibility of Cellular V2V as a potential alternative to radio based V2V communications.

Photos of the equipment setup during the experiment are shown in Figure 23.



Figure 23: Road trial equipment setup

6.2.3. Logistics and Safety

As this experiment involves multiple vehicles and personnel, arranging the logistics plan plays an important role in ensuring the safety and efficiency during the test.

In preparation for this experiment, the detailed test plan was formulated in advance and disseminated to all involved parties. A meeting was scheduled to update all team members with the test objectives, route, equipment operation, as well as selecting a convenient date and time. A dress rehearsal with a shorter test route was planned to familiarise the team members with the setup process, verify functionality of the equipment, as well as to identify any areas worthy of special attention during the main experiment.

Additionally, a comprehensive risk assessment was undertaken to pre-emptively identify any potential hazards that could arise during the experiment. A mandatory safety briefing was also conducted before the start of the test, serving

to familiarise all team members with the safety protocols, and emphasize the importance of adhering to these guidelines at all times.

The leading vehicle was manned by two individuals: a driver, who focused solely on navigating the test route, and the author, responsible for monitoring equipment to ensure correct operation. In contrast, the following vehicle was operated by a single individual, the driver. The team was also equipped with two-way radios allowing for real-time communication and updates.

A significant advantage of V2V communication is its potential to enhance safety, enabling vehicles to detect each other's presence beyond the line of sight, thus allowing earlier warnings and more time to react to possible hazards. However, as this experiment will be taking place on public roads, a high standard of safety must be maintained at all times, therefore the decision was made against executing any specific driving manoeuvres that could pose potential risks, and the devices have been configured to passively gather data for detailed analysis later.

6.2.4. Payload Structure and Data Path

In addition to the longer test route in a more diverse driving environment, this experiment also features an enhanced data transmission and logging methodology, integrating information from a wide range of data sources. A chain of data pathways was also devised, through which data is captured, processed, transmitted, received, and utilised. This section offers a detailed discussion of those topics.

The payload of the DSRC communication is configurable, allowing flexible integration of various types of data as needed. For this particular test, the DSRC message from the leading vehicle to the following vehicle contains the following components:

- Sequence number. This is an integer that increases in a sequential manner with each transmitted message. It can be used to synchronise data between the transmitter and receiver during post-processing, as well as to keep track of any packets that might be delayed or lost.
- GNSS positioning results. Including location in latitude and longitude, velocity in knots, heading in degrees, as well as GNSS time and date. This data plays a crucial role in monitoring the location and behaviour of nearby vehicles, as well as performing safety-related actions and manoeuvres.
- Real-time GNSS correction data. This is the data stream produced from a nearby Continuous Operating Reference Station (CORS) obtained via cellular internet connection. This correction data allows compatible GNSS receivers to enhance positioning accuracy by minimising the impact of various sources of inaccuracies [148]. By broadcasting this correction data over V2V, it is hoped that further improvement in positioning accuracy of nearby vehicles can be achieved, especially in challenging road environments.

The frequency of the DSRC communication is set to 10Hz, and each individual message carries a fixed-size payload of 2048 bytes or 2KB of data.

The GNSS receiver is configured to observe all major satellite constellations, including GPS, GLONASS, BeiDou, and more. However, instead of relying on its built-in positioning engine, the GNSS receiver is configured to output its raw measurement data, containing information such as carrier phase, code phase, Doppler measurements, and pseudorange measurements.

The CORS data are then sent to the STRSVR utility as part of the RTKLIB software suite. This splits the single data stream into two separate streams.

One stream of the raw GNSS measurement data, along with the real-time GNSS correction data sourced from a nearby CORS through the cellular network via the

on-board cellular modem, is sent to RTKNAVI utility for real-time RTK calculation onboard the OBU at the same frequency as the DSRC communication. The resulting enhanced-accuracy location information, as well as related parameters such as velocity, heading, and time and date, are then sent to the DSRC transceiver to be included as part of the outgoing DSRC message. The other stream of raw GNSS measurement data is also sent directly to the DSRC receiver to be included in the payload. Copies of the outgoing DSRC packets are also archived in the onboard storage for post-experiment analysis.

The data path for the DSRC transmitter can be seen in Figure 24.

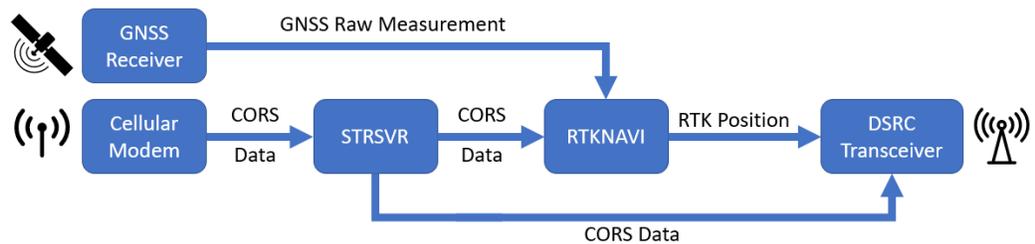


Figure 24: Data Path of DSRC Transmitter

The GNSS receiver on the following vehicle is configured identically to that of the leading vehicle, observing all the major satellite constellations, and programmed to output raw measurement data.

Once a DSRC message is received from the leading vehicle, the payload is extracted, and the GNSS correction data is provided to the RTKNAVI utility along with the raw GNSS measurement data stream. This allows RTK positioning with increased accuracy even in the absence of an active on-board internet connection, enhancing the positioning performance of vehicles within the vicinity. The positioning result, raw DSRC message contents, as well as other KPIs, are also saved to the on-board storage to ensure a comprehensive dataset for subsequent review and analysis.

The data path for the DSRC receiver can be seen in Figure 25.

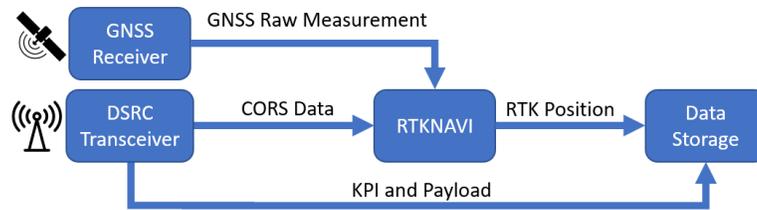


Figure 25: Data Path of DSRC Receiver

The internal clocks of both OBUs are synchronised to the GNSS clock once a valid positioning fix has been obtained. This allows the two devices to operate on the same timebase, allowing valid and accurate comparison of their timestamps in the logged data entries.

The Leica GS10 receivers, one in each vehicle, will also be performing RTK positioning during the duration of the experiment with dedicated antennas and cellular data connections, their positioning results will act as the reference against which the OBU results will be compared to.

6.2.5. Cellular Environment

While C-V2V communication is not the primary focus of this research, it was deemed appropriate and beneficial to log several performance indicators from the cellular module for the duration of the test, including cellular signal quality, service type, latency, and data transmission speed. By collecting and analysing these metrics, they can be used to form an initial observation regarding the behaviour and feasibility of C-V2V communication systems, contributing to a more comprehensive understanding of the different methodologies available for facilitating V2V communication.

6.2.6. Data Processing and Visualisation

A critical aspect of this experiment is the processing and effective visualisation of the collected data, as it allows important insights to be identified and observed, and further actions to be planned as a result.

Python is used to perform the tasks outlined in this section. Initial preprocessing involves performing data cleaning to identify and address missing values, faulty data, and outliers. This is used to establish a credible starting point before further analysis and processing of the data.

Next, raw DSRC packets from both the transmitting and receiving vehicles are matched together based on the sequence numbers. Missing packets can be identified from unmatched sequence numbers. Existing KPIs such as received power level and background noise level will be incorporated into the dataset. New KPIs, including inter-packet gap, packet reception ratio, and signal-to-noise ratio, are calculated and added into the dataset as well. Data from certain part of the experiment, such as during the equipment setup and mid-point break, are removed so the unintended stationary period will not skew the distribution of related parameters.

The data are transformed into Comma Separated Values (CSV) format, where each column signifies an attribute, and each row represents the values at a specific moment in time. This standardised format ensures data interoperability in different applications, and enables ease of exploration in machine learning experiments in a later chapter.

Subsequent stages involve performing detailed statistical analyses on the dataset, as well as visualising the data using tools such as Matplotlib. Those steps aim to identify significant trends and correlations to ensuring accurate interpretation as well as the achievement of the experiment objectives.

6.3. Test Results

The outdoor experiment was carried out according to the test plan without complications thanks to the careful planning as well as the prior rehearsal. The acquired data was retrieved, processed, and formatted in accordance with the methods previously mentioned.

This section gives a thorough examination of the dataset by employing statistical analysis and visualisation techniques to explore relevant parameters. Additionally, notable trends and features identified are discussed and commented.

6.3.1. Driving Environments

The selected test route features a diverse blend of driving environments with different speed limits and road conditions. To ensure an accurate understanding of the OBU's performance, the data analysis will include both an overall assessment and examinations of three distinct driving scenarios when applicable. Namely:

- **Built-up Area**, as defined by the UK Highway Code, with the presence of streetlights at regular intervals and speed limit under 30 MPH [149]. This scenario is characterised by the dense infrastructure such as urban buildings and narrow residential streets can often lead to signal multipath and blockages.
- **Inter-Urban**. This scenario covers the roadway with a speed limit of 40 and 50 MPH, with a blend of open spaces and moderate urban structures. This environment presents a mix of challenges including variable signal availability and occasional interference.
- **High-Speed**. This scenario covers the roadway with a speed limit above 60MPH, and is primarily found on dual carriageways and open roads with minimal obstructions, offering a clear view of the sky, but can be a challenging environment for timely and accurate positioning results.

6.3.2. Dataset Overview

The processed dataset consists of 21 parameters and 47983 instances of data points. 13912 data points, or 29% of the test route, were of Built-up areas. 29801 data points, or 62% of the test route, were of Inter-Urban areas. 4720 data

points, or 9% of the test route, were of High-Speed driving. The parameters are listed and described in Table 5.

Table 5: Parameters of road trial dataset

Parameter Name	Unit	Description
Timestamp	-	Timestamp in HH:MM:SS
Seqnum	Integer	Sequence Number
PowerA	dBm	Received power and background noise level of antenna A and B on the OBU
PowerB		
NoiseA		
NoiseB		
SNRA	Integer	Signal-to-noise ratio of antenna A and B on the OBU
SNRB		
TX_OBU_LAT	Decimal Degrees	Latitude and longitude of OBU positioning result on both vehicles
TX_OBU_LON		
RX_OBU_LAT		
RX_OBU_LON		
TX_REF_LAT		Latitude and longitude of Leica reference receiver on both vehicles
TX_REF_LON		
RX_REF_LAT		
RX_REF_LON		
TX_OBU_speed	MPH	Speed of both vehicles
RX_OBU_speed		
Distance	Meters	Distance between two vehicles at this moment in time
IPG	Seconds	Inter-packet gap between transmission and reception of this packet
PER	Percentage	Packet Error Rate
Road_type	String	One of the three driving scenarios
IS_DROPPED	Boolean	True if this packet is not received by the receiver. False otherwise.

6.3.3. Signal Quality

A strong wireless signal is essential for ensuring robust and stable V2V communication. Unlike the previous test performed in near-ideal conditions, where the transmitter remained stationary and the receiver travelled in a linear path at a consistent speed in open-air conditions, this trial involved both the transmitter and receiver in motion at varying speeds, with the distance between

the two vehicles changing throughout the route. Moreover, the physical environments along the route, characterised by the presence of nearby buildings, foliage, curves in the road that breaks line-of-sight, and possible obstruction of other vehicles, may further affect the signal strength and communication range.

This section thus aims to provide an examination of the signal quality data, the correlations to other parameters, and how these factors influence communication effectiveness under real-world conditions.

The variations in the following distance and vehicle speed can notably influence signal quality, therefore it is logical to begin the investigation with these two key parameters. statistical analysis is shown in Table 6 and Table 7.

Table 6: Statistics of vehicle following distance (meters)

	All	Built-up Areas	Inter-Urban	High-Speed
Mean	54.89	32.99	54.99	131.43
Median	35.26	30.72	37.27	49.07
Max	689.20	138.36	477.29	689.20
Min	4.52	4.52	7.42	14.6
Range	684.68	138.84	469.87	674.60
SD	71.03	19.74	54.23	171.44

Table 7: Statistics of vehicle Speed (MPH)

	All	Built-up Areas	Inter-Urban	High-Speed
Mean	22.63	17.14	23.61	36.20
Median	24.30	20.34	27.11	45.44
Max	72.40	31.38	53.15	72.40
Min	0	0	0	0
Range	72.40	31.38	53.15	72.40
SD	15.14	10.52	14.04	23.53

Over the entire test route, the two vehicles maintained an average proximity of 55 meters, with a median distance of 35 meters. The maximum separation that still allowed DSRC communication was recorded at 689 meters. In built-up areas, the mean and median following distances were notably shorter, approximately

30 meters, indicative of closer spacing in low-speed environments. The maximum distance achieved in Built-up Areas was 138 meters. The standard deviation is also the lowest of all scenarios, suggesting a more consistent following distance in urban areas. In Inter-Urban areas, mean and median following distance saw a moderate increase, and a maximum distance of 477 meters was achieved. High-Speed scenario exhibited the largest amount of vehicle separation, average distance, and the range of the following distance, aligning with the expectation of longer following distances at higher speeds for enhanced safety and response time.

The vehicle speed during the experiment exhibited similar variations in different scenarios, spanning a wide range from stationary to 72MPH, reflecting the typical driving speeds in UK. The average speed overall was around 22MPH, with Built-up Areas showing the lowest median and average speed, but also the lowest variation. High-speed environments demonstrated the highest mean and median vehicle speed and the overall top speed, with the largest amount of standard deviation.

With the diverse span of vehicle speed and following distance, Figure 26 to Figure 28 show the scatter plots of the received power level, noise level, and signal-to-noise ratio for both antenna A and B in dBm, plotted against the distance between the transmitter and receiver in meters. The figures also include a red line indicating the running window average of 20 most recent data points, providing a visual aid for understanding the average trends in the data.

Similar to what was observed in the preliminary experiment, as the distance between the transmitter and receiver increases, the received power decreases, with the communication interruption occurring at around -95dBm. The received noise level also increases with distance, demonstrating a trend similar to logarithmic growth, which plateaus at around -101 dBm once the distance between the two vehicles exceeds 150 meters.

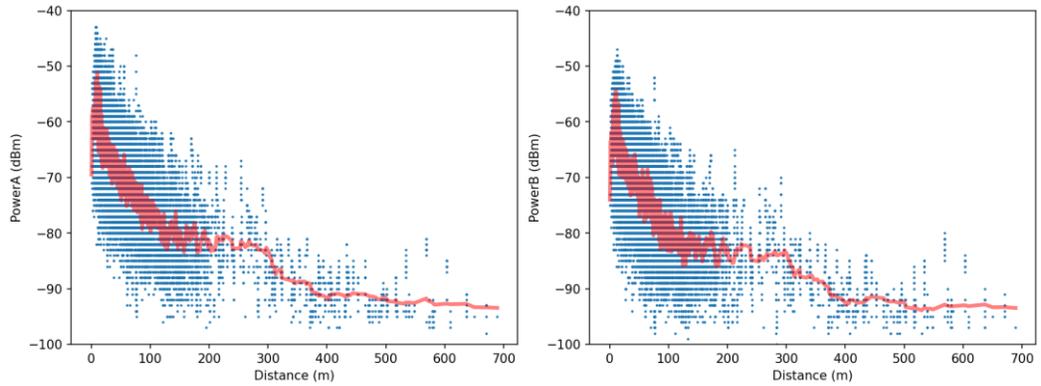


Figure 26: Road trial received power level

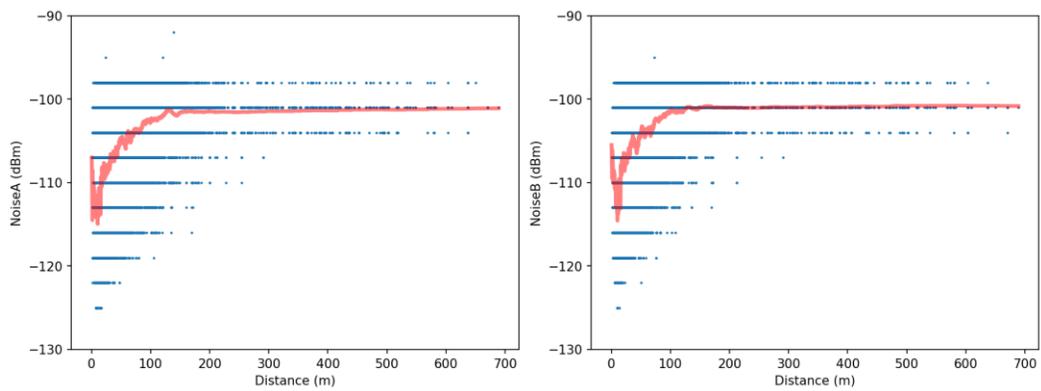


Figure 27: Road trial received noise level

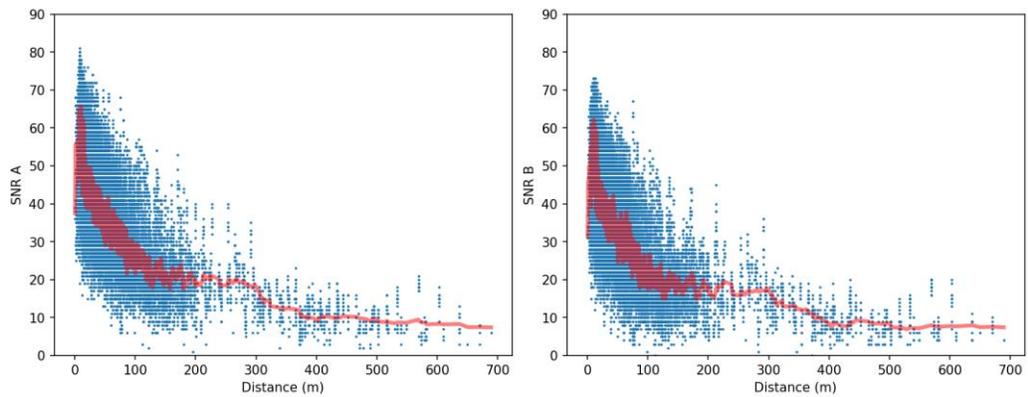


Figure 28: Road trial signal-to-noise ratio

The statistics of signal quality parameters, both overall and in each driving environment, are shown from Table 8 to Table 11.

