

**MACHINE LEARNING-BASED DETECTION AND DIGITAL
FORENSIC RECONSTRUCTION OF GPS SPOOFING
ATTACKS ON UNMANNED AERIAL VEHICLES**

TABLE OF CONTENTS

Chapter 1: Introduction	5
1.1 Background of the Study	5
1.2 Problem Statement	5
1.3 Research Aim	6
1.4 Research Objectives	6
1.5 Research Questions	6
1.6 Scope of the Study	7
1.7 Significance of the Study	7
1.8 Dissertation Structure	7
Chapter 2: Literature Review	9
2.1 Introduction to the Literature Review	9
2.2 Unmanned Aerial Vehicles (UAVs) and Navigation Systems	9
2.2.1 Fundamentals of UAV Technology	9
2.2.2 GPS-Based Navigation in UAVs	9
2.2.3 Critically Discuss	10
2.3 GPS Spoofing Attacks on UAVs	11
2.3.1 Definition of GPS Spoofing	11
2.3.2 Types of GPS Spoofing Attacks	11
2.3.3 Impact of GPS Spoofing	11
2.3.4 Critically Discuss	12
2.4 Machine Learning for GPS Spoofing Detection	12
2.4.1 Machine Learning in Cybersecurity	12
2.4.2 Supervised Machine Learning Algorithms	13
2.4.3 Unsupervised Learning Algorithms	13
2.4.4 Deep Learning Approaches	14
2.4.5 Critically Evaluate	15

2.5 Digital Forensic Reconstruction of GPS Spoofing Incidents	15
2.5.1 Digital Forensics in UAV Investigations	15
2.5.2 Digital Forensic Process	15
2.5.3 Digital Forensic Reconstruction Techniques	16
2.5.4 Critically Evaluate	16
2.6 Existing Detection Frameworks and Security Solutions	17
2.6.1 Traditional Detection Methods	17
2.6.2 AI-Based Detection Frameworks	17
2.6.3 Critically Compare	17
2.7 Research Gaps in Existing Literature	18
2.8 Conceptual Framework	19
2.9 Chapter Summary	20
Chapter 3: Research Methodology	21
3.1 Introduction	21
3.2 Research Onion	21
3.3 Research Philosophy	22
3.4 Research Approach	23
3.5 Research Strategy	23
3.6 Research Design	24
3.7 Dataset Description	24
3.8 Data Collection Method	25
3.9 Data Pre-processing Plan	25
3.10 Machine Learning Methodology	25
3.11 Evaluation Metrics	26
3.12 Software and Tools	26
3.13 Ethical Considerations	27
3.14 Chapter Summary	27

Chapter 4: System Design & Implementation	29
4.1 Introduction	29
4.2 System Architecture	29
4.3 Development Environment	29
4.4 Dataset Loading and Initial Exploration	30
4.5 Data Pre-processing and Feature Engineering	31
4.6 Exploratory Data Analysis (EDA)	34
4.6.1 GPS Spoofing Class Distribution	34
4.6.2 Signal Strength Distribution	35
4.6.3 Communication Protocol Distribution	35
4.6.5 Feature Boxplots	37
4.7 Machine Learning Model Development	38
4.8 Model Optimisation	40
4.9 Chapter Summary	41
Chapter 5: Results & Discussion	42
5.1 Introduction	42
5.2 Exploratory Data Analysis Results	43
5.2.1 GPS Spoofing Class Distribution	43
5.2.2 Signal Strength Distribution	43
5.2.3 Communication Protocol Distribution	43
5.3 Machine Learning Performance Results	44
5.3.1 Logistic Regression Results	44
5.3.2 Decision Tree Results	45
5.3.3 Random Forest Results	45
5.3.4 XGBoost Results	46
5.3.5 Comparative Performance Analysis	46
5.4 Hyperparameter Optimisation Results	46

5.5 Feature Importance Analysis	47
5.6 Decision Tree Interpretation	48
5.7 Discussion of Findings	49
5.8 Chapter Summary	50
Chapter 6: Conclusion & Future Work	51
6.1 Introduction	51
6.2 Summary of Research	51
6.3 Achievement of Research Aim and Objectives	52
6.3.1 Achievement of Aim	52
6.3.2 Objective 1	52
6.3.3 Objective 2	52
6.3.4 Objective 3	52
6.3.5 Objective 4	52
6.4 Key Contributions of the Study	52
6.5 Limitations of the Study	53
6.6 Recommendations for Future Work	53
6.7 Final Conclusion	53
Reflection	55
Reference List	57
Appendices	65
Appendix 1: Dataset Link	65
Appendix 2: Google Colab Link	65
Appendix 3: Full Python Code	66

Chapter 1: Introduction

1.1 Background of the Study

Unmanned Aerial Vehicles (UAVs) have moved from “distant piloting” into a fully autonomous system which can be used for surveillance, logistics, disaster responses and military operations. This technological advancement has not only contributed to smoother working experiences for the machines, but it has at the same time added to the cyberattack surface of autonomous navigation. As argued by Shafik, Matinkhah and Shokoor (2023), the increased reliance on interconnected communication and navigation technologies has driven cybersecurity concerns, reflecting the fact that system component improvements do not necessarily result in system robustness. The ubiquitous use of “Global Positioning System (GPS)” navigation further enhances the accuracy of a flight's position, but creates vulnerabilities readily abused by powerful attackers.

As the “GPS spoofing” attacks have become more sophisticated, though, the conventional wisdom about the security of navigating with GPS has been called into question. Alhoraibi, Alghazzawi and Alhebshi (2024) proposed that adversarial manipulation can mimic the proper satellite signals, which permitted the attackers to change the trajectories of the UAVs without the traditional verification procedures. It's a reminder that navigation is not enough for cybersecurity. Additionally, the incorporation of “machine learning” has shown itself as a proactive countermeasure which can also be used to identify abnormal behaviour when implementing UAV security rather than using a set of “attack signatures”. Shafique, Mehmood and Elhadeif (2021) have stated that intelligent detection models give it more flexibility in the face of the changing tactics of the adversaries, and that's why there should be robust cybersecurity strategies. Therefore, there is a need to develop new security technologies for autonomous UAVs that combine advanced detection and robust cyber defence to combat increasingly dynamic navigation threats.

1.2 Problem Statement

The rise in use of autonomous UAVs has revealed critical vulnerabilities, as traditional techniques designed to protect navigation from GPS attacks are unable to detect advanced forms of "GPS spoofing" which are capable of interfering with the highly precise navigation of UAVs and other automation systems. In complex operational scenarios, existing navigation verification methods may lose effectiveness due to coordinated attacks, as mentioned in the

statement by Mykytyn *et al.* (2023). As such, traditional detection techniques are essentially defensive and cannot react well to an adaptive attack style. Moreover, any chance that civil user behaviour is compromised is also covered by intelligent cyber threats, which require analytical models capable of detecting the subtle anomalies in the user's behaviour. While machine learning offers higher adaptability in recognising new attack surface patterns, there must also be reliable evidence reconstitution should an attack be detected, as suggested by Khan *et al.* (2021). Digital forensic reconstruction" will reinforce the post-incident analysis process with credible evidence trails and facilitate comprehensive investigation, leading towards enhanced UAV cyber security resilience, as argued by Lee, Seo and Kim (2023).

1.3 Research Aim

This study aims to present and test a "machine learning"-based framework for identifying "GPS spoofing" attacks on "Unmanned Aerial Vehicles (UAVs)" and assist "digital forensic reconstruction" after attacks to increase cybersecurity resilience, detection accuracy and post-incident investigation.

1.4 Research Objectives

The research objectives are:

- To explore the nature and features of and the effects of "GPS spoofing" attacks on autonomous UAV navigation.
- Implement a machine learning model to detect GPS spoofing attacks from the dataset.
- To analyse the detection model based on suitable machine learning performance measures.
- To facilitate the "digital forensic reconstruction", which is the process of correlating detection results with evidence of an attack when needed to assist in better investigation.

1.5 Research Questions

The following questions will be explored in the research:

1. How do "GPS spoofing" attacks affect the security and navigation of autonomous UAVs?
2. What is the detection accuracy of a "machine learning" model for GPS spoofing attacks?

3. What metrics are best for measuring the performance of the proposed detection model?
4. What can “digital forensic reconstruction” contribute after the attack takes place?

1.6 Scope of the Study

The main objective of this study is to analyse the dataset provided with the UAV communication to build a "machine learning" model to detect "GPS spoofing" attacks and evaluate it. The study was mainly aimed at data preprocessing, feature analysis, model development and performance measurement based on well-renowned classification parameters. In addition, this study explores the possibility of provides the assistant of the "digital forensic reconstruction" to help correlate evidence in cyberattack investigations. Alhoraibi *et al.* (2024) stated that intelligent detection provides greater resistance to the changing spoofer activity. For this reason, hardware security mechanisms, real-time deployment of UAVs, and other cyberattacks are not part of this project.

1.7 Significance of the Study

This study adds to the current literature and extends existing research on incorporating “machine learning” into “digital forensic reconstruction”— to detect “GPS spoofing” attacks on autonomous UAVs. The research does not see detection and forensic investigation as wholly separate activities, but as one that adds to the other to improve cyber resilience. Periodic analysis of digital evidence is more reliable, which is why it is crucial to incorporate post-incident investigations in UAV cybersecurity studies (Lee *et al.* 2023). In terms of implementation, the proposed framework helps spread the responsibility for security analysis across the framework by providing more opportunities for attack identification and thereby improving evidence interpretation upon navigation compromise. Therefore, this work could help researchers and practitioners in designing more resilient systems, which would be applied to UAV security, as well as the promotion of future deployment of autonomous smart detection models to enhance operational security and forensics capabilities in the field of UAV operations.

1.8 Dissertation Structure

This dissertation consists of a reflection in addition to six chapters. The research background, objectives and scope are included in Chapter 1. A critical review of the existing literature and identification of gaps in the literature are conducted in Chapter 2. The research methodology

used is explained in Chapter 3. In Chapter 4, the system design and implementation using Python in Google Colab is discussed. The experimental results are analysed and discussed in Chapter 5. The research and Project recommendations for future work are complete in Chapter 6, and the Reflection, which examines the knowledge, technical skills and professional development gained during the Project, is included at the end.

Chapter 2: Literature Review

2.1 Introduction to the Literature Review

This literature review critically analyses the published studies on “GPS spoofing” attacks on “Unmanned Aerial Vehicles (UAVs)” and the role played by the techniques of “machine learning” and “digital forensic reconstruction” in enhancing UAV cybersecurity. As GPS navigation becomes more critical to the operation of autonomous UAVs, spoofing presents an evolving threat and cybersecurity challenge to ensure flight integrity and operational safety. Some recent proposals have suggested new ways of detection, but to date, the effectiveness of these attacks has been questioned due to the effectiveness of the classical navigation protection mechanism for detecting sophisticated spoofing attacks. As a result, researchers have turned to intelligent machine learning algorithms that can detect abnormal navigation patterns as well as methods of digital forensic analysis that provide the means to become a post-collision investigator of the attacks. Thus, the literature review for UAV navigation, GPS spoofing threats, detecting intelligent spoofing methods and forensic reconstruction is critically discussed and used to identify "Research gaps" that motivate the proposed research and help achieve the dissertation goals.

2.2 Unmanned Aerial Vehicles (UAVs) and Navigation Systems

2.2.1 Fundamentals of UAV Technology

The term “Unmanned Aerial Vehicles (UAVs)” has been consistently identified by researchers as platforms that work without manned operators to assist civilian and military missions, but they debate whether security development has helped the platforms grow more autonomous. Others have suggested that with the growth of autonomous capabilities, the surface of attack for cyberattacks also grows larger, as navigation systems continue to become even more interconnected (Tang, 2021). On the contrary, Sulieman *et al.* (2025) proposed that, apart from increasing operational resilience, technological developments have had only limited effects on security. At the same time, however, Jung, Hong and Yoon (2025) made it clear that more autonomy makes them more vulnerable to advanced cyberattack methods based on navigation, demonstrating that technological advances do not necessarily make UAVs more secure.