The Built-up Area showcased the most optimal signal quality among the tested scenarios, with the strongest average received signal power level, lowest average received noise level, and lowest standard deviation, resulting in a strong average SNR of approximately 46 with no observed communication interruptions. The superior performance can be attributed to slower driving speeds and closer vehicular proximity.

Conversely, the inter-urban areas displayed a marginal reduction in signal quality. The average received power was slightly lower, with slightly higher noise levels and standard deviation. The average SNR was around 40, with no communication disruptions.

The high-speed driving environments mirrored the inter-urban areas in terms of average received power (-68dBm) and SNR (around 40). However, a notable increase in the standard deviation was observed, indicating a greater variability in signal quality. This inconsistency is likely due to factors such as higher driving speeds, increased vehicle separation, and the intermittent presence of other vehicles disrupting the line of sight. Notably, this scenario was the only one to experience a brief communication interruption from vehicles exceeding the communication range.

Comparison between the two antennas revealed trends similar to the findings in the preliminary test. Antenna A outperformed Antenna B across all scenarios, demonstrating slightly higher average received power and lower noise levels, yielding a marginally higher average SNR. However, the differences were less pronounced in this experiment compared to the earlier test. Antenna B also showed slightly less variance and standard deviation in its performance, suggesting a more stable but slightly less robust communication capability.

Table 8: Road trial signal quality statistics: Overall

	Power (dBm)		Noise (dBm)		SNR	
	ANT A	ANT B	ANT A	ANT B	ANT A	ANT B
Mean	-66	-69	-108	-107	41	38
Median	-66	-68	-107	-104	41	37
Max	-43	-47	-83	-73	81	73
Min	-98	-100	-125	-125	1	-3
Range	55	53	42	52	80	76
SD	9.36	9.34	6.40	6.07	14.73	14.49

Table 9: Road trial signal quality statistics: Built-up Areas

	Power (dBm)		Noise (dBm)		SNR	
	ANT A	ANT B	ANT A	ANT B	ANT A	ANT B
Mean	-64	-66	-109	-108	46	43
Median	-64	-66	-110	-107	45	41
Max	-45	-47	-86	-73	75	73
Min	-95	-98	-125	-122	6	-2
Range	50	51	39	49	69	75
SD	7.16	7.36	6.11	60.4	12.14	12.51

Table 10: Road trial signal quality statistics: Inter-Urban

	Power (dBm)		Noise (dBm)		SNR	
	ANT A	ANT B	ANT A	ANT B	ANT A	ANT B
Mean	-68	-70	-107	-106	40	36
Median	-68	-69	-104	-104	38	36
Max	-43	-50	-83	-74	81	73
Min	-97	-98	-125	-125	2	-3
Range	54	48	42	51	79	78
SD	9.7	9.6	6.45	6.07	15.27	14.81

Table 11: Road trial signal quality statistics: High-Speed

	Power (dBm)		Noise (dBm)		SNR	
	ANT A	ANT B	ANT A	ANT B	ANT A	ANT B
Mean	-68	-73	-108	-105	40	32
Median	-65	-71	-107	-104	43	34
Max	-46	-49	-95	-98	76	72
Min	-98	-100	-122	-122	1	0
Range	52	51	27	24	75	72
SD	11.01	10.22	6.16	5.11	15.91	14.28

The main factor contributing to the observed signal degradation can be attributed to free-space path loss (FSPL), referring to the reduction in signal strength when it is transmitted over relatively long distances, often hundreds of meters or more [150]. As the signal travels through the transmission medium, the energy per unit area diminishes, leading to a decrease in the amount of energy that ultimately arrives at the receiver, and therefore the reduced signal strength.

For ideal isotropic antennas that radiate energy uniformly in all directions, the equation for FSPL can be expressed as:

$$FSPL = \frac{P_t}{P_r} = \left(\frac{4\pi df}{c} \right)^2$$

Where:

- P_t and P_r are the transmitted and received radio power, respectively.
- f is the signal frequency in Hertz.
- d is the distance from the transmitter.
- c is the speed of light in vacuum.

The formula suggests that the attenuation of the signal strength is proportional both to the square of the distance between the transmitter and the receiver and to the square of the signal's frequency. As a result, when the distance or frequency doubles, the signal strength experiences a four-fold loss. This inverse-

square relationship can be observed evidently from the visualisations. The further the signal needs to travel and the higher the frequency it carries, the more significant the reduction in signal strength will result.

The effect of FSPL can be mitigated through several strategies. One approach is to simply increase the transmission power, which would ensure a higher signal power reaching the receiver. Another approach can be enlarging the receiver antenna's capture area, which would capture more signal and consequently increase the signal strength [151]. Utilising directional antennas also serves as an effective solution, as they concentrate the radio energy into a tighter, more focused beam on the transmitter, and allow less unwanted noise and interference to be picked up from the receiver [152]. However, this method requires careful alignment of the antennas, which may add to the complexity of the device.

While the overall relationship between signal quality and distance can be attributed through the FSPL theory, an examination of the received power level visualisation reveals a considerable dispersion of data points on either side of the red trend line, especially for distances under 200 meters. At a given distance between the two vehicles, a substantial variation in received power levels can be found, often reaching a difference of up to 30dBm. This observation indicates that there might be additional influences on the amount of power that ultimately reaches the receiver, beyond the primary effects FSPL.

A potential explanation for this phenomenon could be the existence of reflectors and scatterers along the path of the radio signal, including flat-sided buildings and other vehicles. Such obstacles can exert a detrimental effect on signal energy as the original signal deflects and scatters off the objects. This results in multiple copies of the transmitted signal, each with random amplitude attenuation and phase offsets, reaching the receiving antenna at slightly differing times. These multipath components may combine constructively or destructively at the

receiver, contributing to signal distortion and leading to signal strength fluctuations. This fluctuation in received signal strength over short distances and time periods is known as small-scale fading [153]. Additionally, in 802.11-based wireless radios, multipath propagation often extends the time required for the signal's baseband component to reach the receiver, potentially resulting in signal smearing due to intersymbol interference (ISI), in which the energy of a symbol spills over into succeeding symbols [154].

Techniques such as antenna diversity can be employed to mitigate the errors and distortions induced by small-scale fading. The most prevalent approach involves the use of multiple antennas at the receiving end. The presence of multiple antennas offers the receiver multiple views of the same signal, with each antenna being subjected to a distinct interference environment [155].

Consequently, even if one antenna undergoes significant signal degradation, there would be a high probability that another antenna is receiving a sufficient signal. Taken together, such a system can facilitate a more resilient wireless communication link. However, it is important to note that for optimal functionality, a physical separation of a few wavelengths is required between each individual antenna.

An additional phenomenon known as shadowing can contribute to further degradation of signal quality during transmission. Shadowing is characterised by fluctuations in the received signal power due to the presence of objects obstructing the propagation path between the transmitter and receiver, often breaking their line-of-sight (LOS) [156]. When radio waves encounter these obstructions, part of the signal power is absorbed and blocked, further attenuating the strength of the received signal.

The extent of signal shadowing is influenced by multiple factors, including the frequency of the radio wave and the material composition of the intervening objects [157]. In general, the penetrative power of a radio wave is inversely

proportional to its frequency. As such, lower-frequency signals typically exhibit a better ability to pass through objects with minimal power loss compared to their higher-frequency counterparts, although at the cost of lower data rate.

In the context of V2V communication, signal shadowing is almost inevitable due to elements such as blind corners, other vehicles, and foliage. Nevertheless, strategies can be employed to mitigate its impact, such as installing the antenna in an ideal location, preferably at a higher point on the vehicle's roof, can maximise the possibility of maintaining line of sight. This, when combined with other mitigation methods previously mentioned, can contribute to achieving the best possible signal quality and ensuring stable and reliable communication over wireless links.

6.3.4. Communication Range

Another important performance metric of the OBU is the communication range. Although the experiment recorded a maximum distance of data exchange of approximately 684 meters, it is important to note that the service quality at this distance is likely to be substantially degraded, with considerable packet loss, elevated latency, and frequent disruptions. Therefore, further examination of the collected data is necessary to establish a more reasonable usable communication range where a stable wireless connection can be consistently sustained.

Inter-Packet Gap (IPG) and Packet Error Rate (PER) are two such parameters that can provide valuable insights into the functionality and stability of a wireless network, and can be used to determine the maximum usable range of communication distance between a pair of wireless transmitters and receivers [158].

Packet Error Rate is a metric that quantifies the percentage of transmitted data packets that are either lost or contain errors, compared to the total number of packets sent. The PER is calculated in this experiment by comparing the

sequence number of transmitted and received packets, where unmatched sequence numbers indicate missed packets.

As the distance between the transmitter and receiver increases, the likelihood of encountering errors in packet transmission also increases, due to factors such as FSPL, multipath propagation, and signal shadowing. A high PER may result in higher latency, reduced throughput, and increased risk of communication failures. At some point the PER may increase to an unacceptable level, and the distance at which this maximum acceptable PER is reached can be considered the maximum reliable communication range of the wireless link.

The threshold for PER that determines the maximum communication range is dependent on various factors such as the application requirements, network configuration, and environment. However, a common practice for 802.11-based wireless network is to consider a PER of less than or equal to 10% to be a usable range of the wireless link. At this point, the communication link can still be considered stable enough for many applications without requiring excessive retransmissions and delays. Of course, the threshold can vary depending on the specific needs and tolerance of the system or application. Some mission-critical or real-time applications might require a lower PER, sometimes at or less than 1%, whereas others might be able to tolerate a higher PER.

Inter-Packet Gap, referring to the time interval between the reception of two consecutive packets in a communication system, can also give insights into the communication range. A consistent IPG usually indicate normal operation, whereas high or fluctuating IPG values may suggest issues such as interference, signal loss, or potential communication failures. Similar to PER, a certain threshold of IPG can be set based on the system's performance requirements, and the distance at which this IPG is reached would give an estimate of the maximum communication range.

In this experiment, the internal real-time clock (RTC) on the OBU is synchronised to the GNSS time once a valid position fix has been obtained, ensuring both OBUs are operating on a synchronised and accurate timebase. The DSRC message is broadcast 10 times per second, or every 100 milliseconds. The transmission and reception timestamp are logged on each OBU's internal storage. These two timestamps are then compared to derive the IPG for that specific packet.

The relationship between following distance and IPG and PER is shown in Figure 29.

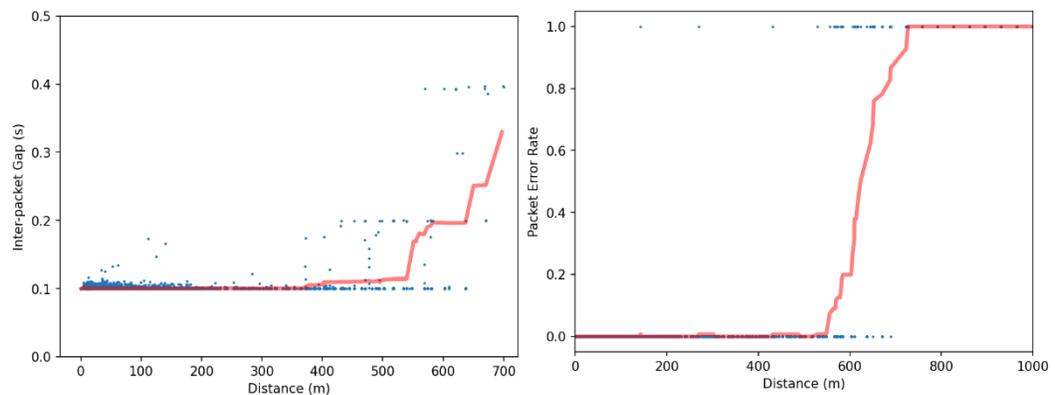


Figure 29 : Road trial IPG and PER

From the visualisation, it can be observed that the PER remains mostly constant near 0% as the distances between the two vehicles increase to approximately 550 meters. Beyond this point the PER starts to increase rapidly, until wireless link between the two vehicle was lost at around 680 meters, at which point PER reaches and remains at 100%. The maximum usable communication distance of the OBU during this test, using a threshold of 10% PER, is around 571 meters, while using the more stringent 1% PER threshold yields the maximum usable distance around 484 meters.

Similarly, the IPG remains predominantly stable at around 100 milliseconds when the vehicles are within a range of 400 meters of each other. As the distance exceeds 400 meters, a gradual increase in the IPG reaching about 110 milliseconds can be observed. The increase in IPG becomes significantly more

pronounced when the distance between the vehicles extends beyond 534 meters. Once the vehicles are approximately 680 meters apart, communication is completely lost, causing the IPG to exceed the y-axis scale of the graph.

It is intuitive to observe that the IPG and PER are closely related with the received signal quality at the receiver end, as a weaker signal may result in missed packets or necessitate re-transmission due to errors in the received packets. The signal quality, in turn, is heavily influenced by the distance between the wireless transmitter and receiver as well as factors such as signal shadowing and multipath interference. Nonetheless, IPG and PER can serve as useful parameters to determine the usable communication range between two wireless radios, with appropriate thresholds based on specific applications. For this test, a usable communication distance of approximately 550 meters was established using a 10% PER threshold.

This distance is notably shorter than the maximum communication range of 680 meters achieved in this test, and substantially less than the nearly 900-meter range reached in the prior test conducted in Wollaton Park. The discrepancy can be accounted for by several contributing factors.

First, the maximum communication range represents a state of barely functional connection between the transmitter and receiver, with considerable packet loss, high latency, and frequent interruptions, and very little chance of sustained and reliable data exchange. The usable range determined by the PER threshold, on the other hand, allows for a more stable data exchange with a reduced risk of errors.

Moreover, the real-world road test introduces many imperfections in addition to the FSPL that contribute to signal degradation, such as previously discussed multipath interference and signal shadowing, as well as potential interference from 5GHz Wi-Fi devices operating at nearby frequencies.

Finally, the positioning of the antennas may have had an effect. In this experiment, the antennas were installed on the OBU inside the vehicles. More optimal positioning, such as mounting on the roof of the vehicle, may have noticeably improved the results by increasing the likelihood of line-of-sight communication. However, appropriate antenna extension cables were not available during the test, and there were safety concerns about the roof-mounted antennas potentially detaching, which could have compromised the test results and posed a safety risk.

6.3.5. GNSS Positioning Accuracy

One of the major objectives of this experiment was to leverage the capability of DSRC to transmit and broadcast GNSS correction data to nearby vehicles. This would allow improved positioning accuracy even for vehicles lacking network connectivity of their own. The tasks involved in this objective were successfully executed during the experiment, and this section presents an examination of the collected data to evaluate its impact on the positioning accuracy.

Highly accurate Leica GS10 receivers were installed on both the transmitting and receiving vehicles, and produced their own Real-Time Kinematic (RTK) positioning outputs, independent of the operations of the OBUs [159], and the positioning results from the Leica receivers serves as a reference to compare the results obtained from the OBUs.

The transmitting OBU broadcasts the real-time GNSS correction data stream as part of the payload via DSRC. Upon reception, the correction data, along with the raw measurement data from the GNSS receiver, was processed through the RTKLIB software suite to perform real-time RTK positioning on the receiver, and the results were subsequently stored.

Consequently, at the conclusion of the experiment, three sets of GNSS positioning data were available from the receiving vehicle:

- Reference result from the Leica receiver.
- OBU single-point positioning result without any aiding.
- OBU RTK positioning result using correction data stream obtained via DSRC.

To enable valid comparison of the GNSS performance, the positioning results from all three sources are combined into a single dataset, synchronised by the timestamp. The performance of the RTK-over-DSRC on the OBU will be examined in three distinct driving environments along the test route, each offering unique challenges and conditions.

Three GNSS solution types are encountered in the dataset: RTK Fixed, RTK Float, and Single. RTK Fixed solutions provides the most accurate location data by resolving the integer ambiguities in order to determine the exact number of wavelengths in the carrier phase signal from the satellite. Centimetre-level accuracy can be achieved with RTK Fixed solution under ideal conditions. In contrast, with RTK Float, the integer ambiguities have not been fully resolved, resulting in a less accurate position solution compared to RTK Fixed. However, RTK Float can still offer notable improvements compared to unaided GNSS solutions, and can achieve sub-meter accuracy in normal conditions [160]. Finally, the Single solution is the simplest form of GNSS positioning, relying on the raw pseudorange measurements without any form of differential correction or ambiguity resolution, and has the lowest level of accuracy of the three solution types, in the range of several meters.

The visualisation of the GNSS solution quality is shown in Figure 30. It can be seen that the OBU on the following vehicle is able to utilise the GNSS correction data received via DSRC to perform real-time RTK for the majority of the test route, with either RTK Fixed or RTK Float solutions. The GNSS solution also reverted to Single for a brief period of time due the lack of correction data

stream from an interruption of DSRC communication near the end of the test route.

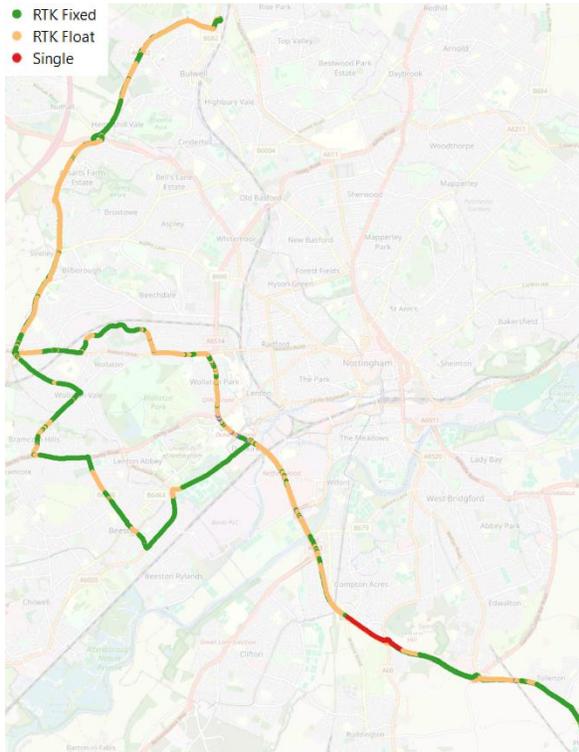


Figure 30: OBU GNSS Solution Type

A breakdown of OBU GNSS solution types over different driving scenarios is shown in Table 12. It can be observed that the OBU successfully achieved RTK solution using data transmitted over DSRC for 98.5% of the test route, with 54% RTK Fixed, 44% RTK float, and 1.5% single solutions. Looking deeper into each category, Built-up Areas with the lowest speed limit achieved 71.75% RTK Fixed and 28.25% RTK Float solutions. On the contrary, Inter-Urban scenario resulted in more RTK Float solutions compared to RTK Fixed, while high-speed driving environment demonstrated roughly equal amount of RTK Fixed and RTK Float solutions, with a small number of single solutions due to the temporary loss of GNSS correction data, which was the only occurrence during the entire trial.