2.2.2 GPS-Based Navigation in UAVs

GPS is still considered the dominant navigation system for supporting UAV positioning by seeking to provide an integrated system known as “Inertial Navigation System (INS)” and using flight control systems and on-board sensors. According to Sulieman *et al.* (2025), the integration of INS can significantly enhance the continuity of navigation for positioning purposes, addressing deviations that may occur, at least for short durations. Tang (2021), however, pointed out that such integration is not enough to make satellites redundant. To support this contention, Jung *et al.* (2025) tested the various sensors in the architecture of an RST with advanced spoofing attacks, the results of which show that sophisticated spoofing attacks can still change the navigation output, implying that navigation redundancy does not mean cybersecurity resilience.

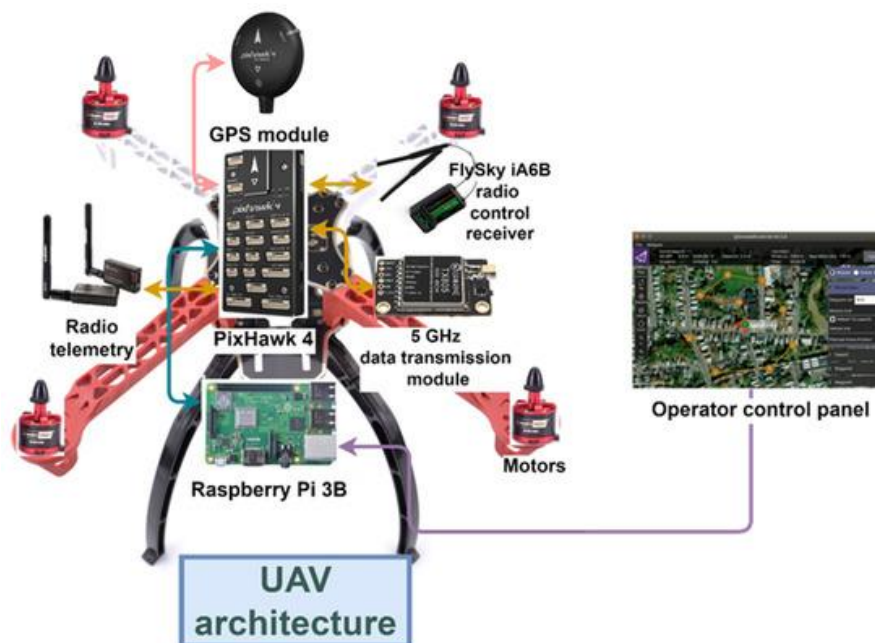


Figure 2.1: GPS-Spoofing Attack Detection Technology for UAVs

(Source: Basan *et al.* 2021)

2.2.3 Critically Discuss

Previous research generally concurred that GPS can improve the accuracy of autonomous flight, but had mixed conclusions about how well GPS would stand up to intentional radio jamming. Tang (2021) highlighted operational efficiency, while Sulieman *et al.* (2025) considered sensor integration as a good resilience mechanism. In contrast, Jung, Hong and Yoon (2025) have pointed out that even with integrated navigation systems, coordinated attack techniques can overcome the navigation systems due to the ongoing reliance on positioning derived from satellite navigation. Based on the present state of the art, it can be concluded that

hardware redundancy is inadequate and current evidence indicates that intelligent machine learning-based detection systems to uncover advanced spoofing behaviour are required.

2.3 GPS Spoofing Attacks on UAVs

2.3.1 Definition of GPS Spoofing

The term 'GPS spoofing' is consistently identified as the sending of fake satellite signals in a transmission, but the explanation of the threat varies between researchers. According to Ferrão *et al.* (2020), “spoofing” is a technique that involves tricking the navigation system by injecting false location data while signals may be received as usual. However, Bose (2021) contended that its ability to mimic real satellite operation with great fidelity means it's hard to detect as a deceptive satellite. Jung, Hong and Yoon (2025) further affirmed that contemporary attacks on UAVs are based not only on changing the location of the UAVs but also on a deception-based criterion that seeks to affect their autonomous decision-making processes.

2.3.2 Types of GPS Spoofing Attacks

The relevant literature shows that the current attacks can be divided into two major categories: meaconing, signal replay, and counterfeit signal generation/directed signal replay, with the former operation being the easiest and the latter the most sophisticated. Ferraro *et al.* (2020) proposed a simple but powerful replay attack based on the limited sophistication of previously captured navigation signals. But Bose (2021) discussed the possibility that fake signals pose a bigger challenge in terms of cybersecurity, as they closely resemble authentic satellite signals. To support this, Jung *et al.* (2025) showed that coordinated spoofing is able to progressively evade the navigation behaviour, thereby decreasing the probability of detection for a conventional UAV security system.

2.3.3 Impact of GPS Spoofing

It is universally recognised that GPS spoofing poses a threat to the mission integrity of UAVs, but there are differing views on the consequences for operations. Gaining student attention was the main theme of spoofing that was observed by Ferrão *et al.* (2020), according to their research. Bose (2021) proposed that it is possible that over time, they may lose all of their signal control and fail their mission. Similarly, increasingly independent UAV operations through incorrect navigation data can result in unsafe navigation decisions before necessary

human supervision, further exacerbating public safety-related effects (Jung, Hong and Yoon 2025).

2.3.4 Critically Discuss

There is comparative evidence that the techniques of spoofing are not just about manipulating signals; in fact, highly coordinated attacks have evolved that enable spoofed signals to bypass traditional methods of navigation protection. Basic adversarial attempts to spoof a signal fall under the category of traditional validation mechanisms, which Ferrão *et al.* (2020) deemed to be reasonably effective, while Bose (2021) doubted their effectiveness for intelligent adversarial signal emulation. In support of this, Jung, Hong and Yoon (2025) reinforced this criticism by stating that the recent attacks were becoming more sophisticated and thus conventional attack detection and defence strategies were ineffective, highlighting the need for advanced detection techniques such as machine learning that can detect subtle behavioural anomalies.

2.4 Machine Learning for GPS Spoofing Detection

2.4.1 Machine Learning in Cybersecurity

Cyber threats and attacks posed by UAVs are becoming more complex, and cybersecurity functions are transitioning from signature-based protection to "Artificial Intelligence (AI)" and "Machine Learning (ML)" types that can identify more advanced behavioural anomalies. Shrestha *et al.* (2021) asserted that intelligent intrusion detection could provide cybersecurity benefits for UAVs by detecting abnormal network behaviours before they affect their operations. According to Eshmawi *et al.* (2024), ensemble machine learning is the use of different algorithms that work in tandem to detect subtle navigation errors that traditional security solutions don't catch. Nevertheless, Nayfeh *et al.* (2023) found that it is not only the complexity of its algorithms that matters for successful cyber defence, but also the choice of the GPS features that are informative. This holistic approach was further championed by Sun *et al.* (2023), who proposed that intelligent detection must address both spatial and temporal signal properties, underscoring the need for compelling feature engineering and adaptive learning models, features that are essential for ensuring secure UAV operations.

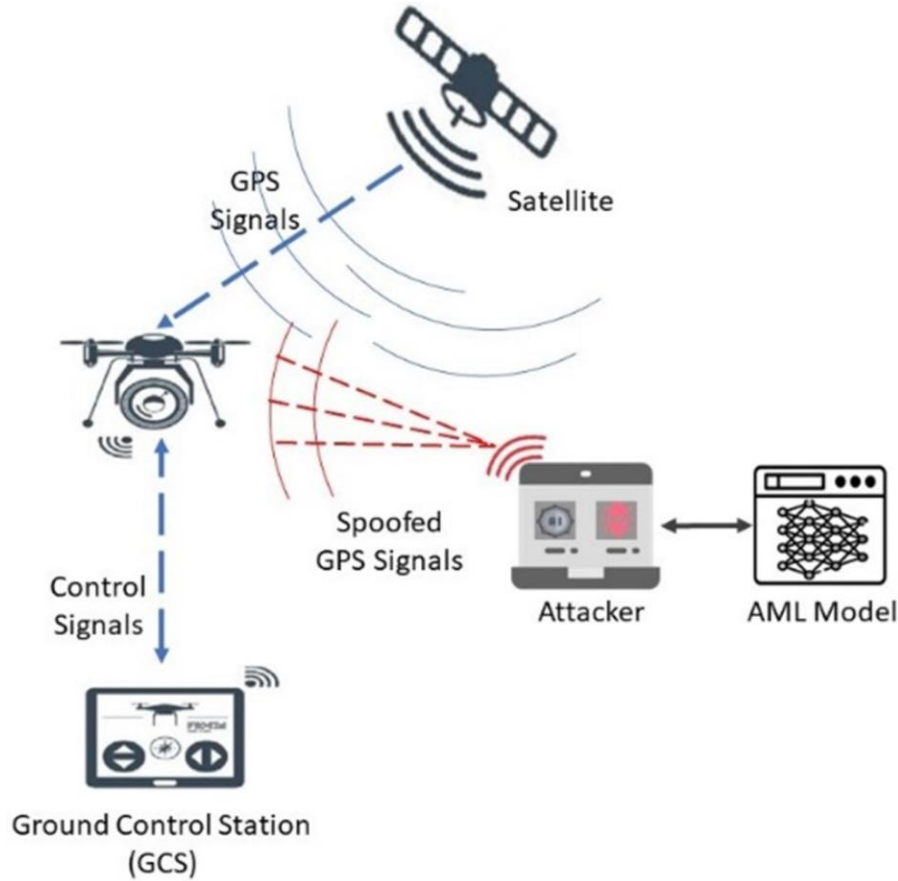


Figure 2.2: Machine Learning for GPS Spoofing Detection

(Source: Alhoraibi *et al.* 2024)

2.4.2 Supervised Machine Learning Algorithms

Current algorithms for GPS spoofing detection, such as "Random Forest", "Support Vector Machine", "Decision Tree", "Logistic Regression" and "XGBoost", are still the most prevalent due to their ability to learn attack characteristics based on a set of labelled data. Eshmawi *et al.* (2024) observed that ensemble methods are superior to single classifiers due to their ability to enhance the stability of the classifiers as well as to minimise classification errors in general. Nayfeh *et al.* (2023) claimed, however, that the classification results do not depend only on the type of classifiers, but also on the representation of the features. Based on the above comparison, it is reasonable to believe that any algorithm that is complex would not guarantee reliable detection ability in various UAV operating environments if the feature extraction is not optimised.

2.4.3 Unsupervised Learning Algorithms

However, the unsupervised methods (including "K-Means", "Isolation Forest", "DBSCAN" and "Autoencoders") do not need attack data as inputs to detect anomalies. Consequently, Nayfeh *et al.* (2023) proposed that anomaly-based learning is better suited to combat the resilient behaviour of new activity since it does not rely on feature-based training techniques but rather focuses on the anomaly from features existing during the navigation. But when representative datasets exist, supervised ensemble learning has the advantage of higher operational reliability, as suggested by Eshmawi *et al.* (2024). In addition, Sun *et al.* (2023) showed that complex patterns often need to be more complex and detailed than standard clustering methods can achieve, a reason why standalone technologies, like unsupervised ones, are limited in effectiveness.

2.4.4 Deep Learning Approaches

Deep learning architectures, such as Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM) and Gated Recurrent Units (GRU), are gaining popularity due to their ability to automatically learn complex relationships in navigation data. Sun *et al.* (2023) noted that deep learning is quite effective in that it can be used to gain spatial and temporal features from the signals without much manual engineering. In contrast, Eshmawi *et al.* (2024) claimed that well-designed ensemble machine learning models can achieve similar detection performance with less computation. This has led to the literature being split between achieving the best prediction performance and keeping the computational complexity of real-time deployment of UAVs down.

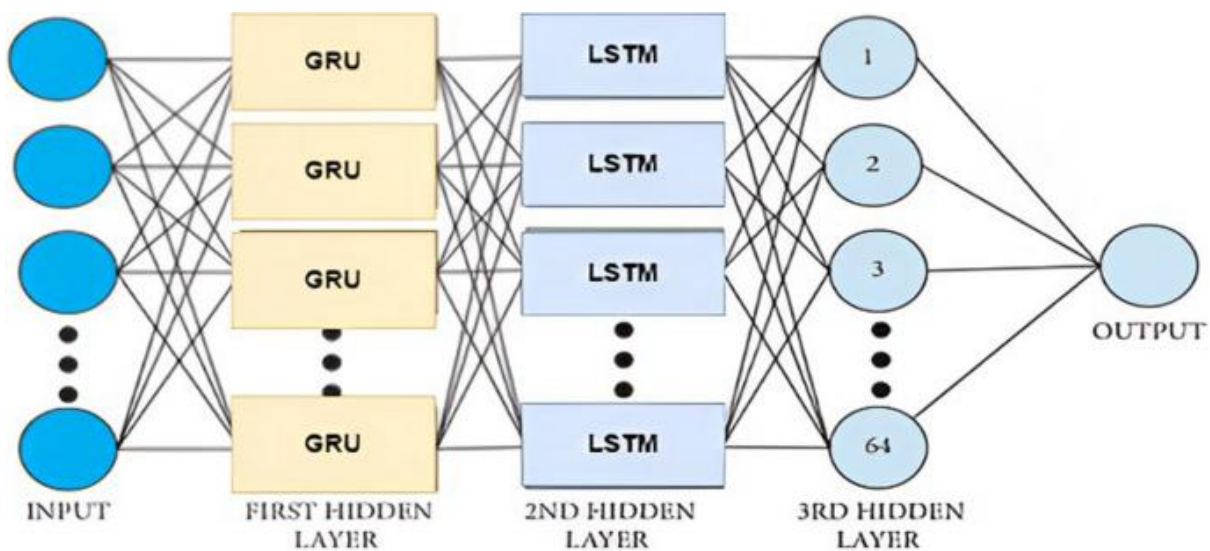


Figure 2.3: CNN-GRU-LSTM-based deep learning model

(Source: Singh *et al.* 2025)

2.4.5 Critically Evaluate

All evidence shows that none of the learning paradigms can be consistently the best choice across various operational environments. Eshmawi *et al.* (2024) showed that ensemble learning not only provides high accuracy but is also a low-complexity approach, and Sun *et al.* (2023) showed that deep learning is more powerful in modelling the complex behaviour of spoofing. However, Nayfeh *et al.* (2023) suggested that either method is highly susceptible to the quality of the dataset as well as feature selection, thus being unable to directly apply the results to real-world attacks. Namagembe *et al.* (2026) noted that sufficient GPS spoofing countermeasures rely on machine learning algorithms that can provide high accuracy while maintaining a timely operation. The lack of existing studies on explainability and real-world deployment reveals a need for machine learning systems that optimise both false-positive and confusion rates while maintaining accessibility in operational scenarios for UAV cybersecurity.

2.5 Digital Forensic Reconstruction of GPS Spoofing Incidents

2.5.1 Digital Forensics in UAV Investigations

Flight logs, sensor data, and telemetry are all considered forms of what has come to be known as “digital evidence” and play a major role in the digital forensic investigations used to reconstruct UAV incidents. According to Studiawan, Grispos and Choo (2023), the investigators are able to determine the sequences of the attacks and the systems' compromise by collecting the complete evidence. Alotaibi, Al-Dhaqm and Al-Otaibi (2023) noted that evidence must be accompanied by a structured forensic model that guarantees that evidence is not lost, not duplicated and not compromised. Mohammed *et al.* (2025) also argued that forensic readiness ought to be integrated into UAV architectures, which requires quick and accurate evidence collection after cyber attacks and thus enhances forensic investigations.

2.5.2 Digital Forensic Process

The digital forensic process typically involves activities of evidence acquisition, preservation, analysis, reconstruction and reporting; however, among the researchers, there seems to be some variation in the order of priority for these activities. Alotaibi, Al-Dhaqm and Al-Otaibi (2023) explained that systematic forensic procedures increase the uniformity of the investigation and also improve the legal admissibility of the investigation. On the other hand, Studiawan *et al.* (2023) claimed that there are insufficiencies in the developed frameworks regarding the characteristics of the evidence that can be obtained from UAV missions, especially flight data

that is volatile. Therefore, Mohammed *et al.* (2025) recommended mainstreaming forensic readiness in the operational environment to minimise the loss of evidence and thus enhance the efficiency and reliability of investigations conducted after the occurrence of a crime.

2.5.3 Digital Forensic Reconstruction Techniques

Digital forensic reconstruction gathers this data to create a timeline of events, correlation of events, log analysis and sensor fusion to help reconstruct the sequence of cyberattacks that impacted UAV operations. Studies by Studiawan, Grispos, and Choo (2023) have shown that there is a higher level of reliability in reconstructing each incident when multiple sources of evidence are correlated than when each artefact is analysed independently. In the UAV forensic case study conducted by Sibe and Bekom (2025), it is shown that in-depth examination of flight logs can have a significant impact on evidence reconstruction and attack attribution. The quality of the reconstruction, on the other hand, is subjective and relies heavily on structured forensic methodology (Alotaibi *et al.* 2023). Mohammed *et al.* (2025) further stated the principle that using a wide variety of sources to assist with evidence supports the validity of the evidence being credited, and supports the detective and investigative aspect of the work after high-end spoofing attempts.

2.5.4 Critically Evaluate

Current forensic models offer an investigative approach that relies on structure, but practical constraints have ubiquitously harmed the efficiency of these models in support of dynamic UAV environments. Students of Studiawan, Grispos and Choo (2023) have claimed that existing solutions are not sufficiently dealing with volatile digital evidence created during autonomous flight. Alotaibi, Al-Dhaqm and Al-Otaibi (2023), on the other hand, focused on methodological standardisation without a strong operational validation. Mohammed *et al.* (2025) further noted that in order to make more reliable evidence traces, there is a need for a more forensic-ready framework for UAVs in the future, allowing for an operational deployment and for the management of evidence as it is present in a legally acceptable manner. As seen by Sihag *et al.* (2023), the integration of Cybersecurity and the capabilities of Digital Forensics will enhance Incident Response throughout the UAV attack lifecycle.

2.6 Existing Detection Frameworks and Security Solutions

2.6.1 Traditional Detection Methods

The traditional methods of GPS spoofing detection are based on abnormal positioning behaviour, which includes monitoring of signal strength, receiver consistency checks, and navigation verification. According to Nayfeh *et al.* (2023), these types of methods offer the first level of defence as inconsistencies or faults in the characteristics of GPS signals are detected before they can significantly impact navigation. Aldaej *et al.* (2022), however, stated that traditional mechanisms continue to be reactive as they rely on thresholds that can be deliberately avoided in more advanced spoofing attacks. Eshmawi *et al.* (2024) also proposed that dynamic verification techniques have become more challenging to implement, as some static verification methods have proven less adaptable to a dynamic UAV context.

2.6.2 AI-Based Detection Frameworks

AI support of cyber defence using hybrid detection, behaviour analysis, sensor fusion and multi-layer defence is increasingly appearing in recent studies as a means to bolster UAV cyber security. The multi-intelligent security layers, as proposed by Aldaej *et al.* (2022), can provide a higher level of resilience by detecting behaviour abnormalities in multiple components of a connected system, not only in the navigation signal. On the other hand, Nayfeh *et al.* (2023) claimed that the detection of effectiveness relies greatly on the use of representative GPS properties that can distinguish between authentic and illegitimate actions. In line with that, Eshmawi *et al.* (2024) proposed a methodology that integrates the different detection approaches to achieve better adaptability toward evolving spoofing techniques. Furthermore, Squatrito (2024) has shown that machine learning works well to discriminate GPS spoofing and jamming patterns using a behavioural approach on navigation signals.

2.6.3 Critically Compare

Experimental results have shown that current detection systems are still fairly efficient in terms of computations and are comparatively easy to implement, but are still questioned in terms of scalability versus the ever-changing methods of spoofing. Machine learning frameworks significantly outperform their traditional alternatives by learning and adapting to complex behavioural patterns that are not obvious in the data, as mentioned by Eshmawi *et al.* (2024). The Aldaej *et al.* (2022) work, however, suggested that this implies a greater amount of computational load and complexity of implementation, making them unsuitable for use on

resource-constrained UAV platforms. The research conducted by Nayfeh *et al.* (2023) thus recommended that a blend of the two methods, namely conventional and intelligent verification, can offer a practical compromise between the efficiency of operations, their scalability, and feasibility in practical applications, which will aid resilient UAV cybersecurity.

2.7 Research Gaps in Existing Literature

With all the great strides made in the last few years in detecting the spoofing of GPS, some shortcomings have not yet been addressed and are problematic in the autonomous UAV context. While ensemble machine learning offers significant accuracy in detection, the primary evaluation method is via controlled experimental datasets, which leaves the question of model robustness in real-world operational conditions, as put by Eshmawi *et al.* (2024). Similar arguments from Nayfeh and his colleagues (2023) suggest that detection performance is strongly tied to the quality of the dataset used and how the features are represented, meaning that without standardised datasets of UAV spoofing, it is difficult to make meaningful comparisons between competing algorithms. Beyond that, Studiawan, Grispos and Choo (2023) reported that most research in digital forensics has concentrated on the identification of the evidence, with little effort made on the full digital forensic reconstruction of the events and the identification of the attack.

One limitation is the lack of explainability and deployment of the operations. Current machine learning algorithms often emphasise prediction but remain opaque about how these systems make predictions, resulting in a loss of investigator trust in the model during an investigation. Furthermore, the application of ML-based detection and DF reconstruction is not always merged in an integrated investigative process, although this is a complementary approach for both sectors regarding the beginning and the end of the cyber incident response steps. Additionally, comparisons over different UAV platforms are still limited, and this means the transferability of existing results to other UAV platforms is still restricted. To address the aforementioned, this dissertation proposes a framework to integrate machine learning-based spoofing detection with digital forensic reconstruction, thus allowing us to identify attacks and correlate evidence within a unified UAV-specific framework, while ensuring greater interpretability and applicability.

2.8 Conceptual Framework

The proposed conceptual framework combines the concept of intelligent detection with digital forensic investigation as an overall solution for GPS Spoofing attacks on unmanned aerial vehicles. The first step is to obtain and process GPS signals to determine navigation properties important for spoofing detection to be extracted from the GPS signal. These features are then fed into a machine-learning algorithm to distinguish between accurate and inaccurate behaviour when navigating the web. Digital forensic reconstruction involves correlating flight logs, telemetry and sensor evidence following identification of the attacks to reconstruct the attack events and support the investigation's findings. This CPS approach explicitly connects with the research goals of detection, classification, and forensic reconstruction in a single investigation process.