Table 12: OBU GNSS Solution Types

	Built-up Area	Inter-Urban	High-Speed	All
RTK Fixed	71.75%	45.32%	46.34%	54.20%
RTK Float	28.25%	54.67%	46.24%	44.30%
Single	0%	0%	7.60%	1.51%

To further evaluate the benefit of RTK-over-DSRC in improving the positioning accuracy of the OBU, the geodesic distance between the reference GNSS receiver and the OBU RTK solutions, as well as the geodesic distance between the reference GNSS receiver and OBU Single solutions without any aiding, are calculated for each epoch and shown in Figure 31.

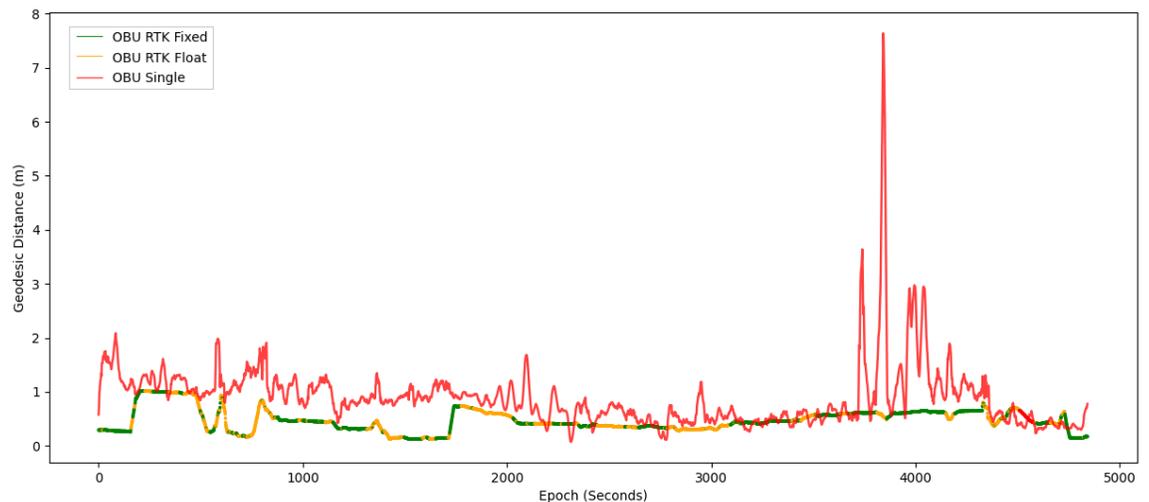


Figure 31: OBU GNSS Performance Comparison.

Observation of the graph reveals notable improvements in positioning accuracy with RTK-over-DSRC compared to unassisted single solutions under the same experimental conditions. The RTK solution demonstrated superior accuracy and reduced fluctuation and noise, which is particularly evident in the significant reduction in the deviation of up to 7 meters near the end of the experiment.

Table 13 shows the statistic comparison between Single solutions and RTK-over-DSRC over the entire test route. It can be seen that RTK significantly reduced the Root Mean Square Error (RMSE) by more than half, from 1.18 to 0.52 meters. The

maximum error and error range also witnessed significant reduction of around 65%. Standard deviation also saw a notable 73% improvement, indicating tighter positioning results with less variation and closer clustering around the mean value.

Table 13: OBU GNSS performance statistics, overall.

	Single (m)	RTK via DSRC (m)	Percent Change
RMSE	1.18	0.52	56%
Mean	0.89	0.47	47%
Median	0.77	0.43	44%
Max	7.63	2.73	64%
Min	0.12	0.10	17%
Range	7.51	2.6	65%
StdDev	0.77	0.21	73%

The statistics are further investigated in each of the driving scenarios, and the comparisons are shown from Table 14 to Table 16.

Table 14: OBU GNSS performance statistics, Built-up Area.

	Single (m)	RTK via DSRC (m)	Percent Change
RMSE	1.57	0.51	68%
Mean	1.07	0.49	54%
Median	0.78	0.48	38%
Max	7.63	1.10	86%
Min	0.12	0.11	8%
Range	7.51	0.99	87%
StdDev	1.15	0.13	89%

Table 15: OBU GNSS performance statistics, Inter-Urban.

	Single (m)	RTK via DSRC (m)	Percent Change
RMSE	0.91	0.44	52%
Mean	0.79	0.39	51%
Median	0.75	0.36	52%
Max	6.65	2.73	59%
Min	0.13	0.10	23%
Range	6.52	2.63	60%
StdDev	0.45	0.18	60%

Table 16: OBU GNSS performance statistics, High-Speed.

	Single (m)	RTK via DSRC (m)	Percent Change
RMSE	0.96	0.67	30%
Mean	0.82	0.61	26%
Median	0.84	0.54	36%
Max	4.13	1.23	70%
Min	0.16	0.14	13%
Range	3.97	1.09	73%
StdDev	0.49	0.29	41%

It can be seen that RTK-over-DSRC notably improved the GNSS positioning performance across all three driving scenarios. However, the highest level of improvement occurred in Built-up Areas, characterised by low speed limit, narrow streets, and dense buildings. The RMSE decreased by more than 1 meter, an almost 70% improvement. The maximum error and error range also saw 85% improvement, as well as an almost 90% reduction in Standard Deviation, signifying significant reduction in noise and variation in the positioning results. Indeed, the enhanced performance can be readily observed in the map visualisation, with Figure 32 demonstrating the improvements in the town centre of Beeston along the test route, where GNSS single solutions exhibited significant variations and drift in the presence of narrow lanes and close by buildings. In contrast, both RTK Fixed and RTK Float solutions resulted in a much more stable

trajectory, likely due to the improved mitigation of multipath signals with the use of more advanced algorithm and additional correction data.

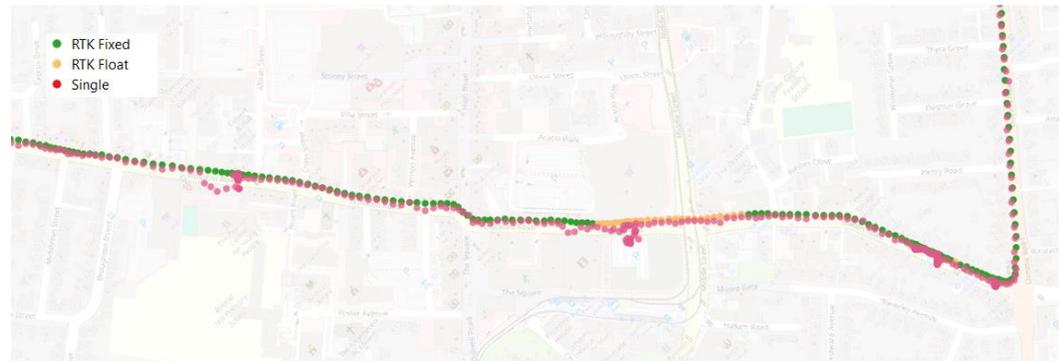


Figure 32: Accuracy improvements in Built-up Areas

RTK-over-DSRC also brought about improvements in Inter-Urban environments, however, the percentage change is not as significant as in Built-up Areas, with around 50% reduction in RMSE and 60% reduction in error range and standard deviation. Finally, the lowest amount of accuracy enhancement occurred during high-speed driving scenarios, with only around 30% of improvements in RMSE and mean errors.

Among the three scenarios, both RTK and Single solutions exhibited smallest RMSE error in Inter-Urban environments, with moderate speed limit and access to open sky. Single solutions fared the worst in Built-up Areas, likely due to multipath interference, while RTK solutions performed worst in High-Speed driving, possibly due to the high-dynamic environments and the large amount of processing required.

However, despite the enhancements in overall performance, there is still room for improvement. While the RTK results show an overall improved performance compared to unaided positioning, the mean error remains at approximately 0.5m away from the reference receiver. Moreover, there were instances where the RTK result exhibited more error than the unassisted positioning. Fluctuations in the data also remain noticeable even after the implementation of RTK processing.

Several factors have been identified that could contribute to the performance deficit between the OBU and the reference receiver. The most obvious factor is the placement of the GNSS antennas of OBU and the Leica receiver. The spatial separation of approximately 40 centimetres between these antennas may have played a role in the observed discrepancies in positioning.

Another factor might be the multipath errors where the GNSS signal bounces off nearby objects and arrives at a slightly different time, amplitude, and phase compared to the original signal, distorting the original signal, and reducing the accuracy [161]. This is similar to the small-scale fading that may have caused similar reduction in quality in DSRC signals.

Additionally, physical obstructions such as buildings and foliage can attenuate the signals reaching the GNSS antenna, as well as reduce the number of satellites in the receiver's field of view. This not only degrades signal quality but also compromises the satellite-receiver geometry, contributing to the reduction of the accuracy.

However, possibly the most important factor might be the quality of the GNSS antenna. A typical ceramic puck antenna was used with the OBU [162], whereas a professional-grade Leica AS10 antenna was used for reference receiver. High-quality antennas usually have a superior gain, reduced noise, and less cable loss due to improved construction, amplifiers, and shielding [163]. This improves the sensitivity and allows them to pick up weaker signals more effectively and acquire signals from a higher number of satellites. Such antennas also often have improved multipath rejection capabilities, as they are designed to receive signals coming directly from the satellites overhead and reject signals coming from other angles, which are likely to be the multipath signals. Compatibility with both L1 and L5 GNSS satellite bands further enhances performance by mitigating sources of inaccuracies such as ionospheric delay, and providing more signal availability and continuity [164].

Another promising avenue for improvement lies in upgrading from the single-frequency GNSS receiver to a multi-frequency GNSS receiver, which allows better correction of ionospheric delays, better distinction between direct and reflected signals, and increased signal robustness and satellite availability [165].

The quality of signal processing algorithms can also influence performance. These algorithms aid in tracking satellite signals, mitigating multipath errors, correcting ionospheric delays, resolving carrier phase integer ambiguity, and modelling potential sources of error. Different manufacturers often employ proprietary algorithms tailored to their specific products, and there might have been room for fine-tuning the RTKLIB algorithms to optimise and improve its performance for this particular setup.

In summary, an analysis of the visualisations and statistics indicates that the OBU is capable of receiving GNSS correction data over DSRC and perform real-time RTK processing to significantly improve its positioning accuracy compared to unaided single point solutions, and broadcasting correction data can indeed be an effective method of improving CAV positioning performance in real-world environments. However, the performance can potentially be further improved by utilising a higher-quality antenna and fine-tuning the open-source signal processing algorithms.

6.3.6. Cellular Environment

Cellular Vehicle-to-Everything (C-V2X) is a similar technology that allows communications between vehicles and other entities. C-V2X can operate in Device-to-Device mode, allowing direct communication without the need for cellular infrastructure, similar to DSRC, or in Device-to-Network mode, allowing communication via cellular network to reach out-of-range vehicles and accessing cloud-based services [29].

Advantages of C-V2X include proposed longer communication range, network scalability, non-line-of-sight performance, and value-added cloud-based services.

Although it also faces challenges such as infrastructure requirements, cost of device and deployment, and doubts about the maturity of the technology.

While it was not possible to test C-V2X Device-to-Device direct communication mode in this experiment due to hardware incompatibility, the onboard cellular modem was able to collect potentially useful data points for evaluating the performance for Device-to-Network mode. Collected parameters include cellular service type, signal quality, latency, and data transmission speed.

While C-V2X is not the main focus of this research, those data points may help form an initial observation regarding the behaviour and feasibility of C-V2V communication systems. The visualisation for cellular signal type and signal quality is shown in Figure 33 and Figure 34, and corresponding statistics are shown in Table 17 and Table 18.

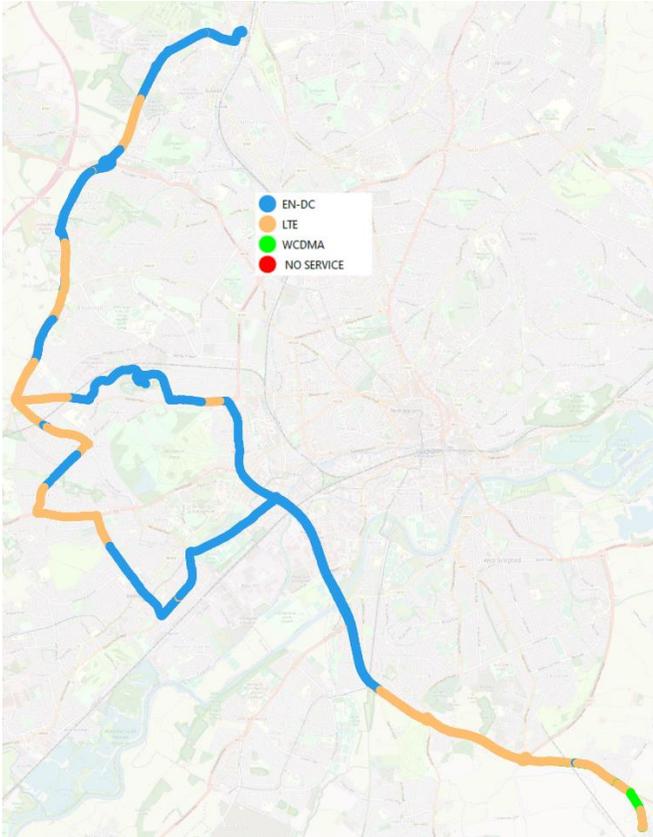


Figure 33: Road Trial Cellular Service Type

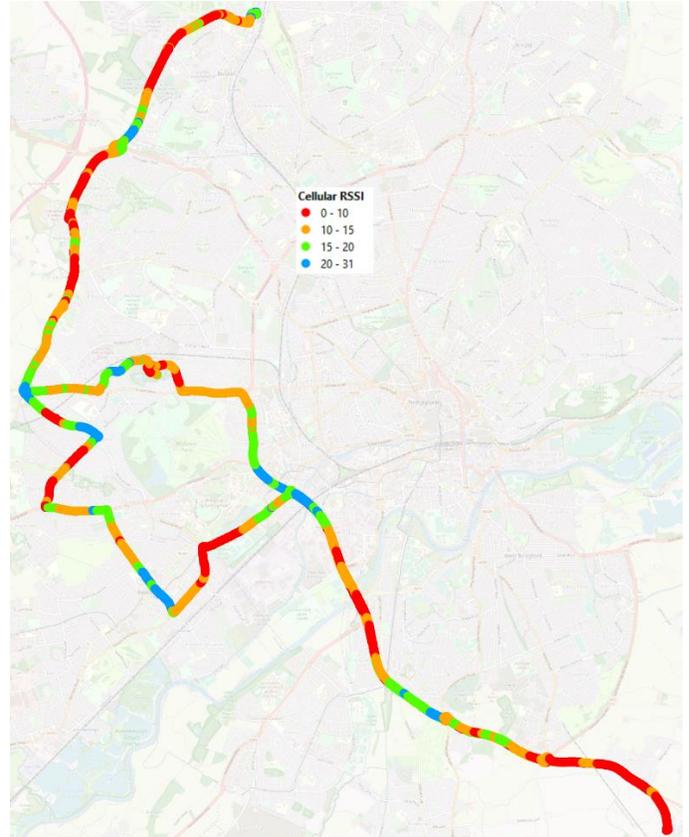


Figure 34: Road Trial Cellular Signal Quality

Table 17: Cellular Service Type Statistics

Service Type	Count	Percentage	Mean Latency (ms)	Mean DL (Mbit/s)	Mean UL (Mbit/S)
EN-DC	2953	60.5%	21	188.7	56.8
LTE	1748	35.8%	28	49.2	23.1
WCDMA	180	3.7%	48	9.01	0.51
No Service	3	0.06%	-	-	-
Total	4884	100%	-	-	-

Table 18: Cellular Signal Quality Statistics

	Cellular Signal Quality
Mean	13.80
Median	13
Max	31
Min	0
Range	31
SD	6.18

Four categories of service type were recorded during the experiment. EN-DC, LTE, WCDMA, and NO SERVICE.

EN-DC, or E-UTRAN New Radio - Dual Connectivity, is a technology allowing cellular device to connect to a 4G network and 5G NR (New Radio) network at the same time, and utilise the bandwidth of both for data transmission [166]. This allows cellular providers to launch 5G services using existing 4G infrastructure as the backbone, also known as the non-standalone (NSA) architecture, which is the basis of most 5G deployments around the world.

LTE, or Long-Term Evolution, is commonly referred to as 4G cellular network, while WCDMA (Wideband Code Division Multiple Access) is used for 3G mobile networks.

Cellular Received Signal Strength Indicator (RSSI) provides an approximation of signal strength and reception quality. In the cellular modem used in this experiment, the RSSI value ranges from 0 to 31. A value of 0 corresponds to -113dBm of received signal power, while a value of 31 indicates -51 dBm [167]. Therefore, a higher RSSI value correspond to a stronger received signal, and may reduce the latency and increase data transfer rate.

The graph illustrates a diverse range of cellular services experienced throughout the test route. The dominant service was 5G, which was available for approximately 60.5% of the route. This was followed by 4G coverage, which was present for about 35.8% of the route. 3G-only service was available for a small portion (3.7%) of the route. Additionally, there was a brief span where no cellular service was detected at all.