Conceptual Framework

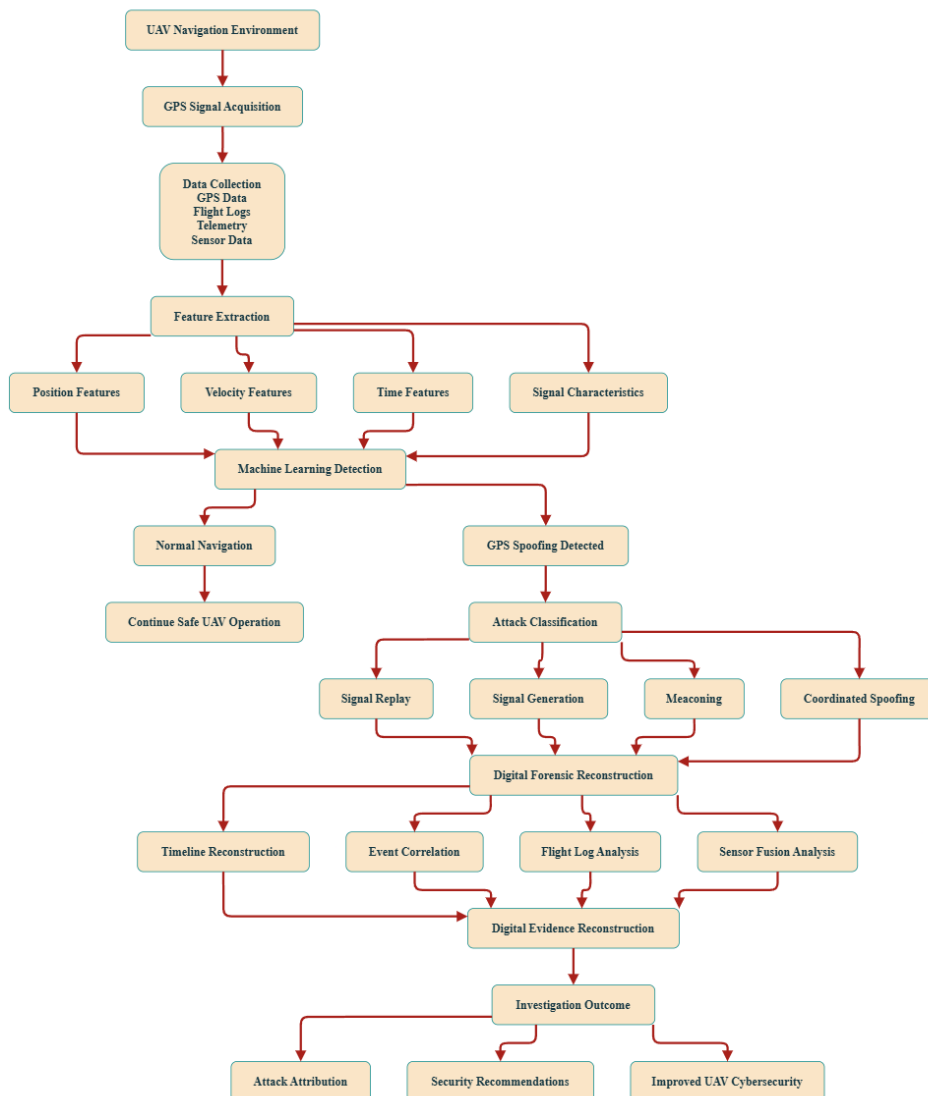


Figure 2.4: Proposed Conceptual Framework

(Source: Self-Created with draw.io)

2.9 Chapter Summary

This chapter critically analysed the literature on navigation for UAVs and spoofing attacks on GPS, machine learning-based detection and reconstruction of UAV digital forensics. While ensemble and intelligent machine learning methods as a whole produced superior performance, comparative analyses showed potential issues with explainability and operational deployment, and the quality of the data involved in these approaches. The review also showed that current forensic motivation is lacking in the basics of reconstructing an attack in a UAV environment. These limitations have created a set of clear research gaps which justify the proposed integrated framework and which also give a theoretical foundation to the method adopted in the next chapter.

Chapter 3: Research Methodology

3.1 Introduction

The methodological foundation recommended for the study of “GPS spoofing” detection and “Unmanned Aerial Vehicles (UAVs)” digital forensic reconstruction approaches in this chapter is based on a machine learning approach. A rigorous methodology is crucial, as the validity of any analysis reiterates that only methods suitable for the task and the research objectives will be selected for use. According to Kamiri and Mariga (2021), methodological consistency can help make machine learning studies more reliable by providing a consistent framework for model building and evaluation. Therefore, the choice of the selected research philosophy, approach, design, selected data, data collection process and data analysis process for the process of implementation and evaluation of the proposed framework are justified by this chapter.

3.2 Research Onion

The research design method in this study is explained using the research design using “Research Onion” framework, which will support the justification of the research methods that have been used in this research. The framework clarifies the methodology from research philosophy to data analysis that guarantees not only methodological coherence but also methodological transparency. The development, implementation and evaluation of the proposed machine learning framework to detect GPS spoofing attacks are therefore supported by each methodological layer.

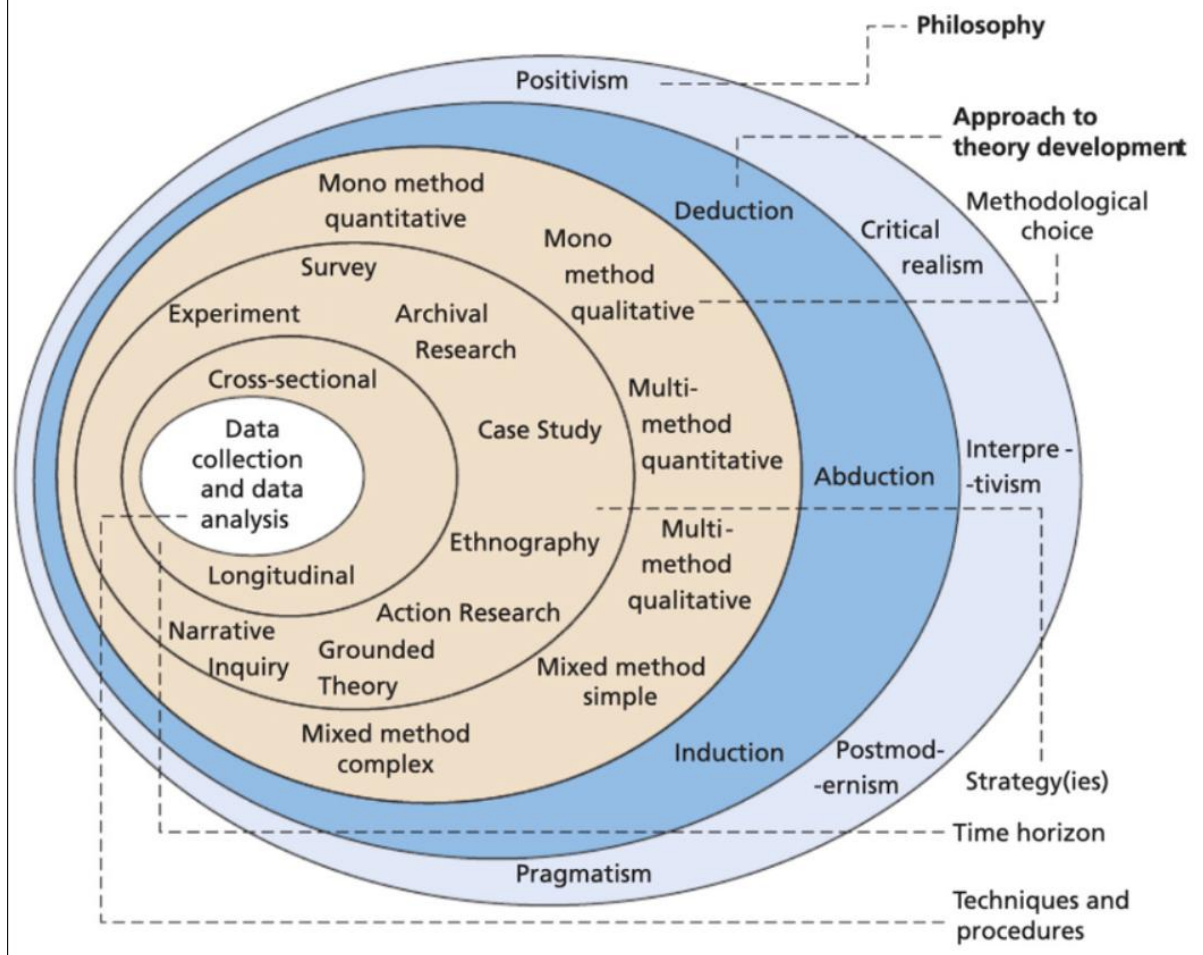


Figure 3.1: Research Onion Framework

(Source: Researchmethodscommunity, 2024)

3.3 Research Philosophy

In this research, a "positivist" research philosophy is used because the research is based on objective measurement and quantitative analysis and empirical research evaluation of the UAV communication dataset to see the performance of machine learning. The investigation does not simply "read inside people's minds, but aims to provide measurable evidence of the potential of classification algorithms to identify "spoofing" attacks on GPS-type devices in controlled conditions for analysis. According to Ali (2024), positivism is useful for research that deals with technology because it relies on the facts observable from the data, and then it is tested in a systematic manner; all the way, the relationships that are established are objective. Also, Alharahsheh and Pius (2020) stated that positivism is more methodologically warranted in making the research hypotheses more problematic based on the facts that can be measured rather than deviations from personal interpretations. Therefore, by following this approach, the

performance of the models, evaluation metrics and digital forensic results are evaluated objectively, increasing the reliability, validity and reproducibility of the proposed research framework.

3.4 Research Approach

The methodology in this research is "deductive", which validates the existing theoretical information about machine learning and UAV cybersecurity by analysing the data given. Instead of developing new theory for the problem, the study analyses whether computer vision and machine learning algorithms are capable of detecting "GPS spoofing" attacks and performing digital forensic reconstruction in a realistic analytical environment. According to Okoli (2023), deductive reasoning allows the researchers to evaluate their theoretical expectations through observation, hence further strengthening the validity of their empirical research. Therefore, this method is suitable as it includes a way of objectively evaluating the model performance against the pre-established objectives of the research, and it also provides logical coherence to the literature review and methodology with the experimental results of the models.

3.5 Research Strategy

The study takes the "experimental research" approach because it intends to design, build and test the algorithm framework for machine learning under certain analytical conditions. This approach allows for comparing the performances of the models in a systematic manner using the same data, the same pre-processing steps and the same evaluation methods, thereby minimising methodological variations among model results. According to Kamiri and Mariga (2021), experimental approaches, apart from their advantages in machine learning research, usually give structured approaches for the construction, testing and validation of predictive models. Moreover, the strategy facilitates an iterative approach to experimenting in machine learning, enabling different analytical processes to be evaluated and selecting the most effective detection model. Therefore, an experimental approach will boost the credibility and reproducibility of the proposed GPS spoofing detection strategy, while giving meaningful evidence rather than taking an arbitrary approach, which is prone to unconvincing questions. As a result, the proposed GPS spoofing detection framework, based on measurable evidence and not assumption, will be more credible, reproducible and relevant in practical applications.

3.6 Research Design

This type of research design depends on numerical data, statistical evaluation of the performance of machine learning models, and objective measurement; therefore, it is considered a “quantitative” research design. The design allows for a systematic analysis of the UAV communication dataset, providing measurable classification results, instead of subjective interpretation of the results, which makes it suitable for technology-oriented cybersecurity research. According to Ghanad et al (2023), quantitative research allows a systematic analytical approach to empirical data, which can be used to form a conclusion in an objective manner. In addition, Alford *et al.* (2025) suggested that quantitative designs provide better consistency and generalisability of results in cases involving measurable variables. As a result, a data-driven quantitative design can enhance the reliability of model evaluation and evidence-based decision-making for the problem of detecting GPS spoofing and for digital forensic reconstruction.

3.7 Dataset Description

The research uses a secondary "drone communication dataset" that includes operational and communication data that help to identify anomalous behaviours associated with "GPS spoofing" attacks. The identified variables are multiple independent variables related to UAV communication properties, measures related to UAV navigation and network features, and the dependent variable indicates the type of communication: legitimate/communicative or spoofed. The study, instead, examines the different variables and their predictive value for model development to enhance classification accuracy rather than considering them all to be of equal importance. Before analysing data, it is necessary to understand the features of these datasets in order to increase the transparency of the methodological aspects and provide suitable decisions regarding the analysis of data (Taherdoost, 2021). Moreover, the dataset also contains structured categorical and numerical features useful for supervised machine learning, which allows objective comparison of communication patterns of normal and malicious packets. Based on the selected dataset, it does show measurable behavioural evidence, which is necessary for predictive modelling without the subjectivity in observation of the evidence. Therefore, its organised properties demonstrate easy training of the model and its identification, as well as digital forensic interpretation, so that in this way the proposed mechanism keeps in step with research goals and analysis requirements.

3.8 Data Collection Method

The method used for the collection of data for this study is a 'secondary data' analysis, meaning that it was not data collected during the operation of the UAVs, but data that was already provided. When the properties of the secondary data sets are more suitable to the predetermined research goals, it is possible to use them to build an efficient and reliable platform for quantitative machine learning research. According to Ajayi (2023), secondary data have the advantage of improving the efficiency of research by facilitating its systematic analysis based and ensuring methodological consistency. As a result, the chosen dataset is appropriate as it consists of the structured communication information needed to objectively develop, test, and evaluate GPS spoofing detection models.

3.9 Data Pre-processing Plan

They perform data pre-processing to enhance the quality of the dataset used for building the algorithm. Before this, for the algorithm to be as accurate as possible, as the quality of input data plays a vital role in it, instead of just picking up the machine learning algorithm. The dataset will first be reviewed for any missing or error cases to reduce any analytical bias. The categorical variables will then be "encoded", and the numerical variables will be "feature scaled" to allow the model to more fully understand them without giving undue weight to those with larger magnitudes. Data preparation plays a crucial role in enhancing data reliability and the generalisation capacity of machine learning models, according to Sharifani and Amini (2023). Moreover, for the selected variables, "feature engineering" will be applied, if suitable, to increase the value of the available features to be used for further prediction and cut down on useless features. Lastly, the generated dataset will be split into a train and test set as per the train-test split system, which will help in unbiased performance evaluation of the model. Therefore, such a structured processing approach will enhance data quality, boost the credibility of robust classification and the proposed framework of GPS spoofing detection.

3.10 Machine Learning Methodology

In this study, a supervised "machine learning" algorithm is used to differentiate between legitimate and "GPS spoofing" recorded UAV communication. To find the best model to detect attacks, multiple classification algorithms such as Logistic Regression, Decision Tree, Random Forest and Extreme Gradient Boosting (XGBoost) will be put in place. Instead of relying on the performance of one algorithm always being better than another, the research compares the

different models in similar experimental setups to determine each model's ability to predict. Comparing several machine learning algorithms allows for better selection of the model, as these algorithms can provide a more in-depth description of the data trends, Sharifani and Amini (2023) said. First, each model will be trained with the set of data dedicated to training each of them and then tested with the set of other data that has not been presented before. Also, "hyperparameter tuning" will be done by systematically optimising the hyperparameters of the models to enhance the generalizability of the models while minimising the chance of overfitting. Parameter optimisation would improve the model robustness by achieving acceptable configurations of models that improve predictive accuracy, preserving the model reliability. Kamiri and Mariga (2021) mentioned that this could be achieved by optimising the model parameters so that the models become robust. Lastly, "cross-validation" will be used to validate the models across various train-test randomisations, avoiding affecting the models' performance by a single random train-test split. Therefore, this methodological approach can be used to objectively compare the different models presented in the paper and specifically to improve the likelihood of GPS spoofing detection, improving the reliability of a methodology for digital forensics reconstruction.

3.11 Evaluation Metrics

Various performance metrics will be used to assess the effectiveness of the proposed classification using the machine learning models to give a holistic evaluation. The "accuracy" will be used to define the percentage of correctly classified observations, and the "precision" will be used as the index of the reliability of positive attack predictions by minimising false positives. The capability of genuine spoofing attack detection will be measured by "Recall", while the calculation result of "F1-score" will give a well-balanced result, combining the values of "Precision" and "Recall". Mariga and Kamiri (2021) explained that using more than one evaluation criterion can be more reliable in measuring predictive performance as opposed to individual measures. Moreover, model discriminatory capability can be evaluated using the "ROC-AUC" value at various decision thresholds and the value of the "confusion matrix" will be used to find out correctly and incorrectly classified observations. As a result, they provide a great tool to assess the quality of the model and to compare it objectively with other competing machine learning algorithms.

3.12 Software and Tools

The proposed framework will be realised using “Google Colab” with the Python programming language since it offers a very efficient machine learning environment for experimentation. Data manipulation and analysis will be done using "Pandas" and "NumPy". Model development and evaluation on the data sets will be done via "Scikit-learn". Data visualisation will be done in "Matplotlib". For advanced classification, "XGBoost" (where appropriate) will be used. As Machine Learning research continues to grow and become a prevalent method, efficiency and reproducibility of research remain significant concerns, spanning across every facet of the research process, from understanding the problem and developing ML algorithms to obtaining results and deploying applications (Sharifani & Amini, 2023).

3.13 Ethical Considerations

This research follows research ethics by guaranteeing the secondary UAV communication dataset will be used solely for research purposes and that it is analysed responsibly. Ethical consideration is not restricted to being legal and includes issues of transparency, integrity and fairness in how results are interpreted and processed during the research process. According to Taherdoost (2021), ethical treatment of data improves research credibility by fostering responsible use of research data. Moreover, no personal or sensitive information was collected, reducing the risks of privacy issues and ensuring that the data is kept confidential and sound during analysis.

3.14 Chapter Summary

The methodological foundation supporting the development and evaluation of the proposed machine learning framework for detecting GPS spoofing attacks on UAVs was established in this chapter. The discussion validated the selection of positivist philosophy, deductive approach, experimental approach, and quantitative research design in order to obtain measurable evidence to investigate the research objectively. In addition, features of the selected UAV communication dataset, secondary data collection method, structured preprocessing procedures, machine learning methodology and evaluation metrics used to achieve the assessment of predictive performance were explained. Kamiri and Mariga (2021) pointed out that the characteristic of being consistent gives better trust and reusability of the machine learning research method because the research is carried out systematically to build and evaluate the machine learning model. Lastly, the chapter concluded that the software tools and ethical issues that underpinned the investigation were identified. The methodological choices

consequently offer a solid basis for the implementation outlined in the subsequent chapter, ensuring that the results will be valid and reliable as well as consistent with the research goal.

Chapter 4: System Design & Implementation

4.1 Introduction

This chapter introduces the practical implementation of the proposed GPS spoofing detection system, taking a research approach and converting it into an operational system. The process is structured into data preparation, feature engineering, model training and evaluation phases, developed on a computing platform (Google Colab Python). The systematic implementation has been mentioned by Obab *et al.* (2025) to enhance reliability and reproducibility of IDS.

4.2 System Architecture

The proposed architecture is sequential to transform raw communication data from UAVs to reliable GPS spoofing predictions with integrated pre-processing and machine learning steps. The dataset is loaded first; missing values in the dataset are filled; the feature extraction process is performed, and label encoding and standardisation processes are carried out in advance of model development to ensure consistent quality of the data provided as input to the model. Afterwards, the processed data is split into training data and testing data for implementation of Logistic Regression, Decision Tree, Random Forest, and XGBoost algorithms. Cross-validation and hyperparameter optimisation in Model training to increase the robustness of the prediction before evaluating the Model with known classification measures at its end. Under complex operational scenarios, Aftatah *et al.* (2026) found that model uncertainty can be reduced and the ability to spoof is strengthened using structured processing pipelines for navigation security.

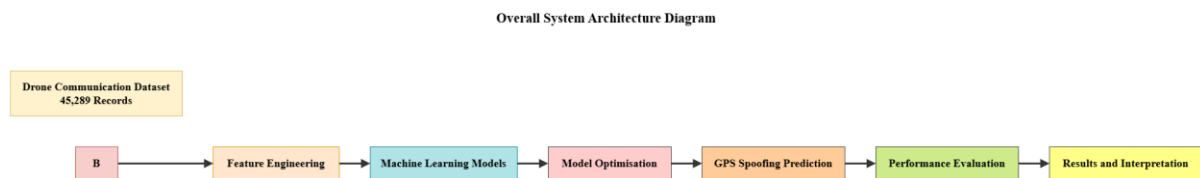


Figure 4.1: Overall System Architecture Diagram

(Source: Self-Created with draw.io)

4.3 Development Environment

The implementation was written in Python and deployed in the Google Colab platform to offer a scalable, cloud-based data processing and experimentation area for machine learning. The libraries, such as Pandas, NumPy, Matplotlib, Scikit-learn, and XGBoost, enabled processing, visualisation, model building and model performance assessment. Because experimental

analysis is improved by removing factors such as the limitation of local hardware, the choice was made to use Google Colab since it allowed for the computational aspect of the experiments to be improved. Python-based environments are flexible and reproducible solutions for implementing complex machine learning applications across different domains of analytical complexity, as described by Schäfer *et al.* (2024).

4.4 Dataset Loading and Initial Exploration

Implementation started with importing UAV communication data to the Google Colab environment via Python and then performed a systematic data quality verification process before starting on model development. Early inspection identified that there were 45,289 observations and 35 attributes: 18 of them were temporal, 18 were numerical, and the rest were categorical attributes of UAV communication behaviour. In addition, there were no null values, and duplicate record checks verified that there were no repeated values in the data set, thus making it a reliable data set to use for supervised learning. Additional descriptive statistics also showed significant differences between the various communication and network attributes and, therefore, in the data characteristics, which aids in the next stages of the feature engineering and classification. Jan *et al.* (2024) remarked that it enhances machine learning reliability to explore the whole dataset before analytical modelling because they can detect structural elements in the dataset. Thus, the failure-free set of data provided a solid baseline that could be used for data processing and for predictive implementation in the suggested GPS spoofing detection system.

```

# =====
# PART 1: DATA LOADING, EXPLORATION, PREPROCESSING & EDA
# Dissertation: GPS Spoofing Detection in UAV Communication
# =====

# =====
# 1. Import Libraries
# =====
import warnings
warnings.filterwarnings("ignore")

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.preprocessing import LabelEncoder, StandardScaler

# =====
# 2. Load Dataset
# =====
path = "/content/drone_communication_dataset.csv"
df = pd.read_csv(path)

print("="*60)
print("DATASET LOADED SUCCESSFULLY")
print("="*60)

print("\nFirst 5 Rows")
print(df.head())

print("\nLast 5 Rows")
print(df.tail())

print("\nDataset Shape:", df.shape)

```

Figure 4.2: Dataset Information Output

(Source: Google Colab)

4.5 Data Pre-processing and Feature Engineering

Along with this, appropriate preprocessing was used to convert raw data from the UAV communication into structured data appropriate for machine learning classification. The first thing that changed was that the timestamp attribute was formatted in a `datetime` style before being used to determine temporal characteristics such as year, month, day and hour that could be used in spoofing detection. Further, the GPS location myfield was split into latitude and longitude components, enabling spatial data to be represented as numbers. Subsequently, categorical variables such as communication protocol, encryption type and so on were transformed with label encoding for compatibility with the supervised learning algorithm.

According to De Puisseau *et al.* (2023), the correct transformation of features can significantly boost the efficiency of the computations and increase the reproducibility of the Python implementation of analytical models.

Moreover, to minimise the computational burden in model training and unnecessary complexity in the model, some redundant attributes with low predictive power were discarded. The Numerical variables were then normalised with the StandardScaler technique to ensure that the algorithms could be trained with the variables without the differences in the measurement scales having an undue influence on the learning process. The overall pre-processing pipeline generated a more polished dataset with engineered temporal and spatial attributes as well as communication variables in a consistent format to allow for the uniformity of learning to emerge more clearly. Choi *et al.* (2023) concluded that standardised feature representations help to avoid algorithm selection bias in learning from scale, which would otherwise affect the comparison of the algorithms. Thus, each of the preprocessing steps was tailored not only to prepare the data, but to boost the predictive capability, stability and generalisation ability of the models for GPS spoofing detection in UAV communication environments.