Certain correlation between the type of cellular service available and the geographical location of the vehicle during the test can be observed from the map. Notably, 5G coverage appears to be predominantly located within or close to the city boundary, while 4G service becomes more prevalent when the vehicle

is outside the city limits. Furthermore, the cellular service tends to downgrade to 3G as the vehicle moves even further away into the rural areas.

Notable data transfer speed improvements between 3G, 4G, and 5G can be observed, although the network latency on 5G service was very similar to the latency values on 4G during this experiment. This might be explained by the use of EN-DC network from which the modem can make its own decisions to use 4G and 5G based on the network load demand. For instance, the modem may opt to employ the 5G network during speed tests to maximise data transfer speed, while during latency tests, it might have used the 4G network to optimise power usage.

The cellular signal quality also demonstrated notable variations during the experiment, covering the whole range of possible values, with a mean value of 13.8. Again, it can be observed that densely populated areas tend to be covered with stronger signals, while rural areas tend to have weaker signals.

It is important to note that the data collected from the cellular modem, and in turn, the derived observations, was not meant to be conclusive or comprehensive, as it was a secondary objective of the main experiment. The cellular network latency and speed can depend on a variety of additional factors that was not covered in this investigation, such as the user density in the area, network traffic levels, and the performance of the server that OBU was connected to. Moreover, the cellular infrastructure is also evolving with time, with ongoing upgrades and the construction of new cellular access points.

Despite the limitations, a few important insights can still be identified. It is clear that superior cellular service quality is generally found in urban and built-up areas, characterised by fast data transfer speeds and low latency. In contrast, rural regions often experience less satisfactory service, with older generation service types, weaker signal strength, and consequently diminished network throughput. For C-V2X technology to demonstrate robust performance in these

circumstances, a comprehensive and efficient cellular infrastructure would be very important, especially considering the fact that vehicles tend to travel at faster speeds compared to urban areas.

6.4. Summary

This chapter gives a detailed description of the second outdoor experiment using the custom OBUs developed for this project. The primary objective was to evaluate how DSRC communication performs under realistic conditions, specifically, on public roads with real vehicles.

Test plans and procedures were discussed in detail, including planning a longer test route with more diverse driving environments, organising logistics, collecting KPI data, broadcasting GNSS correction data stream via DSRC to enhance positioning accuracy, and a look at cellular environment to understand the potential for C-V2X communication as an alternative to DSRC.

The experiment was carried out without complications, the hardware proved to be functional and stable, and all planned objectives were achieved. The collected data was processed, visualised, analysed, and discussed to achieve a comprehensive understanding of the experiment's outcomes.

It was discovered that the received DSRC signal quality is largely inversely proportional to the square of the distance between the two vehicles, as outlined in FSPL and inverse-square law. However, unlike under ideal conditions, additional factors play a big role in further attenuating the signal quality, such as small-scale fading and signal shadowing. Potential mitigation methods were discussed.

The Inter-Packet Gap and Packet Error Rate during the experiment were examined to determine a realistic usable communication range between the two vehicles. With a 10% PER threshold, the communication range was approximately 550 meters, while the maximum range was approximately 680 meters. The

several contributing factors to the shortfall of communication range compared to the preliminary test were discussed.

GNSS positioning data from the reference receiver and OBU were compared, and it was evident that transmitting GNSS correction data over DSRC can indeed be an effective way to increase the positioning accuracy of surrounding vehicles. A number of avenues of possible improvements were identified that might lead to even better results.

Finally, cellular environment data collected along the test route was examined. It was found that urban and built-up areas typically exhibit higher cellular service quality, while rural areas often experience older generation service types, weaker signal strength, and diminished network throughput. As a result, for satisfactory C-V2X performance in rural areas, overhauls and upgrades to the infrastructure might be needed.

In conclusion, this outdoor experiment has demonstrated that the custom OBU is capable of operating in a stable and performant manner, and the collected data provided valuable insights into the real-world performance of V2V communications using DSRC technology. The findings from this experiment will be crucial in informing further research and development efforts in this field.

Chapter 7: Interruption Mitigation with Machine Learning

7.1. Overview

This chapter investigates how machine learning models can be utilised to effectively predict and manage potential interruptions in V2V communication.

Predicting the loss of radio-based V2V communication before it occurs, as opposed to waiting until the actual loss of data link, would allow a larger time window to take appropriate actions such as switching to alternative means of communication, including cellular or satellite-based connection. This, in turn, can minimise communication downtime, a factor especially crucial when dealing with safety-related situations.

Machine learning has proven to be an effective tool in V2V communication research. In [71], Darlan et al. utilised machine learning to determine whether a data packet can be delivered within a desired latency window in cellular V2X scenarios, achieving 85% reliability in predicting Quality of Service (QoS) levels two seconds in advance using a supervised learning approach with an autoregressive integrated moving average filter. Additionally, Qiyi He's study [73] introduced an Anomaly Detection Framework for CAV Cyber Security, employing datasets from multiple CAV attack scenarios to evaluate classifiers like Decision Tree and Naive Bayes, with the Decision Tree algorithm showing superior performance after optimizations to reduce ML model time complexity.

While those studies provided valuable insights, they are limited by their reliance on datasets derived from simulated models, which may not accurately reflect real-world scenarios. This chapter's ML experiments build upon these previous findings, employing real-world experimental data to enhance the validity and applicability of its approach, while also investigating additional models.

First, suitable data preprocessing techniques are identified, introduced, and applied to the KPI dataset from the outdoor experiment. Steps involved include

class variable creation, data cleaning, attribute binning, and class balancing, in order to ensure the optimal input for model training.

Evaluation metrics such as confusion matrix, accuracy, recall, precision, F1 score, and training time are also introduced. These metrics will serve as the basis for the performance comparison between the machine learning models.

Finally, the pre-processed dataset will be used to train a number of machine learning algorithms, and their performance are compared and discussed.

7.2. Dataset Preparation

Before utilising the collected data from the outdoor experiment to train machine learning models, it is essential to perform several pre-processing steps on the dataset, including removing unneeded parameters, creating class variables, and addressing imbalanced class instances. These steps ensure the training data is clean and well-structured to reduce the likelihood of errors and biases during the training process, improving the accuracy, efficiency, and performance of the model.

7.2.1. Class Variable Creation

In the context of classification models, a class variable with multiple possible labels needs to be created and assigned to each entry in the outdoor experiment dataset, in order to facilitate the training of the model to make accurate predictions.

Given the objective of predicting imminent communication interruption, three class labels are devised for the dataset:

- **Disconnected:** This label is assigned to the entries in the dataset where transmitted packets did not reach the receiver.
- **Marginal:** This label is used to denote the entries where the communication quality is notably degraded, and interruption might be likely. This label is assigned to the entries in the dataset where Packet

Error Rate (PER) exceeds 10%. This is the same PER threshold used to determine the useful communication range in the previous chapter.

- **Functional:** This label denotes a normal communication quality with a low risk of communication interruption, and is assigned to all entries in the dataset that do not correspond to either of the two previously mentioned labels.

The categorisations will be instrumental in training the ML models, as the prediction output will correspond to one of the defined labels, and the performance and accuracy of the models can be subsequently analysed.

7.2.2. Parameter Selection

As the original dataset contains a relatively large number of parameters, shown in Table 5, optimisations will be carried out to reduce the parameter count, in order to improve the model's predictive performance as well as the efficiency during the training phase.

Firstly, parameters that do not contribute to the predictive power of the model are removed, such as the sequence number and the positioning result from the Leica reference receiver. The coordinates from both OBU's on-board GNSS receiver are also removed, as the classification process should not be dependent on the location of the device. By eliminating irrelevant parameters, the models can focus on the more meaningful features, improving the accuracy and predictive power, as well as decreasing the computational cost.

Furthermore, as PER was used to determine one of the class labels, it will also be removed to avoid potential bias or overfitting, where the model exhibits good performance during training but fails to generalise on new and unseen data. With the PER removed, the models are encouraged to learn from a broader range of features, enhancing its predictive capacity and generalisability.

Finally, the Pearson product-moment correlation coefficients (PPMCC) between received power, received noise, and SNR parameters are calculated. PPMCC is a statistical measure that evaluates the strength and direction of the linear relationship between two variables, producing a value between -1 and 1 [168]. A value close to 1 indicates a strong positive relationship, meaning as one variable increases, so does the other. Conversely, a value close to -1 indicates a strong negative relationship, signifying that as one variable increases, the other decreases. A coefficient close to 0 indicates a weak or no linear relationship. The correlation values between signal quality parameters are shown in Table 19.

Table 19: Correlation coefficient between signal quality parameters.

	PowerA	PowerB	NoiseA	NoiseB	SNRA	SNRB
PowerA	1	-	-	-	-	-
PowerB	0.847	1	-	-	-	-
NoiseA	-0.739	-0.645	1	-	-	-
NoiseB	-0.664	-0.758	0.737	1	-	-
SNRA	0.956	0.819	-0.903	-0.742	1	-
SNRB	0.824	0.961	-0.725	-0.907	0.838	1

It can be observed that the received signal power on antenna A and B demonstrates strong positive correlation, with a PPMCC of 0.84. Similarly, the values of received noise and SNR also exhibit strong correlation between the two antennas. Furthermore, there is an almost perfect correlation between the received power and the SNR value, with a PPMCC of 0.956 for antenna A and 0.961 for antenna B. This is not surprising given that SNR is calculated as the difference between the received power value and the received noise value.

Visualisations of relationships between several signal quality parameters are shown in Figure 35. As the PPMCC value suggests, strong linear correlation between the parameters can be observed. The drop-off of the trend line in graph a and c reflects the observation that antenna B tends to exhibit lower average received power compared to antenna A during the two outdoor experiments.

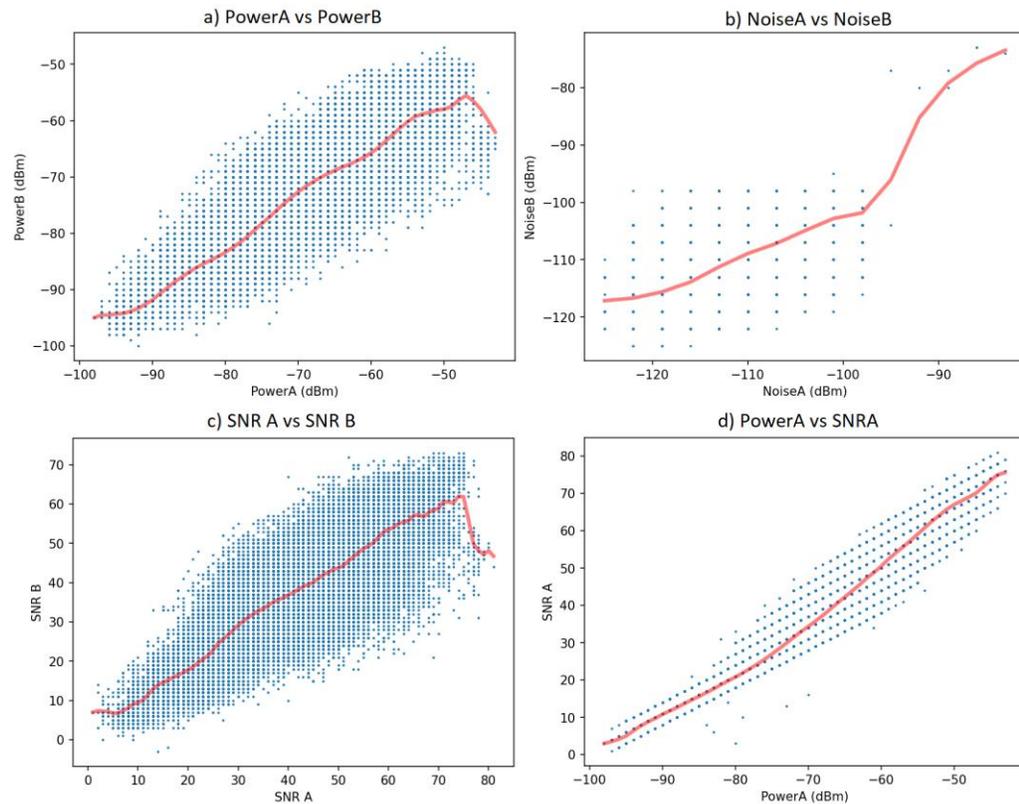


Figure 35: Visualisation of correlation between signal quality parameters.

Overall, strong correlations are observed over parameter pairs between antenna A and B, as well as between received power and SNR. Therefore, it would be beneficial to remove some of the parameters, as the redundant information may not improve the models' predictive power, and can even lead to negative effects including overfitting, difficult to interpret results, and unstable outputs [169]. By consolidating highly correlated parameters, the training models can be simplified with easier to understand results, improved training efficiency with reduced computational cost, memory footprint and faster training time.

Given the strong correlation observed between antenna A and B, the parameters from antenna B are excluded from the dataset. Additionally, considering that the

average noise level from the outdoor experiment was -108dBm compared to the average received power level of -66dBm, the received signal power overpowers the noise by a large margin, over 15848 times in this instance. Thus, the received noise levels for both antennas A and B are deemed insignificant and discarded as well. Finally, as the SNR was calculated by subtracting signal noise level from received power level and therefore demonstrate almost perfect correlation, it was determined that the SNR parameters will also not be used for training the models. Consequently, only the received power level from antenna A will be incorporated into the training dataset. This parameter effectively represents the information from all other signal quality indicators without adding unnecessary complexity to the dataset.

The updated training dataset parameters for the machine learning experiments are shown in Table 20, the number of parameters was reduced from 21 to 5, allowing more efficient training and reducing the chance of overfitting.

Table 20: Updated parameters of the training dataset.

Parameter Name	Unit	Description
PowerA	dBm	Received power on antenna A
TX_OBU_speed	MPH	Speed of both vehicles
RX_OBU_speed		
Distance	Meters	Distance between two vehicles
IPG	Seconds	Inter-packet Gap
IS_DROPPED	Integer	1 if this packet did not reach the receiver. 0 otherwise.
Class	Text	Functional Marginal Disconnected

7.2.3. Class Balancing

After determining the appropriate class labels, discarding irrelevant parameters, and consolidating strongly correlated parameters, the next pre-processing step involves examining the distribution of each class instances. In Figure 36, the test

route is coloured with class labels and overlaid on the map for an intuitive visual representation of the dataset.

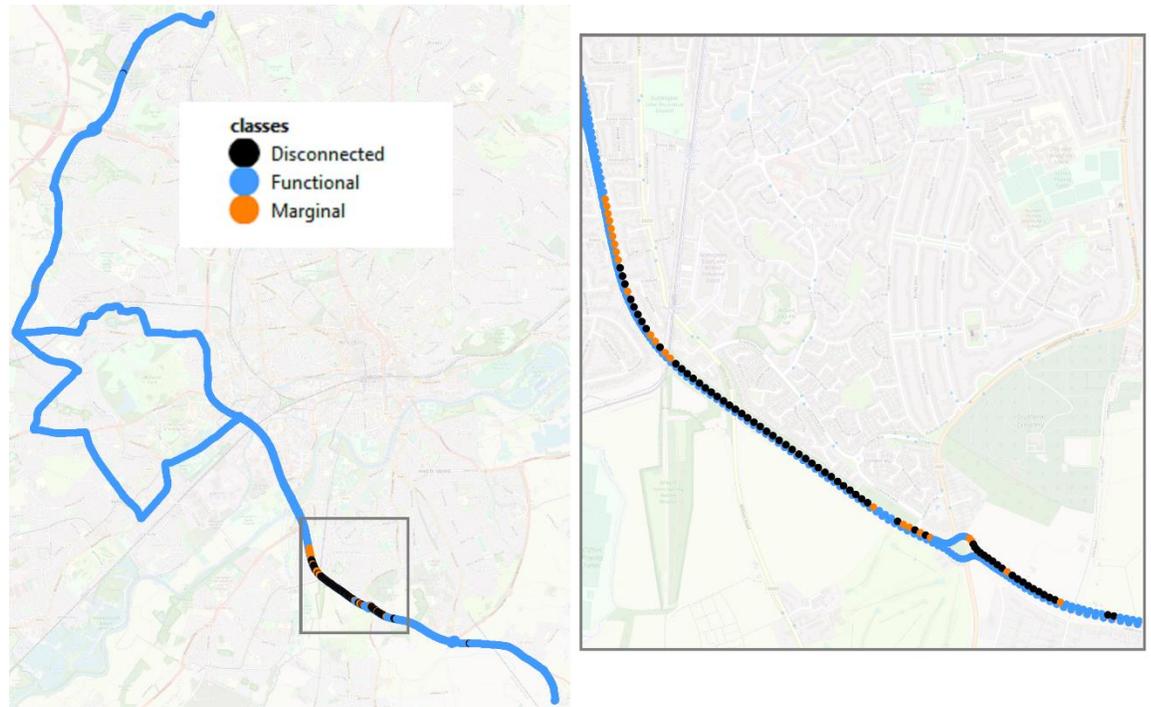


Figure 36: Class labels visualisation

It is apparent that for the majority of the test route, the communication quality is labelled as "functional", with stable connection and low risk of interruption. Most of the communication dropouts during the outdoor experiment occurred on a single stretch of dual carriage way with relatively high vehicle speed, extended following distance, as well as obstructing vehicles that may have further contributed to the signal degradation. It can be observed that the instances that are labelled as "Marginal" almost always precede and follow the instances that are labelled "Disconnected", reflecting the pattern of signal quality deterioration and recovery. This pattern makes the "Marginal" label an ideal predictor for imminent V2V communication interruption, suggesting its potential utility in making the decision to switch to alternative communication methods.

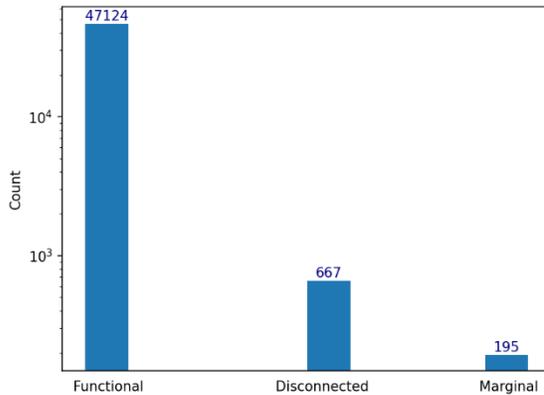


Table 21: Statistics of class labels

	Count	Percentage
Functional	47121	98.20%
Marginal	195	0.41%
Disconnected	667	1.39%
Total	47983	100%

Figure 37: Histogram of class labels

The histogram of class labels is shown in Figure 37, with the Y-axis in logarithmic scale. The statistics regarding the distribution of class labels are shown in Table 21. It is evident that the "Functional" class dominates the dataset, constituting over 98% of all entries. Conversely, the "Marginal" class has the lowest number of instances, with only 195 entries taking up 0.41% of the dataset, while the "Disconnected" class has 667 instances and comprises 1.39% of the training data.