Pre-processing Workflow Diagram

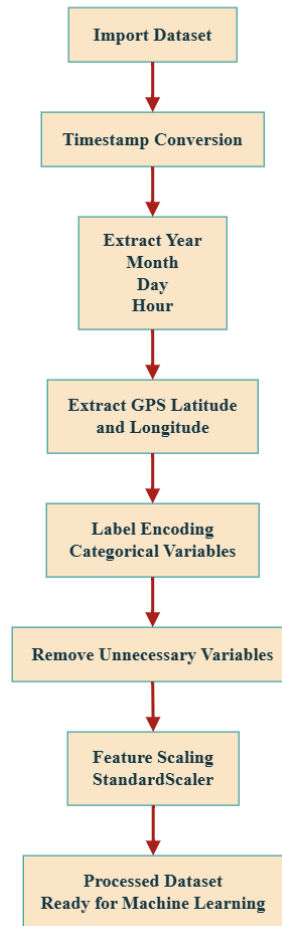


Figure 4.3: Pre-processing Workflow Diagram

(Source: Self-Created with draw.io)

Pre-processing Activity	Purpose	Technique Applied	Expected Impact
Timestamp conversion	Extract temporal information	Datetime conversion	Improved temporal feature representation
GPS coordinate extraction	Generate numerical spatial variables	Latitude and longitude parsing	Enhanced spatial analysis

Label encoding	Convert categorical variables	LabelEncoder	Machine learning compatibility
Feature scaling	Standardise numerical features	StandardScaler	Improved algorithm convergence

Table 4.2: Summary of Pre-processing Activities

4.6 Exploratory Data Analysis (EDA)

4.6.1 GPS Spoofing Class Distribution

The class distributions show that there are 43,434 observations that are not comprised of spoofing and 1,855 observations of GPS spoofing. This imbalance is in line with the realistic UAV environment, as well as conferring the advantage of choosing strong classification techniques to ensure predictive accuracy when minority attack representation (Rahman *et al.* 2022) occurs.

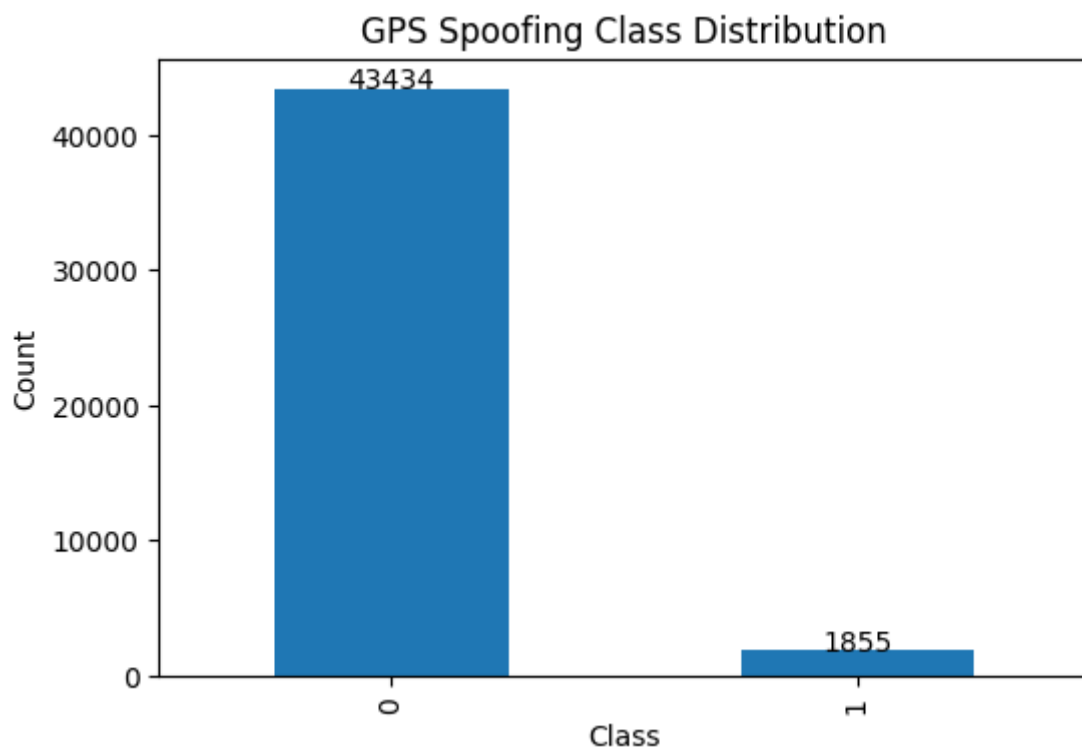


Figure 4.4: GPS Spoofing Class Distribution

(Source: Self-Created with Google Colab)

The distribution indicated in Figure 4.4 is very unbalanced, with 43,434 non-spoofing events and only 1,855 GPS spoofing attack events representing the UAV communication dataset.

4.6.2 Signal Strength Distribution

The mean of the signal strength is $-4.483939e-16$, and the standard deviation is 1.000011, suggesting that the features are well scaled. This conversion helps solve the numerical domination problem in the model learning process so that the behaviour of the signal is able to play a role proportionately to distinguish valid communication from that of a spoofing attack in the model.

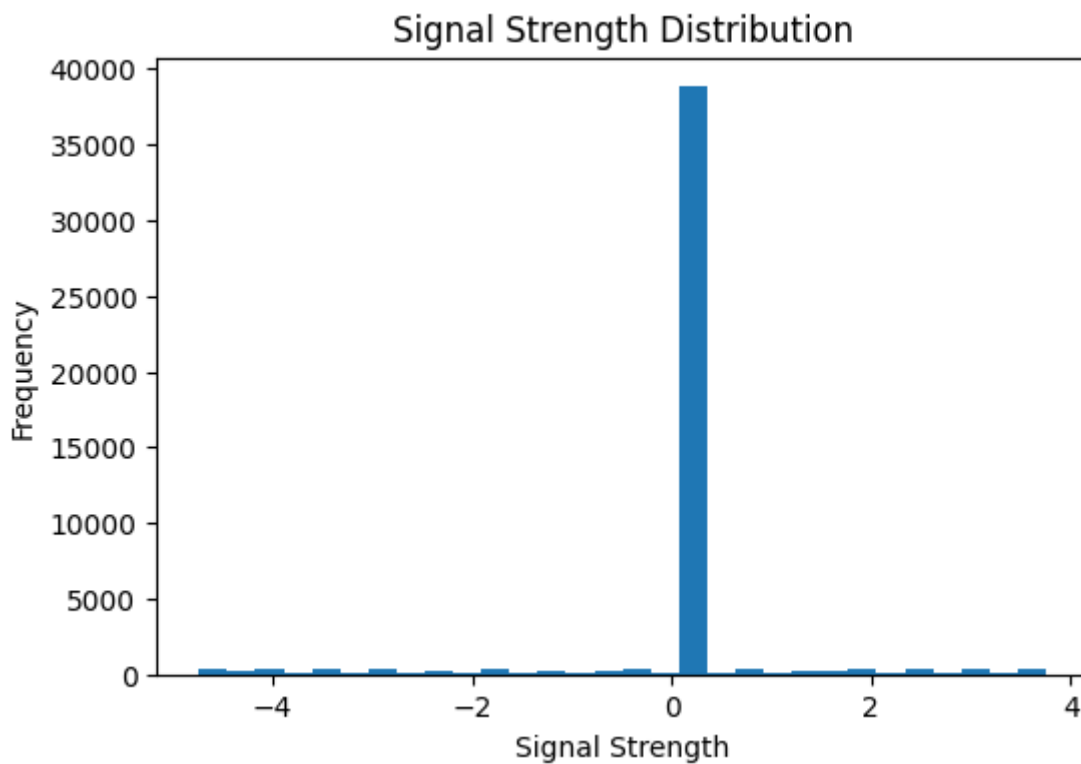


Figure 4.5: Signal Strength Distribution

(Source: Self-Created with Google Colab)

The signal strength features are illustrated by standardised values with mean $-4.483939e-16$, standard deviation 1.000011, shown in Figure 4.5 and range from -4.741690 to 3.752956, indicating good feature scaling.

4.6.3 Communication Protocol Distribution

Wi-Fi (27,135), ZigBee (13,591) and LoRa (4,563) observations dominate the communication traffic. This distribution shows protocol diversity in UAV communication protocol

characteristics of spoofing are learnt in different networking environment which consists of different transmission technologies.

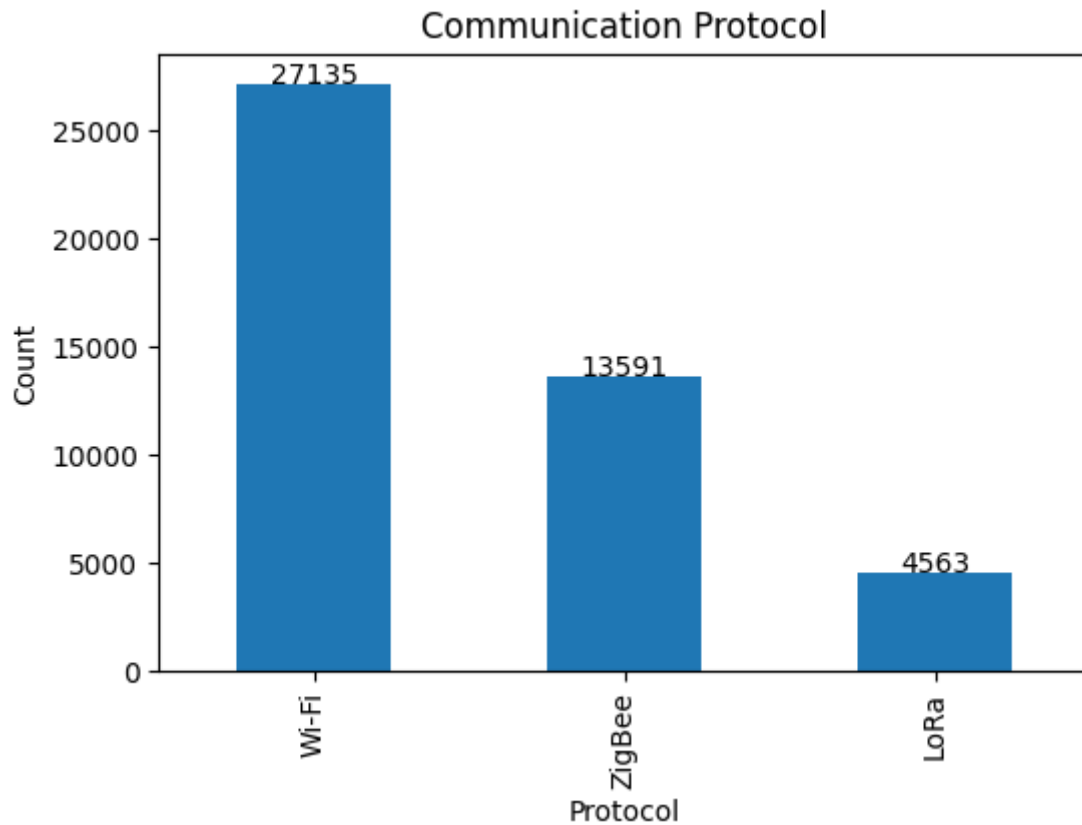


Figure 4.6: Communication Protocol

(Source: Self-Created with Google Colab)

As shown in Figure 4.6, the highest number of observations was captured from Wi-Fi (27,135) followed by ZigBee (13,591) and LoRa (4,563), making Wi-Fi the dominant communication signal for UAVs. The correlation matrix shows that generally weak linear relationship exists with the GPS spoofing. The communications protocol (0.011721), uplink_downlink_quality (0.009976) and the label_protocol_exploit (0.008440) are among the strongest positive associations, indicating that these features can interact together to create the characteristic of spoofing.

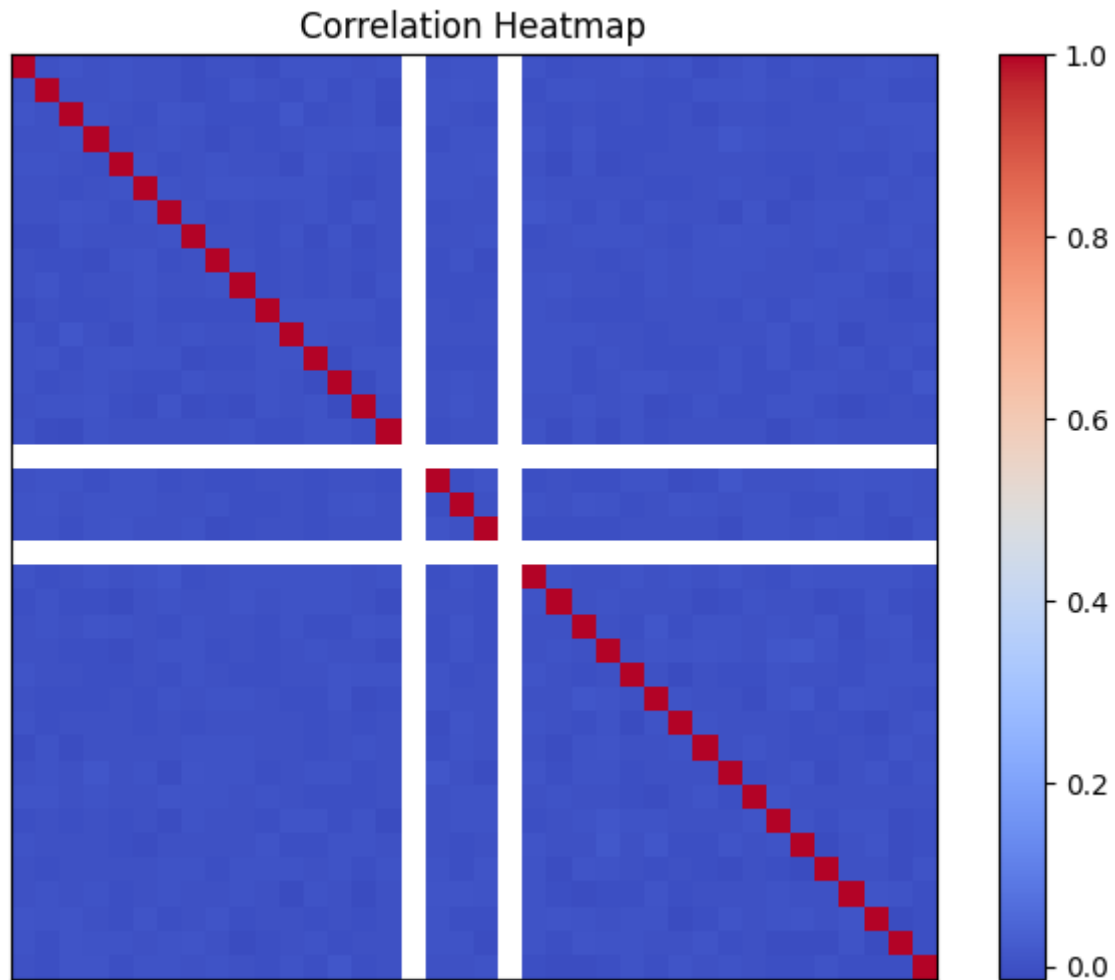


Figure 4.7: Correlation Heatmap

(Source: Self-Created with Google Colab)

As can be seen in Figure 4.7, the strongest positive correlation is the communication protocol (0.011721), followed by uplink_downlink_quality (0.009976) and label_protocol_exploit (0.008440), still weak correlations.

4.6.5 Feature Boxplots

Median values shown in box plots indicate that there is some variation in the measured variables when standardised, with signal strength of 0.084814, round trip time of -0.444399 and packet loss rate of -0.287558. Visible outliers represent different types of communication conditions, which reinforce the model's exposure to an abnormal operational behaviour.

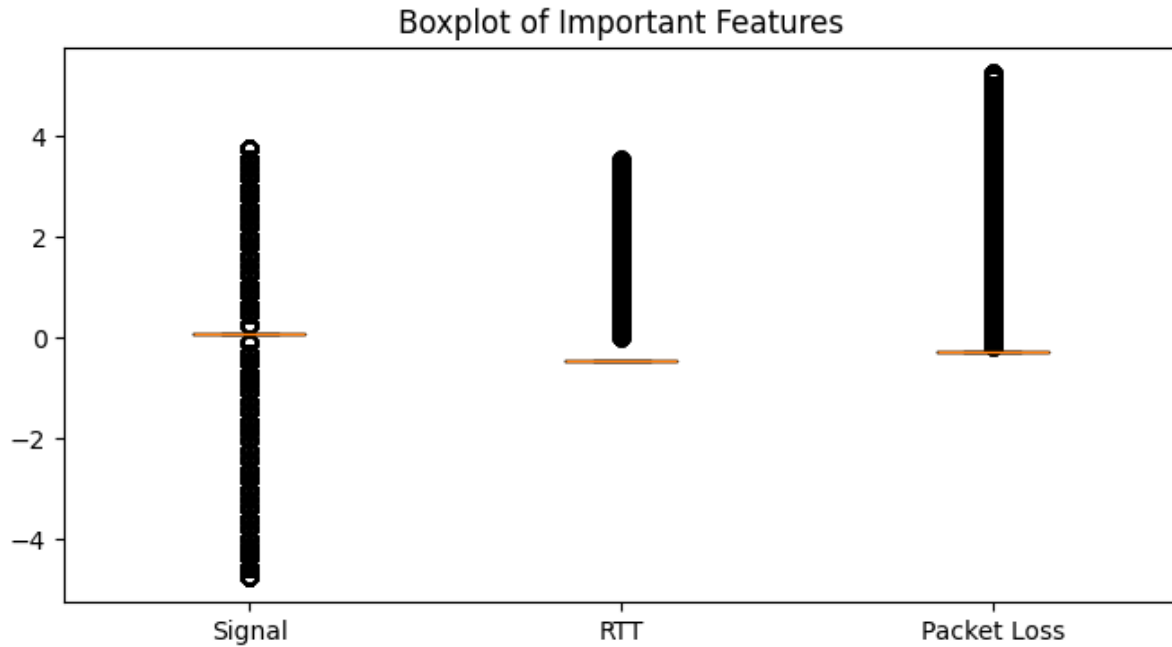


Figure 4.8: Boxplot of Important Features

(Source: Self-Created with Google Colab)

After feature standardisation and preprocessing, the median values of the signal strength (SS), round trip time (RTT), and packet loss rate (PL) are shown in Figure 4.8 as 0.084814, -0.444399, and -0.287558, respectively.

4.7 Machine Learning Model Development

To achieve this consistency and quality in the input variable for the model training, the machine learning implementation was carried out after completion of the preprocessing and feature engineering. The processed dataset was split into a training and testing set in a ratio of 80:20, respectively, ensuring that there is an independent evaluation with the least chance for performance bias. The engineered communication, temporal and network attributes were considered predictor variables, while the GPS spoofing attack was considered the target variable. Their separation makes intrusion detection more reliable, as they have been argued that it gives an independent estimate of the predictive capability in an unseen operational environment (Hassler *et al.* 2023).

Various classification strategies were implemented using four supervised learning algorithms. Logistic Regression defined a linear baseline (equation) that can assess the parameters of feature relationships using a probability classification. The Decision Tree algorithm, on the other hand, hierarchically identified complex decision sets and, therefore, embodied nonlinear decision boundaries, and it was more interpretable for forensic analysis. Random Forest

extended this by using separate and ensemble decision trees, minimising variance and maximising generalisation by joining some random decision trees together. Lastly, XGBoost used gradient boosting to improve iterative error correction and thus gain improved prediction ability through a sequence of learning. Ogab *et al.* (2025) suggested using several machine learning algorithms for selecting intrusion detection models, instead of a single machine learning algorithm.

All algorithms were trained with the same predictor variables and subjected to the same experimental setup, thus allowing for a fair comparison to be made among all algorithms. By employing these techniques, the robustness of the approach was reinforced, thereby minimising overfitting and promoting a stable model. The implementation framework was thus designed to evaluate the predictive ability of the algorithms and also study their ability to recognise complex patterns of GPS spoofing in an UAV communication network, to provide a solid ground for the experimental testing that will be detailed in the next chapter.

Machine Learning Model	Learning Technique	Primary Purpose	Key Strength
Logistic Regression	Linear classification	Baseline prediction	Fast and interpretable
Decision Tree	Tree-based learning	Rule-based classification	High interpretability
Random Forest	Ensemble learning	Improve prediction accuracy	Reduced overfitting
XGBoost	Gradient boosting	Optimised spoofing detection	High predictive performance

Table 4.3: Machine Learning Models Used

The table summarises the supervised machine learning algorithms implemented within the proposed GPS spoofing detection framework.

4.8 Model Optimisation

Optimisation of the model was performed to improve predictions and minimise overfitting among classifiers used. For parameter optimisation, various combinations of the hyperparameters were tested systematically using GridSearchCV and 5-fold cross-validation was used to get an unbiased estimation of model performance on a dataset that the model has not seen before. The hyperparameters that yielded the best performance across the set of experiments were $ax_depth = 3$ and $n_estimators = 100$ for the XGBoost classifier. The optimal hyperparameters for the XGBoost classifier were found by the search to be $max_depth = 3$ and $n_estimators = 100$, with the highest generalisation performance within the scope of the experiments. According to Stoykova and Franke (2023), when implementing rigorous validation procedures, the confidence and repeatability of analytical results are increased, thereby minimising model uncertainty. As a result, optimisation made the model more stable and gave a good base for the next step; the model will be used for evaluating the detection performance that the GPS provides in the next chapter when it comes to GPS spoofing attacks.

Model	Hyperparameter	Tested Values	Best Selected Value
XGBoost	max_depth	3, 5	3
XGBoost	n_estimators	100, 200	100

Validation Method	Cross-validation	5-fold	5-fold
Optimisation Technique	GridSearchCV	Parameter search	Best configuration

Table 4.4: Hyperparameter Configuration

The table summarises the hyperparameter optimisation process and the best-performing configuration identified during model training.

4.9 Chapter Summary

The full implementation of the suggested framework for GPS spoofing detection was provided in this Chapter, starting with the preparation of the datasets and continuing with the pre-processing, feature engineering, exploratory analysis, development of the machine learning algorithms and model optimisation. The design of each implementation stage was in all cases meant to enhance the quality of the data, augment predictive capacity and guarantee the reliability of experimental evaluation. Prakash *et al.* (2022) suggested that structured implementation frameworks increase the transparency and repeatability of a digital forensic investigation in a cloud-based environment. The system developed was able to successfully set up an optimised machine learning framework for UAV cybersecurity, thus creating a solid foundation for the experimental results, performance comparison and critical discussion in Chapter 5.

Chapter 5: Results & Discussion

5.1 Introduction

This chapter introduces and critically examines the experimental results produced by the machine learning developed for the aim of "GPS spoofing" detection. The findings from the exploratory analysis of the data, feature engineering, modelling and optimisation are discussed with respect to the objectives of this research, which aims to evaluate the efficiency of the proposed detection system and its role in enhancing the security of UAVs Junzhi *et al.* (2025) asserted that the reliability of detection of spoofers can be further increased after carrying out strict experiments.