It is apparent that dataset under examination presents a clear and pronounced imbalance across its three classes. This uneven distribution can notably affect the performance of machine learning models, as the algorithms are overexposed to the majority class during training, in turn making them more likely to predict the majority class, resulting in inflated accuracy and lower predictive performance for the underrepresented classes. A balanced dataset ensures a fair representation of all classes during the model training phase, thereby promoting the model's ability to generalise and predict each class accurately. Therefore, steps need to be taken to balance the class in this dataset [170].

There are a number of strategies available to address the issue of class imbalance in a dataset, each with their own strengths and limitations [171]. Under-Sampling is one of the easiest methods to alleviate class imbalance, which involves discarding instances from the majority class in an attempt to balance out the

class distribution. The primary advantage of under-sampling is its simplicity and speed. However, there is also a risk of discarding potentially useful data from the majority class, which could affect the model's performance.

Similarly, over-sampling can be used to increase the number of minority classes instead, which is often achieved by duplicating the instances from the minority class until a more balanced distribution is attained. Over-sampling can help to improve the model's performance on the minority class in a simple and computationally inexpensive manner, but it may also increase the risk of overfitting due to the repetition of instances.

Weighted Sampling is another method used to handle class imbalance, where different weights are assigned to each class based on their representation in the dataset. The minority class is usually assigned a higher weight, thereby encouraging the learning algorithm to pay more attention to it. This technique does not usually result in loss of data or overfitting, but careful tuning of the weights might be required.

Advanced techniques such as SMOTE (Synthetic Minority Over-sampling Technique) and ADASYN (Adaptive Synthetic Sampling) offer more sophisticated ways of handling class imbalance [172] [173]. Rather than simply duplicating instances, these methods generate synthetic instances of the minority class, adding more diversity to the data and helping the model to learn more complex patterns.

For this experiment, a combination of oversampling and undersampling will be used to balance the classes in the dataset, due to their straightforward nature to implement and compute. Additionally, those methods generally do not introduce additional hyperparameters to be fine-tuned, and the new dataset will be compatible with all the algorithms as before, unlike more complex techniques that introduce additional hyperparameters.

The majority class "Functional" will be undersampled, reducing its size via random discards until it matches the quantity of "Disconnected" instances. At the same time, the "Marginal" class will be oversampled, where its instances will be randomly duplicated until the total matches that of the "Disconnected" class as well. As a result, each of the three classes in the training dataset will possess an equal number of instances, specifically 667 each. This balanced distribution should result in a more equitable learning process, potentially improving model's performance and reduce the risk of bias towards the dominant class.

7.3. Evaluation Methods

To ensure a systematic and accurate assessment of the machine learning algorithms involved in this experiment, a number of commonly used evaluation metrics are discussed in this section. These metrics provide a quantitative understanding of how well an algorithm can predict or classify data points, and offer insights into the strengths and weaknesses of each model, aiding the decision-making in selecting the most optimal model for this experiment.

7.3.1. Prediction Outcomes

The prediction generated from classification algorithms may yield four possible outcomes, characterised as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [174].

A True Positive denotes when a model correctly predicts the positive class. For example, in a model trying to predict if a patient has a certain disease, a True Positive signifies that the model predicts the patient has the disease, and the patient does have the disease.

Similarly, a True Negative occurs when the model correctly predicts the negative class. Here, the model asserts the patient is disease-free, and it was indeed the case in reality.

A False Positive occurs when the model incorrectly identifies a negative instance as positive. With the disease example, the model would predict a patient of having a disease, while in reality the patient is disease-free.

Finally, False Negative is when the model incorrectly classifies a true positive instance as negative. With the example, the model would predict the patient is disease-free, while the patient does have the disease.

These four outcomes form the foundation of many additional performance metrics used to evaluate machine learning models, and offers detailed insights into a model's performance. Therefore understanding these outcomes is integral to the development and tuning of effective and accurate machine learning models.

7.3.2. Validation Methods

Testing and validation are critical steps in the machine learning pipeline, allowing the estimation of predictive performance, and helps to ensure that the model generalises well and performs effectively on unseen data. Two common validation techniques are Holdout and Cross-Validation.

The Holdout method involves splitting the available data into a training set and a test set. The model is trained on the training set, and then its performance is evaluated on the unseen test set [175]. This method is straightforward and computationally inexpensive, however, the assessment of the model using this method can depend heavily on how the data is split, which can lead to high variance in model performance. Additionally, a less optimised model might result from the fact that a portion of the data is not being utilised during training, especially when dealing with datasets with small number of instances.

Cross-Validation, or 'k-fold cross-validation' is a more robust method of evaluating the performance of a model. The process involves dividing the dataset into 'k' equally sized folds or subsets. The model is then trained 'k' times, each

time using a different fold as the test set and the remaining fold as the training set. The model's performance is evaluated as the average of the performance measures from the 'k' iterations [176]. This method reduces the variance associated with holdout method, providing a more reliable measure of how the model is expected to perform on unseen data. It also features an efficient use of data, as every instance is used in training and validation. Although as the model needs to be trained multiple times, more computational resources and time is often required, especially when the dataset is large, and the model is complex.

7.3.3. Confusion Matrix

The confusion matrix is a table that gives comparison between actual and predicted values from a classification model. It is a very popular method to visualise and summarise the results of a classification algorithm [177]. A sample confusion matrix is shown in Table 22.

Table 22: A sample confusion matrix

Disease	Healthy	← Predicted As ↓ Actual Class
88	12	Disease
5	95	Healthy

Each row of the table represents the instances of an actual class, while each column represents the instances of a predicted class. The diagonal elements of the matrix represent correct predictions, while the off-diagonal elements are incorrect predictions. By summing all the diagonal elements and dividing by the total, we can calculate the overall accuracy of the model.

The confusion matrix can also help in identifying the weakness in predicting certain classes in a model, where more mistakes were made compared to other classes. Such insights can help in identifying where additional training data or a different algorithm or model configuration might increase the performance.

The confusion matrix allows the calculation of several other important metrics like precision, recall, and the F1 score, which will be discussed below. These metrics provide a more nuanced understanding of the model's performance, especially in scenarios where the data might be imbalanced.

7.3.4. Accuracy, Recall, and Precision

Accuracy, Recall, and Precision are important concepts in understanding and assessing the performance of a machine learning model [178].

Accuracy is one of the simplest and most intuitive performance metrics in the context of classification model results. It is defined as the number of all correct predictions, including True Positive and True Negative, divided by total number of instances. Accuracy gives an overview of the classification performance, and works best when the dataset is balanced in terms of the number of instances in each class. However, a high accuracy does not always indicate a good overall performance of a model, and a number of other metrics need to be examined for a comprehensive evaluation of an algorithm.

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

Recall, also known as sensitivity or true positive rate, represents the model's ability to correctly identify all positive instances from the total actual positives. It is calculated by dividing the number of true positives by the sum of true positives and false negatives. Recall answers the question of "For a given class, how often does the algorithm correctly predict it?" A high recall implies a low rate of false negatives, meaning the model is good at detecting positive instances.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Precision quantifies the model's ability to correctly identify positive instances out of all instances it classified as positive. It is calculated by dividing the number of true positives by the sum of true positives and false positives. Precision answers the question of "When the model predicts an instance as positive, how often is it

correct?” A high precision score signifies a low false positive rate, meaning when the model predicts an instance to be positive, it is likely to be correct.

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

A model with high recall but low precision may make many positive predictions, but many of them may be incorrectly labelled. Conversely, a model with low recall but high precision might make fewer positive predictions, but a higher proportion of those predictions are likely to be correct. An ideal system should have both high precision and recall, but in real life, a trade-off between the two metrics is often needed based on the specific requirements of the problem at hand.

7.3.5. F1 Score

F1 score, also known as F-measure, is a metric that combines both Precision and Recall through their harmonic mean. When the Precision and Recall are similar, this measure is approximately the average of the two. However, if either precision or recall is significantly lower, the F1 score will decrease significantly, reflecting potential issues with the model's performance.

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

A F1 Score of 1.0 means that the model has perfect precision and recall, while a score of 0 suggests that the model has zero precision, zero recall, or both. This makes the F1 score a robust metric for scenarios where both precision and recall are crucial to the model's performance.

7.3.6. Training Time

Training time refers to the amount of time it takes for a machine learning model to learn patterns from a given training dataset. This metric can vary based on the complexity of the algorithm, parameter settings, the number of attributes and instances in the dataset, and the capabilities of the computing hardware. A

shorter training time can indicate a more efficient algorithm and lower usage of computational resources. However, a short training time does not necessarily lead to a better model, as its quality is ultimately determined by its ability to make accurate predictions. Therefore, training time is typically considered alongside other performance metrics when evaluating a model's effectiveness.

7.4. Model Training and Evaluation

In this section, several machine learning models are trained and evaluated using the prepared dataset. The predictive capabilities of these algorithms are investigated based on the evaluation measures introduced in the previous section.

The balanced dataset is saved in comma-separated values (CSV) format, a simple file format often used to store tabular data. A header at the beginning of the file defines the names of attributes at each column, and the data is organised into rows, with each attribute separated by commas. Its uncomplicated nature and compatibility with different data analysis tools have made CSV a popular choice for storing training datasets for machine learning experiments.

The dataset contains 7 parameters and a total of 2001 instances, with 667 instances in each of the three classes. Each row in the CSV file corresponds to a unique data point, and each column represents a specific parameter. The final parameter, as per machine learning customs, is the 'class' which the algorithms aim to predict. The values of all other parameters are in floating point numerical format. A snippet of the dataset, showing its structure and content, is illustrated in Figure 38.

```

rx_powA,rx_ipg,rx_speed_mph,tx_speed_mph,distance,is_dropped,class
-93,0.09974899999997433,42.81131755999999,59.840559999999996,471.42458640239613,0,Functional
-94,0.09980100000029779,42.81131755999999,59.840559999999996,471.42458640239613,0,Functional
-95,0.0998410000001968,42.81131755999999,59.840559999999996,471.42458640239613,0,Marginal
-92,0.19964099999924656,42.81131755999999,59.840559999999996,471.42458640239613,0,Marginal
-92,0.09993100000065169,43.784877439999995,57.0901958,482.10471527879395,0,Marginal
,,43.784877439999995,57.0901958,482.10471527879395,1,Disconnected
,,43.784877439999995,57.0901958,482.10471527879395,1,Disconnected
,,43.784877439999995,57.0901958,482.10471527879395,1,Disconnected
-94,0.09983899999951973,43.784877439999995,57.0901958,482.10471527879395,0,Marginal
,,43.784877439999995,57.0901958,482.10471527879395,1,Disconnected
-95,0.3996600000000399,44.61228826,59.23064659999999,491.981799144051,0,Marginal
-94,0.2044770000002245,44.61228826,59.23064659999999,491.981799144051,0,Marginal
-93,0.0949270000000979,44.61228826,59.23064659999999,491.981799144051,0,Marginal
-93,0.09999599999991915,44.61228826,59.23064659999999,491.981799144051,0,Marginal
-92,0.10002199999962613,44.61228826,59.23064659999999,491.981799144051,0,Functional
-92,0.10075000000051659,44.61228826,59.23064659999999,491.981799144051,0,Functional
    
```

Figure 38: Instances in the training dataset

Weka software suite is employed in this experiment for model training, validation, and analysis, due to its free and open-source nature, extensive collection of different machine learning models, friendly user interface, and robust community support from its widespread use in the machine learning community.

During the experiment, each machine learning model undergoes training using the identical dataset, and 10-fold cross validation is used for assessing the performance of each predictive model. Data is randomly partitioned into 10 equal sized subsamples, where 9 of them will be used to train the model, and the remaining one for validation. This process is then repeated 10 times, and the results are averaged to produce a single estimation. The confusion matrix is shown, and the performance is evaluated and discussed via the metrics including precision, recall, F1 score, accuracy, and training time.

7.4.1. ZeroR

ZeroR is one of the simplest classification algorithms that predicts all unseen instances as the majority class from the training dataset. Despite being simple and fast, its inability to leverage any features from the training set results in a lack of predictive power, typically leading to very low accuracy. Although ZeroR is not commonly used for practical applications due to its limitations, it is often

used to establish a baseline performance measure [179]. Any well-functioning classifier should surpass ZeroR's performance, otherwise it could indicate potential issues with the model or data.

Table 23 presents the training results obtained using the ZeroR method.

Table 23: ZeroR Classification Results

Functional	Disconnected	Marginal	← Predicted As ↓ Actual Class		Precision	Recall	F1 Score
667	0	0	Functional	Functional	0.333	1	0.5
667	0	0	Disconnected	Disconnected	-	-	-
667	0	0	Marginal	Marginal	-	-	-
Accuracy					33.33%		
Training Time					0 Seconds		

Given the equal distribution of classes, the algorithm arbitrarily selected "Functional" as the most common class, and subsequently classifying all instances as "Functional".

As a result, for the "Functional" class, ZeroR demonstrated perfect recall rate as it identified all instances of this class correctly. However, it also exhibited low precision due to classifying every single instance as "Functional", when only 667 were truly "Functional", resulting in a low F-1 score. For the remaining two classes, precision and F-1 score could not be calculated as all their instances were misclassified. Consequently, those classes have a recall rate of zero.

Overall, the ZeroR classifier achieved an accuracy of 33.3%, a sub-optimal outcome for this project, and typical of its performance. However, the training time was extremely short, and its result can serve as the baseline for the other models.

7.4.2. OneR

The OneR algorithm is another simple yet effective classification algorithm in machine learning. It generates a rule for each parameter in the dataset, and calculates the error rate of classifying instances based on that parameter's values. The parameter with the smallest total error rate is selected, and the rules generated from that parameter are used for classification of unknown data [180]. Despite its simplicity, OneR often demonstrates good performance in real-world situations. However, by only focusing on a single parameter, it may overlook complex relations between parameters in the dataset.

The rules generated by OneR method are shown in Figure 39, and the training results are shown in Table 24.

```

=== Classifier model (full training set) ===

rx_powA:
  < -83.5 -> Marginal
  < -80.5 -> Functional
  < -78.5 -> Marginal
  >= -78.5 -> Functional
  ? -> Disconnected
(1894/2001 instances correct)
    
```

Figure 39: OneR model rules

Table 24: OneR Classification Results

Functional	Disconnected	Marginal	← Predicted As ↓ Actual Class		Precision	Recall	F1 Score
622	0	45	Functional	Functional	0.909	0.933	0.921
0	667	0	Disconnected	Disconnected	1	1	1
62	0	605	Marginal	Marginal	0.931	0.907	0.909
Accuracy					94.65%		
Training Time					0 Seconds		

It can be observed that the OneR model selected the "received power level antenna A" parameter as the basis for its classification rules. Out of the 667

instances of the "Functional" class, it accurately predicted 622 of them, while misclassifying 45 instances as "Marginal". For the "Marginal" class, 605 out of the 667 instances were correctly classified, with 62 instances incorrectly labelled as "Functional".

The OneR model also correctly predicted all instances in the "Disconnected" class without making any mistakes. This level of accuracy can be explained by the nature of the chosen parameter, as when the V2V communication is interrupted, the received signal strength has no valid values, and the model capitalised on this fact to make a perfect prediction for this class.

Looking at statistics, the "Disconnected" class exhibits perfect precision, recall, and F1 score, due to its flawless prediction.

For the "Functional" class, a precision of 0.909 suggests that when the model predicts an instance to be "Functional", it is correct about 90.9% of the time. A recall of 0.933 indicates that the model identifies 93.3% of all actual "Functional" instances correctly. The F1 score of 0.921 suggests a balanced performance between precision and recall for the "Functional" class.

Similarly, for the "Marginal" class, a precision of 0.931 means that when the model predicts an instance to be "Marginal", it is correct approximately 93.1% of the time. A recall of 0.907 means the model correctly identifies 90.7% of all actual "Marginal" instances. The F1 score for the "Marginal" class is 0.909, again showing balanced performance between precision and recall.

The statistics suggest that the OneR model exhibits excellent result on all classes. However, a more detailed inspection of the generated rules indicates some counterintuitive classifications. As the signal strength increased from -83.5dBm to -78.5dBm, the model transitioned from labelling instances as "Marginal", to "Functional", back to "Marginal", and then finally to "Functional" again for signals stronger than -78.5dBm.

This pattern contradicts the expectation that a stronger signal would correlate with a lower likelihood of communication quality degradation. One possible explanation for this could be a minor instance of overfitting, where the model learns specific patterns in the training data so well that it performs poorly when applied to unseen data due to its inability to generalise.

This behaviour also highlights a potential limitation of the OneR method. By relying on a single parameter for decision-making, the method might have overlooked the complex relationships among other parameters in the dataset, generating rules that might be oversimplified and lack generalisability.

Despite the observed limitations, the OneR model's nearly 95% overall accuracy is notably impressive. Coupled with its extremely short training time, similar to the ZeroR model, it serves a strong benchmark for comparing the performance of other algorithms.

7.4.3. K-Nearest Neighbour

The k-Nearest Neighbour (k-NN) algorithm classifies unknown instances based on the majority vote of their 'k' nearest neighbours. k-NN is computationally inexpensive to train, but can be resource-intensive during prediction, especially with large, high-dimensional datasets, as the entire training dataset is stored in the model [181].

The number of neighbours, denoted by 'k', plays a crucial role in determining the predictive performance of a K-NN model. A small k value can lead to a fragmented decision boundary, causing the model to be highly sensitive to noise and outliers in the dataset, potentially resulting in overfitting.

Conversely, a large k value tends to produce a smoother decision boundary, which can help to reduce overfitting. However, this may cause the model to overlook finer details and patterns in the data, leading to over-generalisation or underfitting.

The selection of 'k' is often an experimental process. A common starting point is the square root of the number of instances in the dataset. This provides a balance between overfitting and underfitting, offering a solid foundation for further fine-tuning of the k value.

A range of k values were tested with the dataset, and the accuracy results are presented in Table 25.

Table 25: Influence of K value on KNN model accuracy

K	1	5	15	25	35	45	55	65	75	85
Accuracy (%)	98.00	96.50	95.80	96.10	95.25	95.35	94.05	92.85	92.70	93.20

The result suggests that a k value of 1 provides the best performance with an accuracy of 98.00%. While the result is impressive, it may indicate a potential risk of overfitting. As the k value increases, a gradual decrease in accuracy can be observed, reaching 93.20% when k is 85. This trend illustrates a reduction in overfitting and the enhanced generalisation of the model as the k-value increases. Table 26 presents the KNN training results with a more balanced K value of 45.

Table 26: KNN Classification Results with K = 45

Functional	Disconnected	Marginal	← Predicted As ↓ Actual Class		Precision	Recall	F1 Score
620	0	47	Functional	Functional	0.931	0.930	0.930
0	667	0	Disconnected	Disconnected	1	1	1
46	0	621	Marginal	Marginal	0.930	0.931	0.930
Accuracy						95.35%	
Training Time						0 Seconds	

For the "Functional" class, the model accurately classified 620 out of 667 instances, while 47 instances were incorrectly identified as "Marginal". Similarly, for the "Marginal" class, out of 667 instances, 621 were correctly labelled, and 46 were misclassified as "Functional". Again, the model correctly identified all instances in the "Disconnected" class, with no errors made.

The statistics reveal that the precision, recall, and F1 score for both "Functional" and "Marginal" classes are approximately 0.930, indicating a more balanced performance compared to the OneR method. The overall accuracy has also improved slightly to 95.35%. The training time for the KNN model remained minimal, consistent with the expectation that KNN requires minimal computation during the training phase.

7.4.4. J48 Decision Tree

The Decision Tree algorithm constructs a tree-like model by recursively partitioning data into subsets based on attribute values. The resulting tree consists of root, internal, and leaf nodes. To make a prediction, one traverses the tree from the root following branches according to the attribute values, and arrives at a leaf node, which represents the decision. The output of a decision tree algorithm is easily interpretable and can handle both numerical and categorical data. It can also perform well with large datasets [182].

However, decision trees can be sensitive to minor changes in the training data, which may significantly alter the tree structure. They also risk overfitting the training data by creating overly complex trees. Techniques like pruning are used to combat overfitting, improving the tree's performance on unseen data.

Many algorithms can be used to construct decision trees. The ID3 algorithm, as one of the earliest, primarily handles categorical attributes and utilises information gain for decision-making. This was later enhanced by C4.5 algorithm, which supports to both categorical and numerical data, can handle missing

values, and uses gain ratio for node splitting [183]. The model in Weka uses the J48 algorithm, which is a java implementation of C4.5.

The tree generated by J48 algorithm is shown in Figure 40, and the training results are shown in Table 27.

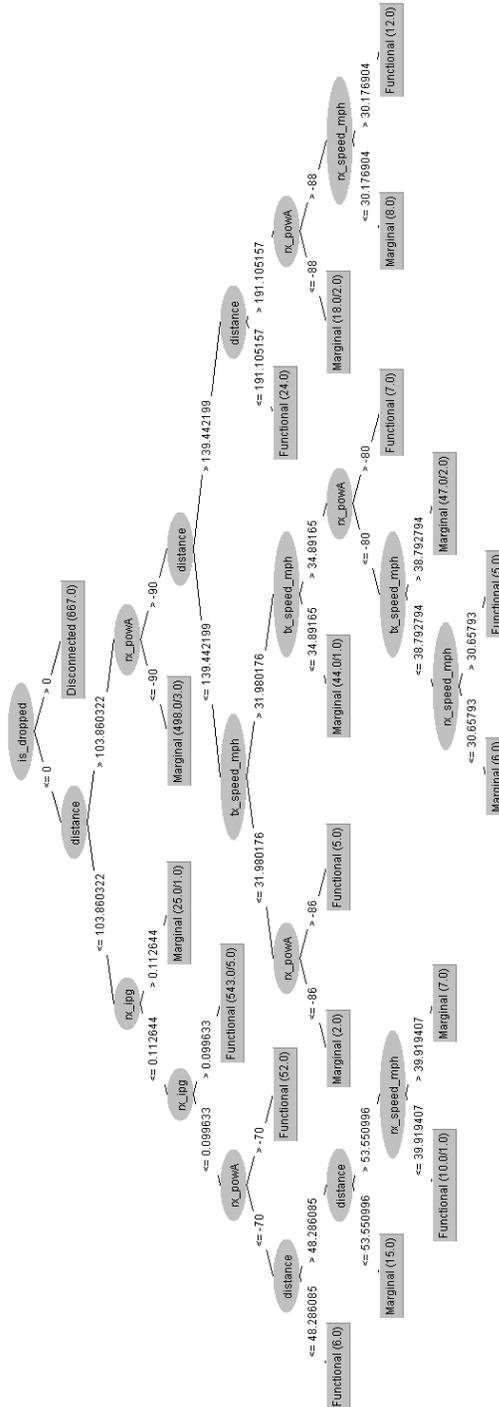


Figure 40: Constructed J48 Decision Tree

Table 27: J48 Decision Tree Classification Results

Functional	Disconnected	Marginal	← Predicted As ↓ Actual Class		Precision	Recall	F1 Score
637	0	30	Functional	Functional	0.974	0.955	0.964
0	667	0	Disconnected	Disconnected	1	1	1
17	0	650	Marginal	Marginal	0.956	0.975	0.965
Accuracy						97.65%	
Training Time						0.2 Seconds	

It can be seen that unlike OneR, the J48 algorithm is able to discern relationships among the parameters, and utilises all parameters for decision making. As a result, this approach increases the predictive performance with an overall accuracy of 97.65%. Only 30 instances in the "Functional" class were misidentified, and 17 instances in the "Marginal" class were misidentified. The training time has a slight increase to 0.2 seconds due to the computational load associated with the more sophisticated algorithm.

However, despite the impressive accuracy, it is easy to observe the complexity of the tree that it generated, with a depth of 9 levels, 19 internal nodes and 20 leaves. This is a very visual representation of overfitting, where the model over-learns the training data, and the overly complicated and rigid tree might limit the model's ability to generalise effectively to unseen data.

A technique known as pruning can be used to simplify and reduce the size of decision trees, in order to make them more efficient and to help mitigate overfitting. This is done by removing sections of the tree that provide little predictive power. The J48 algorithm incorporates pruning, and a parameter known as "confidenceFactor" can be adjusted to change aggressiveness of the pruning [184]. However, despite setting this parameter to its most aggressive level, the resulting tree remains complex.

7.4.5. REPTree

The REPTree (Reduced Error Pruning Tree) is another decision tree learning algorithm for classification and regression tasks. Similar to J48, REPTree builds a tree-like model of decisions, with each node representing a test on an attribute, branches representing the outcome of the test, and leaf nodes corresponding to class labels.

However, one of the major differences between REPTree and J48 lies in the pruning techniques. While J48 constructs a full decision tree before applying a post-pruning step, REPTree employs a strategy known as reduced-error pruning. This involves pruning the tree during its construction by removing subtrees if doing so reduces the estimated error rate. This bottom-up pruning is performed as the tree is being built, rather than as a separate step after the tree is fully constructed [185].

By incorporating pruning during tree construction, REPTree can reduce complexity and potentially reduce training times compared to J48. Additionally, the continuous pruning can lead to simpler, more understandable trees which are less prone to overfitting.

An additional benefit of the REPTree implementation in Weka is the ability to set a maximum tree depth. Unrestricted, REPTree yields a tree of similar complexity to the J48 algorithm, with a maximum depth of 11, 43 internal nodes, and a similar accuracy of 97.05%. However, by imposing a maximum tree depth, a simpler and more generalisable model can be achieved.

The tree generated by REPTree algorithm with a depth limit of 4 is shown in Figure 41, and the training results are shown in Table 28.

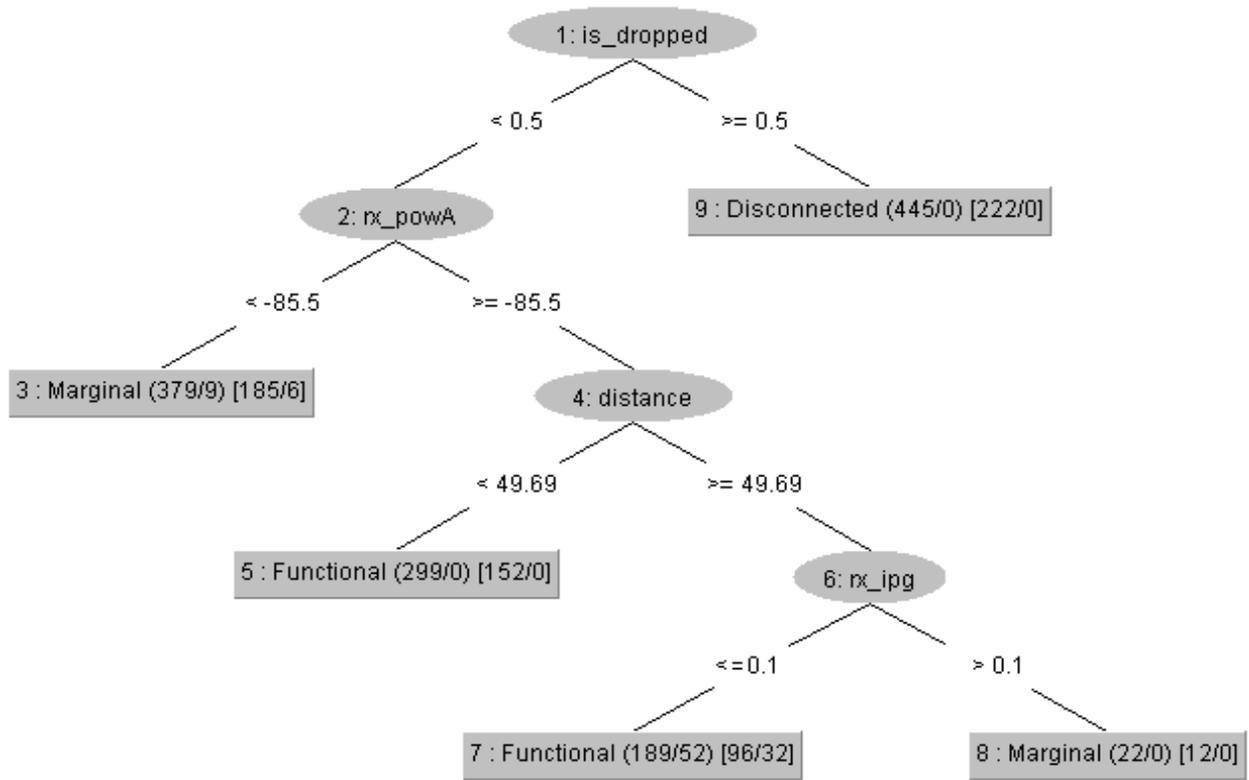


Figure 41: REPTree with depth limit of 4

Table 28: REPTree with depth limit of 4 classification results

Functional	Disconnected	Marginal	← Predicted As ↓ Actual Class		Precision	Recall	F1 Score
625	0	42	Functional	Functional	0.915	0.937	0.926
0	667	0	Disconnected	Disconnected	1	1	1
58	0	509	Marginal	Marginal	0.935	0.913	0.924
Accuracy						95.00%	
Training Time						0.3 Seconds	

Similar to the J48 decision tree, multiple parameters were used to construct the REPTree, so the maximum amount of information from the dataset can be extracted. However, with a depth limit of 4, the tree was pruned to be much simpler and more intuitive to interpret. Initially the model checks the "is_dropped" parameter to assess if the connection is interrupted. If not, it proceeds to verify the received power level, labelling anything weaker than -85.5dBm as "Marginal". Subsequent checks are made on distance and IPG, and it only deems a scenario as "Marginal" if the distance between the vehicles exceeds 49.69 meters and the IPG is greater than 0.1 seconds. This reflects the intuitive expectation that longer distances and a deteriorating IPG could indeed compromise communication.

The REPTree model with depth limit of 4 has an overall accuracy of 95%, with 42 misidentified instances in the "Functional" class, and 17 misidentified instances in the "Marginal" class. The training time is similar to the J48 tree. Although it exhibits a slightly lower accuracy compared to J48 tree's 97.65%. The marked simplification of the model and mitigation of overfitting enhance its real-world applicability. This is achieved by providing more understandable results, easier interpretation, and greater generalisability.

Attempts to reduce complexity and overfitting were carried out by further limiting the tree depth. A depth limit of 2 and 3 both produced the same tree, shown in Figure 42, and the training results are shown in Table 29.

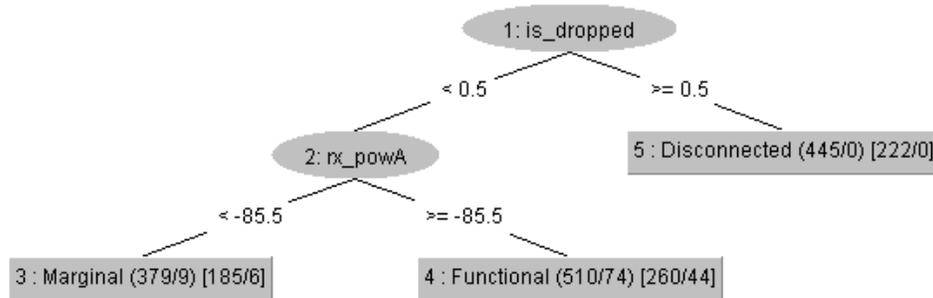


Figure 42: REPTree with depth limit of 2

Table 29: REPTree with depth limit of 2 classification results

Functional	Disconnected	Marginal	← Predicted As ↓ Actual Class		Precision	Recall	F1 Score
610	0	57	Functional	Functional	0.905	0.915	0.910
0	667	0	Disconnected	Disconnected	1	1	1
64	0	603	Marginal	Marginal	0.914	0.904	0.909
Accuracy						93.95%	
Training Time						0.3 Seconds	

The updated REPTree is only two levels deep, and utilised only two parameters. The "is_dropped" parameter is checked first to determine if the connection is interrupted, then the model labels all instances with a signal stronger than -85.5dBm as "Functional", and otherwise "Marginal". Compared to the previous tree with a depth limit of 4, it is apparent that the branch involving distance and IPG checks has been pruned, leading to further simplification and intuitive comprehension. The pruning of the branch also led to a minor decrease in predictive performance, resulting in an overall accuracy of 93.95%. However, the approximately 1% difference may be an acceptable trade-off for a simpler and more flexible model.

7.5. Summary

Following the successful execution of the outdoor V2V communication experiment, this chapter investigates the potential use of machine learning to effectively mitigate potential disruptions. By utilising classification models to predict potential communication dropouts before they occur, timely action to switch to alternative communication channels, such as cellular or satellite, can be facilitated, minimising downtime particularly in safety critical scenarios.

Several data pre-processing steps were carried out. Class labels were created for the training dataset, and the large number of parameters were reduced by discarding irrelevant parameters, consolidating parameters with strong correlations to each other, and balancing the dataset with a combination of undersampling and oversampling. Those steps ensured a balanced and robust dataset for the best possible performance in model training. A number of evaluation metrics were also introduced to provide a quantitative understanding of the effectiveness of each algorithm.

The investigation subsequently trained five machine learning models: ZeroR, OneR, KNN, J48 decision tree, and REPTree, using the pre-processed dataset. Excluding ZeroR, all models demonstrated impressive levels of performance, exceeding 90% accuracy. The J48 decision tree exhibited highest level of accuracy at 97.65%, at the expense of significant overfitting, a tendency also observed in KNN and, to a lesser extent, OneR. Initially, the REPTree algorithm also displayed similar overfitting behaviours, but upon adjusting the tree depth limit, much more simplified and generalised trees were produced with minimal accuracy loss, and it was deemed to be the most robust choice among the models investigated, due to its excellent performance, tunable tree complexity, and intuitive and interpretable output format.

In conclusion, the original objective of this chapter indeed appears to be viable. The machine learning model can effectively interpret real-time key performance

indicators data to predict early signs of communication interruption. While the results have been generally positive, there are several areas identified for improvement. The communication disruption during the on-road experiment was relatively brief, lasting less than two minutes, providing a limited dataset. Moreover, most of the communication dropouts during the experiment occurred during relatively high-speed driving on dual carriageways, leading to a potential gap in data from urban areas. Another observed limitation was the insignificant difference in training times across the models, likely attributable to the modest size of the dataset. Therefore, gathering additional data from diverse scenarios in future experiments could significantly refine and optimise the model, enabling it to perform at its highest potential.

Chapter 8: Conclusions

8.1. Main Contributions

The rapid growth of global vehicular traffic has necessitated intelligent transportation systems to combat congestion, ensure road safety, and maintain environmental sustainability. In light of this, the Connected and Autonomous Vehicles (CAV) stand out as a promising technology for the future of transportation, with Vehicle-to-Vehicle (V2V) communication serving as a critical component. V2V communication allows vehicles to share dynamic information amongst one another, allowing increased road safety through reduced collision risks, improved traffic flow via enhanced vehicle coordination, and a minimised environmental footprint by optimising driving habits, fuel consumption, and emissions.

Although the potential of widespread V2V networks have been theorised over the past 40 years, it was only in the recent decades that technological progress has made real-world V2V deployments feasible. Numerous pilot projects and field trials have been carried worldwide to evaluate the practicality and efficacy of V2V networks, interoperable standards have been proposed and developed by regulatory bodies, and automotive manufacturers are increasingly adopting V2V capabilities in their new vehicles.

However, while there has been numerous research on V2V communication, a significant portion has been conducted within simulated environments. While these studies provide important insights into V2V performance, the results may be limited by the assumptions and simplifications inherent in simulation models, and may not accurately reflect the realities and challenges associated with V2V system deployment in real-world conditions.

This thesis aims to bridge the gap by performing an in-depth evaluation of V2V communication systems in real-world scenarios, involving a comprehensive review of existing literature, investigation and development of a custom OBU hardware platform with wide range of connectivity capabilities and modular form factor, planning and execution of outdoor experiments with road-legal vehicles on public roads, processing, examination, and discussion of collected data, and experimentation with machine learning algorithms for the proactive prediction of communication disruptions.

A detailed review of existing literature regarding V2V communication was carried out in Chapter 2, including its background and history, working principle and benefits, competing standards of DSRC and C-V2X and their strengths and weaknesses, infrastructure requirements, and its deployment status and future challenges. The working principle and the merits and weaknesses of several popular machine learning classification algorithms were also discussed. The literature review provides an enhanced understanding of the topics involved, and establishes a theoretical foundation for subsequent chapters.

Chapter 3 documents the development process of a custom OBU hardware platform. The design goals and requirements were first identified, with a discussion of the benefits of developing the custom hardware to suit the exact needs of this project. The purpose, requirements, and justification of the major components of the OBU were discussed, as well as an introduction of Printed Circuit Boards and several important design considerations for optimal performance involving high-speed signals. The design of two iterations of the custom OBU board was introduced, including the novel incorporation of PC/104 form factor to improve modularity and upgradeability.

With the completion of the OBU hardware and in anticipation of the outdoor experiments, Chapter 4 introduced a number of software packages and digital tools to be used for data processing, visualisation, and analysis, as well as a

selection of Key Performance Indicators for quantifiable assessment of system performance, laying the groundwork for concepts in later chapters.

Chapter 5 documents the first outdoor test utilising the completed OBU hardware with the goal of confirming the OBU and its subsystems were performing in a stable and reliable manner. Test plans and procedures were described in detail, and the data collected from the preliminary test was visualised, and the observed KPI characteristics are discussed. The experiment validated the hardware design's functionality and stability, and the KPI data revealed a maximum communication range nearing 900 meters under open-air conditions with line of sight.

After the successful preliminary experiment, Chapter 6 documents the planning, execution, and the results of a new outdoor experiment that was considerably more comprehensive than its predecessor, featuring real-world driving scenarios with multiple road-legal vehicles navigating public roadways under a diverse range of traffic conditions, road environments, and vehicular speeds.

The experiment was carried out successfully according to plan, and the test results were examined and discussed. It was found that the two vehicles were able to maintain stable V2V communication for the majority of the experiment, although compared to the preliminary test in near-ideal conditions, the signal quality in this test exhibited noticeably more variance due to the real-world imperfections such as buildings, foliage, curves, and obstruction of other vehicles. FSPL was responsible for most of the signal degradation, although additional factors such as small-scale fading and signal shadowing also contributed to the power loss.

Inter-Packet Gap (IPG) and Packet Error Rate (PER) parameters were also investigated, and a usable communication range of the V2V setup was determined to be approximately 550 meters, with a 10% PER threshold. This was

notably shorter than maximum communication range due to the aforementioned factors.

Another important part of the experiment was to leverage the capability of V2V to transmit and broadcast GNSS correction data to nearby vehicles, allowing improved positioning accuracy even if they lack network connectivity of their own. The collected data was compared to the reference positioning result from a Leica receiver, and improvements of 25% in median accuracy and 80% in standard deviations were observed, validating the feasibility of this objective.

Finally, cellular environment data along the test route was examined, revealing more optimal cellular service quality in urban and built-up areas, and outdated service types, weaker signals, and reduced network throughput in rural areas, suggesting that infrastructure improvements may be required in order to achieve optimal Cellular V2X performance.

Chapter 7 covers the experiments to investigate the potential use of machine learning classification models to mitigate potential V2V communication disruptions. The dataset was pre-processed, and five models were trained, with REPTree displaying the most optimal balance between accuracy and generalisability, demonstrating the feasibility of the objective.

Overall, this research project features a multidisciplinary approach for a detailed investigation of the construction, deployment, and performance of V2V communication system in real-world environments, while also exploring additional areas of innovation such as GNSS correction data broadcasting, and machine learning-based pre-emptive interruption prediction. The results of this research project make valuable contributions to the field of V2V research, and provides a robust foundation for further explorations.

8.2. Limitations and Areas for Improvement

While this research project has yielded meaningful and valuable results, several limitations and areas for improvement have been identified, which might further enhance the completeness and quality of this project.

A notable area for improvement involves broadening the quantity and diversity of data collection. While the test route incorporated a variety of driving scenarios, including different speed limits, following distances, and road types, it did not cover several other typical scenarios such as congested urban traffic, rural areas, motorway driving, and weather conditions such as rain, fog, or snow. Additional experiments incorporating these scenarios could provide further insight into system performance. Moreover, during the on-road experiment, communication disruptions were brief and mostly occurred during high-speed travel on dual carriageways. Gathering more data on communication interruption scenarios in urban environments could also offer deeper understanding of V2V performance and improve the effectiveness of machine learning models.

Another potential improvement involves the use of higher-quality equipment, such as professional-grade, dual-frequency GNSS antennas to enhance positioning accuracy, and 5.9GHz DSRC antennas for improved signal strength and extended communication range. The incorporation of such upgrades may have enhanced the results even further.

Finally, a closer and more detailed exploration of C-V2X technology could potentially yield more insights into the evolving V2V landscape. As an emerging technology, C-V2X offers both device-to-device and device-to-infrastructure communication options, attracting the attention of many automobile manufacturers. Therefore, it would have been beneficial to evaluate and compare the performance of both C-V2X and DSRC-based V2V systems across various real-world scenarios, in order to provide valuable perspectives on the

respective strengths and limitations of these technologies, and inform the strategies for future V2V system development and deployment.

8.3. Future Work

As this research project concludes, the achieved goals set a robust groundwork for future investigations.

As previously suggested, conducting additional experiments in diverse driving environments, as well as comparing DSRC-based V2V communication with C-V2X, would contribute to a deeper understanding of V2V performance in real-world settings. Gathering more data would also encourage the development of more accurate and robust machine learning models capable of predicting communication disruptions, as well as developing plans to observe how they perform in real time during experiments.

The OBU hardware can also benefit from upgrades to keep pace with the latest technology, such as updating to the faster and more capable Raspberry Pi Compute Module 4, integrating a new C-V2X transceiver, adding 5G cellular capability, upgrading to a dual-frequency GNSS receiver, and incorporating additional sensors. The PC/104 stackable and modular form factor of the OBU ensures these upgrades can be easily incorporated by adding or replacing the relevant circuit boards.

Moreover, the experience gained from the hardware design process, along with the intellectual property accrued, lays a strong foundation for the potential commercialization of the device. Additional steps, such as designing an appropriate enclosure, performing further testing, and obtaining necessary certifications, may be needed to make the device market-ready. This research thus not only contributes to academic knowledge but also holds promise for tangible, real-world impact.

8.4. Summary

This research project features a multidisciplinary approach of a detailed investigation of the construction, deployment, and performance of V2V communication system in real-world environments, while also exploring additional areas of innovation such as GNSS correction data broadcasting, and machine learning-based pre-emptive interruption prediction.

The project began with an extensive literature review on V2V communication, considering its historical development, functioning, benefits, competing standards, and future challenges, serving as a foundation for the investigation.

The research then covered the development of a custom on-board unit (OBU) hardware platform designed to meet the project's specific requirements.

Subsequently, outdoor experiments were planned and carried out using the custom OBU. Various software packages and tools were selected for data processing, visualisation, and analysis. Key Performance Indicators (KPIs) were outlined to assess the system's performance quantitatively.

Comprehensive experiments were planned and carried out, testing the OBU under real-world driving scenarios. Additionally, the experiment explored a custom implementation of GNSS correction data broadcasting that allows nearby vehicles to improve their positioning accuracy. The feasibility was confirmed by a significant improvement in positioning accuracy.

The research also investigated the application of machine learning classification models to predict and mitigate potential V2V communication disruptions, testing a number of classification models with excellent results.

This project delivers a detailed investigation into V2V communication systems, with an emphasis on practical application and real-world scenarios. It features important aspects of the design, deployment, and performance of these systems while also exploring innovative concepts such as GNSS correction data broadcasting and disruption prediction through machine learning.

Furthermore, the research led to the development of substantial intellectual property, including the custom OBU hardware platform and the machine learning models, which holds potential not only for further academic progression, but also for potential commercialisation, as well as making a contribution to the future of intelligent transportation systems.

Chapter 9: Bibliography

- [1] D. T. M. B. S. R. S. B. Ehsan Moradi-Pari, "DSRC Versus LTE-V2X: Empirical Performance Analysis of Direct Vehicular Communication Technologies," *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, 2023.
- [2] National Highway Traffic Safety Administration, "U.S. DOT advances deployment of Connected Vehicle Technology to prevent hundreds of thousands of crashes," 2016.
- [3] L. F. Coleman Bazelon, "The Economic Costs and Benefits of a Federal Mandate that All Light Vehicles Employ 5.9 GHz DSRC Technology," 2016.
- [4] European Telecommunications Standards Institute, "ETSI EN 302 663 V1.2.0 - Intelligent Transport Systems; Access layer specification for Intelligent Transport Systems operating in the 5 GHz frequency band," 2012.
- [5] Federal Communications Commission , "FCC Adopts New Rules For The 6 GHz Band, Unleashing 1,200 Megahertz Of Spectrum For Unlicensed Use," 2020.
- [6] S. Chen, J. Hu, Y. Shi, L. Zhao and W. Li, "A Vision of C-V2X: Technologies, Field Testing and Challenges with Chinese Development," *IEEE Internet of Things Journal (Early Access)*, 2020.
- [7] Toyota USA, "Toyota and Lexus to Launch Technology to Connect Vehicles and Infrastructure in the U.S. in 2021," 2018.

- [8] EE Times, "VW Pits its New Golf Against the 5G Lobby," 2019.
- [9] Harding, J., Powell, G., R., Yoon, R., Fikentscher, J., Doyle, C., Sade, D., Lukuc, M., Simons, J., & Wang, J., "Vehicle-to-vehicle communications: Readiness of V2V technology for application," National Highway Traffic Safety Administration, Washington, DC, 2014.
- [10] C. Weiß, "V2X communication in Europe – From research projects towards standardisation and field testing of vehicle communication technology," *Computer Networks*, vol. 55, no. 14, pp. 3103-3119, 2011.
- [11] M. D. M. N. Eduard Zadobrischi, "The Utility of DSRC and V2X in Road Safety Applications and Intelligent Parking: Similarities, Differences, and the Future of Vehicular Communication," *Sensors*.
- [12] Kenney, J.B., "Dedicated Short-Range Communications (DSRC) Standards in the United States," in *Proceedings of the IEEE*, 2011.
- [13] L. F. Coleman Bazelon, "The Economic Costs and Benefits of a Federal Mandate that All Light Vehicles Employ 5.9 GHz DSRC Technology".
- [14] "Programme for a european traffic system with highest efficiency and unprecedented safety," EUREKA website, 1987.
- [15] Mercedes-Benz Group AG, "From the intelligent to the autonomous car," [Online]. Available: <https://group.mercedes-benz.com/innovation/case/autonomous/intelligent-drive-2.html>. [Accessed 22 10 2023].
- [16] N. Congress, "The Automated Highway System," *Public Roads*, vol. 58, no. 1, 1994.

- [17] M. Novak, "The National Automated Highway System That Almost Was," *Smithsonian Magazine*, 16 5 2013. [Online]. Available: <https://www.smithsonianmag.com/history/the-national-automated-highway-system-that-almost-was-63027245/>.
- [18] Federal Communications Commission, "Proposed Rules," *Federal Register*, vol. 85, no. 25, pp. 6841-6856, 2020.
- [19] U. D. o. Transportation, "5.9 GHz Transportation Safety Band Testing".
- [20] K. Ansari, "Joint use of DSRC and C-V2X for V2X communications in the 5.9 GHz ITS band," *IET Intelligent Transport Systems*, 2020.
- [21] A. K. K. O. S. I. Zachary MacHardy, "V2X Access Technologies: Regulation, Research, and Remaining Challenges," *IEEE COMMUNICATIONS SURVEYS & TUTORIALS*, vol. 20, no. 3, p. 1858, 2018.
- [22] 5G Automotive Association, "The Case for Cellular V2X for Safety and Cooperative Driving," 2016.
- [23] Qualcomm Technologies, Inc., "How NR based sidelink expands 5G C-V2X to support new advanced use cases," 2020.
- [24] Autotalks Ltd., "Autotalks Launches the World's First Global V2X Solution Unifying DSRC and C-V2X on its Deployment-Ready Chipset," 2019.
- [25] Quectel, "Quectel AG15 Datasheet".
- [26] I. Qualcomm Technologies, "Cellular-V2X Technology Overview," 2019.
- [27] I. Qualcomm Technologies, "Introduction to Cellular V2X," 2019.

- [28] 3rd Generation Partnership Project, "Technical Specification Group Services and System Aspects, Release 14 Description," 2018.
- [29] J. J. V. K.-L. H. Lili Miao, "PC5-Based Cellular-V2X Evolution and Deployment," *Sensors*.
- [30] 3GPP Standard TS 22.185, "Service Requirements for V2X Services," 2016.
- [31] 3GPP, "3GPP Release 15," 26 4 2019. [Online]. Available: <https://www.3gpp.org/specifications-technologies/releases/release-15>.
- [32] S. S. H. H. A. B. S. H. A.-z. Ahmed Elbery*, "To DSRC or 5G? A Safety Analysis for Connected and Autonomous Vehicles," *2021 IEEE Global Communications Conference*, 2021.
- [33] 5G Automotive Association, "An assessment of LTE-V2X (PC5) and 802.11p direct communications technologies for improved road safety in the EU," 2017.
- [34] 5G Americas White Paper, "Cellular V2X Communications Towards 5G," 2018.
- [35] 5. A. Association, "List of C-V2X Devices," 2021.
- [36] TOYOTA MOTOR NORTH AMERICA, INC. , "Letter to Federal Communications Commission," 2018.
- [37] J. T. F. F. Y.-B. K. ZEESHAN HAMEED MIR, "Enabling DSRC and C-V2X Integrated Hybrid Vehicular Networks: Architecture and Protocol," *IEEE Access*, 2020.

- [38] B. L. Q. S. L. G. A. J. Weijing Qi, "Traffic Differentiated Clustering Routing in DSRC and C-V2X Hybrid Vehicular Networks," *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, 2020.
- [39] I. HARMAN INTERNATIONAL INDUSTRIES, "HARMAN Savari MobiWAVE Factsheet".
- [40] INFORMATION SOCIETY TECHNOLOGIES (IST) PROGRAMME, "Dangerous Good Transportation Routing, Monitoring and Enforcement," 2006.
- [41] Savari, "MobiWAVE MW-1000 OBU Datasheet," 2017.
- [42] u-blox AG, "THEO-P173 Host-based V2X transceiver module datasheet," 2016.
- [43] R. Orlove, "Mercedes E Classes Can Talk To Each Other Now," *Jalopnik*, 2016.
- [44] Cadillac, "V2V Safety Technology Now Standard on Cadillac CTS Sedans," 2017.
- [45] J. Lowy, "Gov't won't pursue talking car mandate," *Associated Press*, 2017.
- [46] House Committee on Transportation and Infrastructure, "Full TI Letter to FCC," Washington, 2020.
- [47] Wi-Fi Alliance, "Wi-Fi Alliance introduces Wi-Fi 6," 2018.
- [48] J. Yoshida, "Pandemic Shouldn't Justify Killing DSRC," *EE Times*, 2020.
- [49] Volkswagen AG, "With the aim of increasing safety in road traffic, Volkswagen will enable vehicles to communicate with each other as from 2019," 2017.

- [50] O. Blanchard, "Why C-V2X May Yet Become The Global Automotive Connectivity Standard," Futurum Research, 2019.
- [51] S. Abuelsamid, "Ford Breaks With Auto Rivals By Committing To C-V2X Vehicle Communications Tech," Forbes, 2019.
- [52] S. Lawson, "BMW, Ericsson & Vodafone Pushing C-V2X Adoption in EU," Informa PLC., 2018.
- [53] M. Allevan, "Report: DSRC mandate moving off the table for automakers," fiercewireless, 2017.
- [54] G. Elinoff, "Is C-V2X Overtaking DSRC in Vehicle-to-Vehicle Communications?," EETech Media, LLC, 2019.
- [55] R. Kohavi and F. Provost, "Glossary of terms," *Machine Learning*, vol. 30, pp. 271-274, 1998.
- [56] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [57] Weiss, M., Indurkha, N., "Rule-based Machine Learning Methods for Functional Prediction," *Journal of Artificial Intelligence Research*, vol. 3, pp. 383-403, 1995.
- [58] Altman, N, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," *The American Statistician*, vol. 46, no. 3, pp. 175-185, 1992.
- [59] Everitt, B. S., Landau, S., Leese, M. and Stahl, D, "Miscellaneous Clustering Methods," in *Cluster Analysis*, John Wiley & Sons, 2011.
- [60] Rokach, L. and Maimon, O., "Data mining with decision trees: theory and applications," World Scientific Pub Co, 2008.

- [61] Maimon, O. and Rokach, L., *Data Mining and Knowledge Discovery Handbook*, 2006.
- [62] J. R. Quinlan, "Induction of decision trees," in *Machine Learning*, 1986, p. 81.
- [63] Barros, R. C., de Carvalho, A. C. P. L. F. and Freitas, A. A., "Automatic Design of Decision-Tree Induction Algorithms," Springer, 2015.
- [64] Gareth, J., Witten, D., Hastie, T. and Tibshirani, R., *An Introduction to Statistical Learning*, Springer, 2015, p. 315.
- [65] Breiman, L., "Random Forests," *Machine Learning*, pp. 5-32, 2001.
- [66] Ho, T. K., "Random Decision Forests," in *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Montreal, 1995.
- [67] Ho, T. K., "The Random Subspace Method for Constructing Decision Forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832-844, 1998.
- [68] Ho, T. K., "A Data Complexity Analysis of Comparative Advantages of Decision Forest Constructors," in *Pattern Analysis and Applications*, 2002, pp. 102-112.
- [69] M. K. H. Elmustafa Sayed Ali, "Machine Learning Technologies for Secure Vehicular Communication in Internet of Vehicles: Recent Advances and Applications," *Machine Learning and Applied Cryptography*, vol. 2021, 2020.

- [70] Z. L. H. L. Md. Noor-A-Rahim, "6G for Vehicle-to-Everything (V2X) Communications: Enabling Technologies, Challenges, and Opportunities," *Proceedings of the IEEE*, p. 172, 2022.
- [71] I. M. G. W. S. C. C. D. A. S. Darlan C. Moreira, "QoS Predictability in V2X Communication with Machine Learning," in *IEEE 91st Vehicular Technology Conference*, 2020.
- [72] S. Y. S. Farooque Hassan Kumbhar, "DT-VAR: Decision Tree Predicted Compatibility-Based Vehicular Ad-Hoc Reliable Routing," *IEEE WIRELESS COMMUNICATIONS LETTER*, vol. 10, no. 1, p. 87, 2021.
- [73] Q. He, "A Machine Learning-based Anomaly Detection Framework for Connected and Autonomous Vehicles Cyber Security," 2021.
- [74] X. M. C. X. X. W. Y. C. G. Y. L. B. Simon Roberts, "CoDRIVE – Delivering high accuracy, ubiquitous positioning through combined radio navigation and inertial sensing technologies," 2020.
- [75] Savari, "MobiWAVE2000 Dual Mode On-Board-Unit," May 2019. [Online]. Available: https://savari.net/wp-content/uploads/2019/05/MW-2000_May2019-1.pdf.
- [76] Commsignia, "Commsignia ITS-OB4," [Online]. Available: https://shop.commsignia.com/index.php?route=product/product&path=64&product_id=54.
- [77] Unex Technology Group, "RSU-101E V2X Roadside Communication Unit, ITS-G5 stack," 2020. [Online]. Available: <https://www.unex.com.tw/products/v2x/v2isolution/v2x-rsu/etsi-tc-its-stack-1/detail/rsu-101e>.

- [78] M. Levin, "Cost-benefit analysis of fuel-efficient speed control using signal phasing and timing (SPaT) data: evaluation for future connected corridor deployment," University of Minnesota, 2023.
- [79] HARMAN INTERNATIONAL INDUSTRIES, INCORPORATED, "HARMAN Savari MobiWAVE Data Sheet".
- [80] A. Ltd, "At the Heart of the Smartphone," ARM Ltd, [Online]. Available: <https://www.arm.com/markets/consumer-technologies/smartphones>.
- [81] NVIDIA CORPORATION, "NVIDIA Jetson TX2 NX System-on-Module Datasheet".
- [82] D. D. A. V. V. Harsha Vasudev, "Computers and Electrical Engineering 82 (2020) 106555," *Computers and Electrical Engineering*, vol. 82, 2020.
- [83] "Analysis of Low Cost Communication Technologies," *applied sciences*, no. 10, p. 1249, 2020.
- [84] V. D. D. D. S. K. D. Harsha Vasudev, "A Lightweight Mutual Authentication Protocol for V2V Communication in Internet of Vehicles," *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, vol. 69, no. 6, 2020.
- [85] Raspberry Pi (Trading) Ltd., "Raspberry Pi Compute Module 3+ Datasheet," Jan 2019. [Online]. Available: https://www.raspberrypi.org/documentation/hardware/computemodule/datasheets/rpi_DATA_CM3plus_1p0.pdf.
- [86] Cohda Wireless Pty Ltd, "MK6C EVK Specs," [Online]. Available: https://cohdawireless.com/wp-content/uploads/2019/07/3.-CW_Product-Brief-sheet-MK6C-EVK-v2.docx.pdf.
- [87] Cohda Wireless, "MK5 RSU Specs," 2018.

- [88] Cohda Wireless, "MK5 OBU Specs," 2018.
- [89] S. Y. Zhen Li, "Design and Research of a Vehicle Approaching Reminder Device Based on ZigBee Wireless Sensor Network," *Artificial Intelligence for Evaluation Decision-making in Modern Product Design*, 2022.
- [90] ublox, "THEO-P173 Host-based V2X transceiver module System Integration Manual".
- [91] u-blox AG, "NEO/LEA-M8T Data Sheet," June 2016. [Online]. Available: https://www.u-blox.com/sites/default/files/NEO-LEA-M8T-FW3_DataSheet_%28UBX-15025193%29.pdf.
- [92] N. Nadarajah, "Multi-GNSS PPP-RTK: From Large to Small-Scale Networks," *Sensors*, 2018.
- [93] E. Y. H. M. N. C. A. R. I. S. Y. A. S. Yuwono, "Assessment of the Single Frequency Low Cost GPS RTK Positioning," in *The 4th International Conference of Indonesian Society for Remote Sensing*, Makassar, 2018.
- [94] X. M. Y.-T. W. Fang-Shii Ning, "Low-Cost Receiver and Network Real-Time Kinematic Positioning for use in Connected and Autonomous Vehicles," *The Journal of Navigation*, vol. 72, no. 4, p. 917, 2019.
- [95] "LZERO: A Cost-Effective Multi-Purpose GNSS Platform," *Advanced Sensing Technology for Environment Monitoring*, vol. 22, 2022.
- [96] Y. G. Zhitao Lyu, "An SVM Based Weight Scheme for Improving Kinematic GNSS Positioning Accuracy with Low-Cost GNSS Receiver in Urban Environments," *GNSS Data Processing and Navigation in Challenging Environments*, vol. 20, 2020.

- [97] M.-C. P. E.-J. Y. D.-W. S. JEONG-KYU BAE, "Implementation and Performance Evaluation for DSRC-Based Vehicular Communication System," *IEEE Access*, 2020.
- [98] A. N. Muhammad Hassan, "An IoT based Structural Health Monitoring System for Critical Infrastructures," in *IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)*, 2021.
- [99] Y. Z. Yuan Zhou, "Design of Vehicle Remote Monitoring System Based on 4G and FlexRay," in *IEEE 18th International Conference on Communication Technology (ICCT)*, 2018.
- [100] L. Quectel Wireless Solutions Co., "EC20 Hardware Design".
- [101] S. A. N. A. J. Ashish Daniel Cherian, "Development of a Novel Embedded Board: CANduino," in *Intelligent Technologies and Robotics Intelligent Technologies and Robotics*, 2020.
- [102] A. J. Ha, "Passive Balancing Battery Management System for," California Polytechnic State University, 2018.
- [103] M. T. Inc., "MCP2561/2 High-Speed CAN Transceiver".
- [104] Analog Devices, Inc, "ADIS16460 Six Degrees of Freedom Inertial Sensor Datasheet".
- [105] S. L. Z. X. X. N. Quan Zhang, "Velocity-Based Optimization-Based Alignment," *IEEE Sensors*, vol. 20, no. 10, p. 5527, 2020.
- [106] G. L. M. G. D. L. R. W. Wenhao Zhao, "INS-assisted inter-system biases estimation and inter-system," *GPS solutions*, 2023.

- [107] Analog Devices, Inc., "ADIS16488 Tactical Grade Ten Degrees of Freedom Inertial Sensor Datasheet".
- [108] "THE DIRECT GEOREFERENCING APPLICATION AND PERFORMANCE ANALYSIS OF UAV HELICOPTER IN GCP-FREE AREA," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vols. XL-1/W4, p. 151, 2015.
- [109] M.-L. T. E.-S. N. A. H. a. C.-H. C. Kai-Wei Chiang, "A New Calibration Method Using Low Cost MEM IMUs to Verify the Performance of UAV-Borne MMS Payloads," in *UAV Sensors for Environmental Monitoring*, 2015.
- [110] M. Ali, "Compensation of temperature and acceleration effects on MEMS gyroscope," in *13th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, 2016.
- [111] www.allaboutcircuits.com, "What Is a Printed Circuit Board (PCB)?," [Online]. Available: <https://www.allaboutcircuits.com/technical-articles/what-is-a-printed-circuit-board-pcb/>.
- [112] B. Olney, "Multilayer PCB Stackup Planning," 2011.
- [113] I. o. P. Circuits, "Design Guide for High-Speed Controlled Impedance Circuit Boards," 1996.
- [114] c. halford, "Controlled Impedance".
- [115] PCBWay, "Impedance Calculation," [Online]. Available: https://www.pcbway.com/pcb_prototype/impedance_calculator.html.
- [116] T. Instruments, "HDMI Design Guide".

- [117] S. L. Inc, "AN0046: USB Hardware Design Guidelines".
- [118] M. T. Inc, "LAN9514 USB 2.0 Hub and 10/100 Ethernet Controller".
- [119] PC/104 Embedded Consortium, "PC/104 Specification Version 2.6," 2008.
- [120] Guido van Rossum, Python development team, "Python Tutorial".
- [121] NumPy Community, "NumPy User Guide".
- [122] D. D. E. F. M. D. John Hunter, "Matplotlib Release 2.0.2".
- [123] QGIS Project, "QGIS Desktop 3.16 User Guide".
- [124] K. B. F. H. D. T. Anders Nielsen, "An open source QGIS-based workflow for model application and experimentation with aquatic ecosystems," *Environmental Modelling & Software*, p. 258, 2017.
- [125] M. R. D. Meyer, "Open source QGIS toolkit for the Advanced Research WRF modelling system," *Environmental Modelling & Software*, p. 166, 2019.
- [126] " QGIS related repositories on GitHub - NSA," [Online]. Available: <https://github.com/NationalSecurityAgency?q=qgis>. [Accessed 12 11 2023].
- [127] " QGIS related repositories on GitHub - LINZ," [Online]. Available: <https://github.com/linz?q=qgis>. [Accessed 15 11 2023].
- [128] R. Richmond, "Digital Help for Haiti," 27 1 2010. [Online]. Available: <http://gadgetwise.blogs.nytimes.com/2010/01/27/digital-help-for-haiti/>.

- [129] "Launching the Facebook Map"., 25 2 2021. [Online]. Available: <https://hi.stamen.com/launching-the-facebook-map-8d028c4f0e0e>.
- [130] Apple, "Apple Maps Legal Information," [Online]. Available: <https://www.apple.com/legal/internet-services/maps/legal-en.html>. [Accessed 12 11 2023].
- [131] D. S. L. P. Jennings Anderson, "Corporate Editors in the Evolving Landscape of OpenStreetMap," *ISPRS International Journal of Geo-Information*, vol. 8, no. 5, p. 232, 2019.
- [132] B. L. S. P. d. A. J. e. a. Herfort, "The evolution of humanitarian mapping within the OpenStreetMap community," *Sci Rep*, vol. 11, p. 3037, 2021.
- [133] S. S. D. T. a. A. X. F. J. E. Vargas-Munoz, "OpenStreetMap: Challenges and Opportunities in Machine Learning and Remote Sensing," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 1, p. 184, 2021.
- [134] M. C. S. H. B. J. Feldmeyer D, " Using OpenStreetMap Data and Machine Learning to Generate Socio-Economic Indicators," *ISPRS International Journal of Geo-Information*, vol. 9, no. 9, p. 498, 2020.
- [135] H. Q. J. W. a. J. L. J. Li, "OpenStreetMap-Based Autonomous Navigation for the Four Wheel-Legged Robot Via 3D-Lidar and CCD Camera," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 3, p. 2708, 2022.
- [136] RTKLIB , "RTKLIB ver. 2.4.2 Manual".
- [137] L. K. H. K. Christian ELING, "Development of an RTK-GPS system for precise real time positioning of lightweight UAVs," in *Internationalen Ingenieurvermessungskurs*, 2014.

- [138] A. Y. Tomoji Takasu, "Development of the low-cost RTK-GPS receiver with an open source program," in *Tokyo University of Marine Science and Technology*, 2009.
- [139] T. Everett, "Smartphone Decimeter Challenge: An RTKLIB Open-Source Based Solution," in *35th International Technical Meeting of the Satellite Division of The Institute of Navigation*, 2022.
- [140] M. A. H. I. H. W. Eibe Frank, "WEKA Workbench".
- [141] M. J. Swasti Singhal, "A Study on WEKA Tool for Data Preprocessing,," in *International Journal of Innovative Technology and Exploring Engineering*, 2013.
- [142] T. M. S. Y. Mohd Fauzi bin Othman, "Comparison of Different Classification Techniques Using WEKA for Breast Cancer," in *3rd Kuala Lumpur International Conference on Biomedical Engineering*, 2016.
- [143] H. S. J. K. T. K. Bo Wei, "RSSI-CSI Measurement and Variation Mitigation with Commodity WiFi Device".
- [144] T. W. Y. L. W. D. H. O. J. Wang, "Modeling and performance analysis of dynamic spectrum sharing between DSRC and Wi-Fi systems," *Wireless Commun. Mobile Comput*, vol. 2743, no. 16, p. 16, 2016.
- [145] C. L. Y. Z. D. Y. M. Shi, "DSRC and LTE-V communication performance evaluation and improvement based on typical V2X application at intersection," in *Proc. Chin. Autom. Congr.*, 2017.
- [146] X. C. H. H. R. Xiaomin Ma, "Performance and Reliability of DSRC Vehicular Safety Communication: A Formal Analysis," *EURASIP Journal on Wireless Communications and Networking*, vol. 1, no. 13, 2008.

- [147] Cisco, "Antenna Patterns and Their Meaning".
- [148] L. Geosystems, "GPS Reference Stations and Networks: An introductory guide".
- [149] UK Government Department for Transport, "The Highway Code - Introduction," [Online]. Available: <https://www.gov.uk/guidance/the-highway-code/introduction>.
- [150] M. S. M. A. K. S. A. M. S. A. K. Imran Israr, "Path Loss Modeling of WLAN and WiMAX Systems," *International Journal of Electrical and Computer Engineering*, vol. 5, no. 5, pp. 1083-1091, 2015.
- [151] University of Hawaii, "ANTENNA INTRODUCTION / BASICS".
- [152] L-com, "Choosing the Right WiFi Antenna for your Application".
- [153] S. Filtikakis, "Performance analysis of IEEE 802.11g signals under different operational environments," 2005.
- [154] R. H. a. R. V. P. Kishor Chandra Joshi, "Association, Blockage and Handoffs in IEEE 802.11ad based 60 GHz Picocells- A Closer Look," 2019.
- [155] J. K. M. S. Sukhdeep Kaur, "Antenna Diversity Techniques," *International Advanced Research Journal in Science, Engineering and Technology*, vol. 3, no. 4, 2016.
- [156] D. B. Ibrahim Heydarov, "Experimental Evaluation of the Shadowing of a Planar Antenna Caused by a Quadcopter Frame," *Electrical, Control and Communication Engineering*, vol. 16, no. 1, pp. 37-43, 2020.
- [157] P. M. TSAI, "Path-loss and Shadowing," National Taiwan University.
- [158] Rohde & Schwarz, "802.11 Packet Error Rate Testing Application Note".

- [159] L. Geosystems, "Leica Viva GS10 Data sheet".
- [160] OXTS, "What is RTK?," 13 10 2020. [Online]. Available: <https://www.oxts.com/rtk/>.
- [161] B. K. M. J. Samer Khanafseh, "GNSS Multipath Error Modeling for Automotive Applications," in *31st International Technical Meeting of The Satellite Division of the Institute of Navigation*, 2018.
- [162] C. i. C. Ltd, "GPS ACTIVE 28dB MAGNETIC ANTENNA Datasheet".
- [163] G. WORLD, "2017 GNSS Antenna Survey".
- [164] C. Hegarty, "A Brief History of GPS L5," MITRE Corporation, 2010.
- [165] B. W. X. L. J. H. H. L. X. H. Xingxing Li, "Principle and performance of multi-frequency and multi-GNSS PPP-RTK," *Satellite Navigation*, vol. 3, no. 7, 2022.
- [166] ETSI, "3GPP TS 38.101-3 version 15.2.0 Release 15," 2018.
- [167] A. P. SODERINI, "OUTDOOR POSITIONING ALGORITHMS BASED ON LTE AND WIFI MEASUREMENTS," Tampere University of Technology, 2016.
- [168] J. D. Chee, "Pearson's Product-Moment Correlation: Sample Analysis," The Queen's Medical Center, 2015.
- [169] X. Ying, "An Overview of Overfitting and its Solutions," *Journal of Physics Conference*, vol. 2, p. 1168, 2019.
- [170] C. K. Nongnuch Poolsawad, "Balancing Class for Performance of Classification with a Clinical Dataset," 2014.
- [171] S. Kotsiantis, "Handling imbalanced datasets: A review," 2005.

- [172] K. W. B. Nitesh V. Chawla, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.
- [173] Y. B. E. A. G. Haibo He, "ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning," in *IEEE World Congress on Computational Intelligence*, 2008.
- [174] N. Z. N. W. Wen Zhu, "Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations," 2010.
- [175] L. E. T. ,. H. U. L. I. P AYAM R EFAEILZADEH, "Cross-Validation," Arizona State University, 2008.
- [176] D. Berrar, "Cross-Validation," 2018.
- [177] B. R. A. R. E. v. d. K. Sofia Visa, "Confusion Matrix-based Feature Selection.," in *22nd Midwest Artificial Intelligence and Cognitive Science Conference*, 2011.
- [178] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation," 2011.
- [179] D. L. D. C, "Effectiveness Analysis of ZeroR, RIDOR and PART Classifiers for Credit Risk Appraisal," 2014.
- [180] H. v. Jouanne-Diedrich, "Package 'OneR': One Rule Machine Learning Classification Algorithm with Enhancements," 2017.
- [181] S. J. D. Padraig Cunningham, "k-Nearest neighbour classifiers," *ACM Computing Surveys*, 2007.

- [182] O. M. Lior Rokach, "Decision Trees," in *The Data Mining and Knowledge Discovery Handbook*, 2005, pp. 165-192.
- [183] N. Ramakrishnan, "C4.5 Algorithm," 2009.
- [184] The Schank Academy , "J48 Classifier Parameters".
- [185] E. FRANK, "Reduced-error pruning with significance tests," University of Waikato.
- [186] S. S. Haykin, "Neural networks : a comprehensive foundation," Prentice Hall, 1999.
- [187] Egmont-Petersen, M.; de Ridder, D.; Handels, H., "Image processing with neural networks – a review," *Pattern Recognition*, no. 35, pp. 2279-2301, 2002.
- [188] M. Levin, "Cost-benefit analysis of fuel-efficient speed control using signal phasing and timing (SPaT) data: evaluation for future connected corridor deployment," University of Minnesota, 2023.
- [189] Savari, "Savari SW-1000 Road-Side-Unit (RSU)".