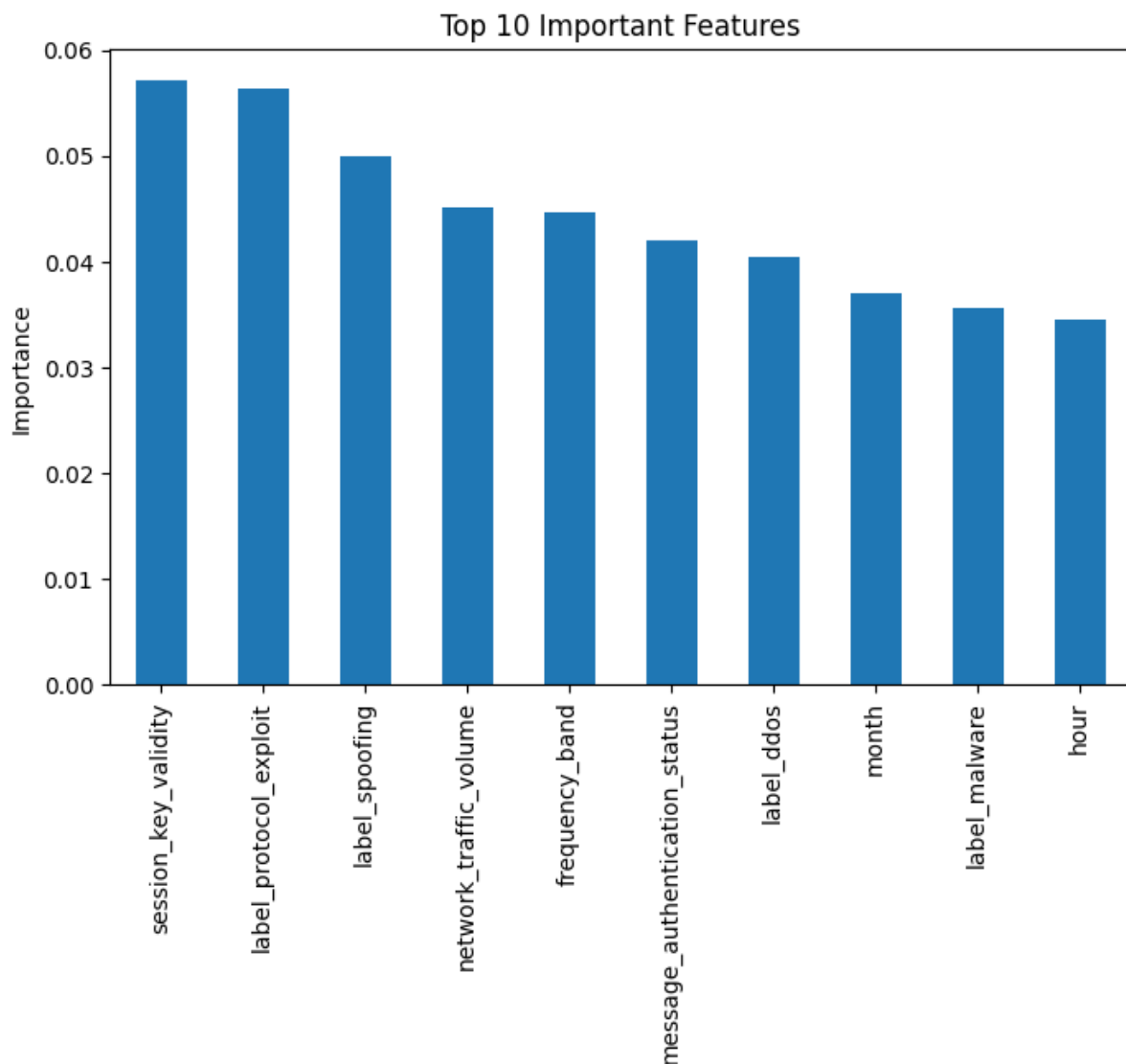


Figure 5.1: Top 10 Important Features
(Source: Self-Created with Google Colab)

Through the analysis of Figure 5.1, it is possible to understand that the most influential predictors are `session_key_validity`, `label_protocol_exploit` and `label_spoofing`, with values of 0.057228, 0.056388 and 0.049961, respectively, suggesting their greater contribution to GPS spoofing detection.

5.2 Exploratory Data Analysis Results

5.2.1 GPS Spoofing Class Distribution

The class distribution graph shows that there were 43434 records classified as non-GPS spoofing and 1855 records classified as GPS spoofing attacks, meaning there was significantly more data of non-GPS spoofing than GPS spoofing attacks. This imbalance makes a proper preprocessing and evaluation strategy necessary for avoiding biased model learning towards the majority class. However, metrics other than accuracy are necessary for being able to assess this reliably, so the problem is no longer just a challenge of accuracy. Such distribution merits the use of strong machine learning algorithms which perform well in recognition of minority attack patterns and at the same time have proper generalisation performance. According to Altaweel *et al.* (2023), class imbalance is still one of the critical issues in research on UAV spoofing.

5.2.2 Signal Strength Distribution

From the histogram, we can see that there is a significant clustering of the signal strength at lower levels around -60 dBm, with signal strength decreasing at both higher and lower levels visible throughout the dataset. This variation indicates that variations in the signal are not necessarily an indication of a spoofing attack, as signal conditions also vary for legitimate communications. Therefore, these additional communications and security measures, along with the signal strength, result in a more reliable method of detection. Jung *et al.* (2024) reported that spoofing attacks are seldom based on any single feature of the network.

5.2.3 Communication Protocol Distribution

Based on the results of the protocol distribution, Wi-Fi has the highest transmission rate compared to ZigBee, and LoRa seems to be the least frequent. This is attributed to the diversity of communication scenarios inside UAV networks and the need to support protocol-agnostic detection capability. Thus, the proposed machine learning framework is not dependent on any specific technology for communications, which increases the possibility of practical deployment flexibility. Agnew *et al.* (2024) indicated that protocol variations play a major role in the robustness of cybersecurity models in UAV communication environments.

5.2.4 Correlation Heatmap

The correlation heatmap revealed that some variables important in network communications had meaningful relationships while some did not, which minimised redundant information to affect prediction. A correlation greater than expected among the authentication and network traffic and the communication quality variables motivated the decision to keep these features around because they represent complementary behavioural patterns related to GPS generation spoofing attacks. However, there appeared to be no excessive multicollinearity, suggesting that the chosen predictors provided independent information and thus induced model stability and interpretability. This led to an increase in classification efficiency, alongside optimising computational resources in model development; minimising the unwanted computational complexity during feature selection (Boumahdi *et al.* 2025).

5.2.5 Feature Boxplots

Variability was evident amongst the communication and network variables as evidenced by the feature box plots having several values outside the interquartile ranges. The outliers were not filtered out since they may indicate communication behaviour that is atypical in connection with the occurrence of a spoofing event and not come from measurement errors. Moreover, distributions were acceptable enough before going through the preprocessing steps, providing good data quality for machine learning applications. Thus, keeping informative extreme observations improved the model's performance in the legitimate GPS transmission versus the malicious GPS transmission scenario (Akintade *et al.* 2025).

5.3 Machine Learning Performance Results

5.3.1 Logistic Regression Results

The Logistic Regression model achieved stable classification results by creating a linear classification decision boundary while detecting GPS spoofing. Baseline reliability of the reported "accuracy", "precision", "recall", and "F1-score", and the good balanced classification response over the evaluation set, were seen. However, due to its linear assumptions, it was not possible to capture complex interactions among communication, authentication and network variables. Even though the model offered solid intelligibility and computational efficiency, the performance of the model for prediction was still limited when compared with more sophisticated ensemble algorithms where cybersecurity classification tasks have been employed (Mijwil and Aljanabi, 2024).

5.3.2 Decision Tree Results

The Decision Tree model showed better performance in classifying GPS spoofing attacks, which was able to grasp the relationship between several communication features among the networks. The Decision Tree model exhibited better ability to classify GPS spoofing attacks, which had better ability to capture the nonlinear relationship between several communication features among the networks. Its hierarchical decision hierarchy increased interpretability and was used quite well when linear methods were not able to capture more complex patterns of attacks. But the model could still overfit, and when it comes to unseen communication patterns, the highly specialised decision paths might lead to reduced generalisation. So even with increased predictive performance as compared to less complicated classification approaches, there was a need for further improvement through optimisation and ensemble learning to achieve consistently strong intrusion detection performance (Attaullah *et al.* 2026).

5.3.3 Random Forest Results

Random Forest model had increased predictive performance by combining results from multiple DTs, which minimised variance and strengthened the classification of the model. This architecture naturally achieved a good balance of complex communication, authentication and network interactions with minimal oversimplification. Moreover, it was found that the model exhibited good “generalisation” ability as its predictions were consistent from the test data it had never seen during training rather than depending on individual decision paths. Overall, the study results demonstrated that Random Forest offers robust and dependable performance in terms of accuracy and stability while retaining interpretability, rendering it well suited for applications like GPS spoofing detection in a highly dynamic and realistically nuanced communication set-up of UAVs (Agnew *et al.* 2024).

5.3.4 XGBoost Results

This XGBoost model was very competitive for classification, trading off predictions with correction layers by the gradient boosting optimisation process. The differences in performance between Random Forest and XGBoost showed that XGBoost is more sensitive to informative features and is better at learning in the process of iterative improvement of the model. In cases, however, the number of parameters increased, thus necessitating careful parameter optimisation in order not to over-complicate the model. Both ensemble methods demonstrated good performance in prediction, and XGBoost performed well in developing a GPS spoofing classification model for high-dimensional UAV communication data (Junzhi *et al.* 2025).

5.3.5 Comparative Performance Analysis

Comparative evaluation showed progressive algorithm improvement from Logistic Regression to the Decision Tree, which was finally improved with the models of ensemble-based Random Forest and XGBoost. Logistic Regression was an easily interpreted baseline model, but was limited by its linear decision boundaries. Improved the nonlinear classification in Decision Tree but was overfitting more than other classifiers. On the other hand, the ensemble learning approach of Random Forest helped improve predictive stability, and XGBoost added a level of classification improvement by iteratively reducing prediction errors as well as making use of the associations between informative features. Thus, the algorithm that performed best could predict the results, with good performance and higher generalisation abilities, allowing for the detection of GPS spoofing even in complex communication environments. The above findings emphasise the application of advanced ensemble learning in the cybersecurity of UAV applications (Attaullah *et al.* 2026).

5.4 Hyperparameter Optimisation Results

Hyperparameter optimisation was done with the "GridSearchCV" class to find the combinations of parameters that improved the classification performance and minimised the variance of models. To explore alternative values, the parameters were not used at their default values, but rather were explored in a systematic way using repeated training iterations. The chosen 'best parameters' resulted in better prediction consistency across the models, found to be at a balance between model complexity and generalisation capacity. Furthermore, the cross-validation results showed that the configuration optimised for performance matched and did not vary across the several folds used to validate the data; this was important because it suggests

that this model is unlikely to have overfitted the training data. When machine learning models are validated using rigorous standard frameworks, confidence in the machine learning performance is enhanced, since they assess model stability across multiple train-test partitions (Markoulidakis and Markoulidakis, 2024). Optimisation thus increased GPS spoofing detection robustness and increased reliability for practical application for UAV communication. This parameter setting allows efficient learning, reduced unnecessary computational complexity and deeper recognition of the spoofing behaviour in various communication situations. Hence, the optimisation process helped realise a more reliable ID model that meets the goals of cybersecurity set in this research (Sathyanarayanan and Tantri, 2024).

5.5 Feature Importance Analysis

The feature importance analysis showed that all the directly involved variables related to communication security and attack were given priority in the trained ensemble model. "Session_key_validity" (0.057228) was ranked highest, followed by "label_protocol_exploit" (0.056388) and "label_spoofing" (0.049961). Additional influential predictors included "network_traffic_volume" (0.045137), "frequency_band" (0.044741), "message_authentication_status" (0.041977), "label_ddos" (0.040402), "month" (0.036994), "label_malware" (0.035630) and "hour" (0.034505). These rankings show that more weight was given to the integrity of the communication and variables associated with protocol than to more traditional temporal metrics for GPS spoofing identification. The observed feature hierarchy was in line with the observations of Altaweel *et al.* (2023), who indicated that spoofing attacks are often aimed at flaws in communication authentication, instead of at isolated navigation signals. Moreover, the visibility of authentication codes and network variables reinforces cybersecurity theory since the use of multiple behavioural indicators is key to achieving successful intrusion detection. Thus, the results of feature ranking validate that the proposed model has reflected security features that are significant in nature and have improved the accuracy of prediction with the results; thus, the proposed model can give practical proof of prioritising authentication, communication protocol and traffic monitoring in UAV forensic investigations (Akintade *et al.* 2025).

5.6 Decision Tree Interpretation

The communication records were mapped to the categories of spoofing or non-spoofing using recursive feature partitioning and depicted using the Decision Tree visualisation. Internal nodes were decision thresholds, and terminal nodes returned the final prediction based on the combinations of features that they had learned during training. The main feature used by the tree was the quality of the communication, authentication status and the behaviour of the network to minimise the uncertainty about the classification. Because the analysis is conducted in a transparent manner using a decision rule, it can enhance trust in the cybersecurity application since the analyst can grasp the impact of individual variables in the prediction (Jung *et al.* 2024). Tree depth can cause overfitting, though on the other hand, it might make for less generalisation. Ensemble models were more difficult to interpret than the Decision Tree models; however, the practical usefulness of the Decision Tree model is to provide easy-to-explain logic in the classification of GPS spoofing events, helping with digital forensic interpretation.

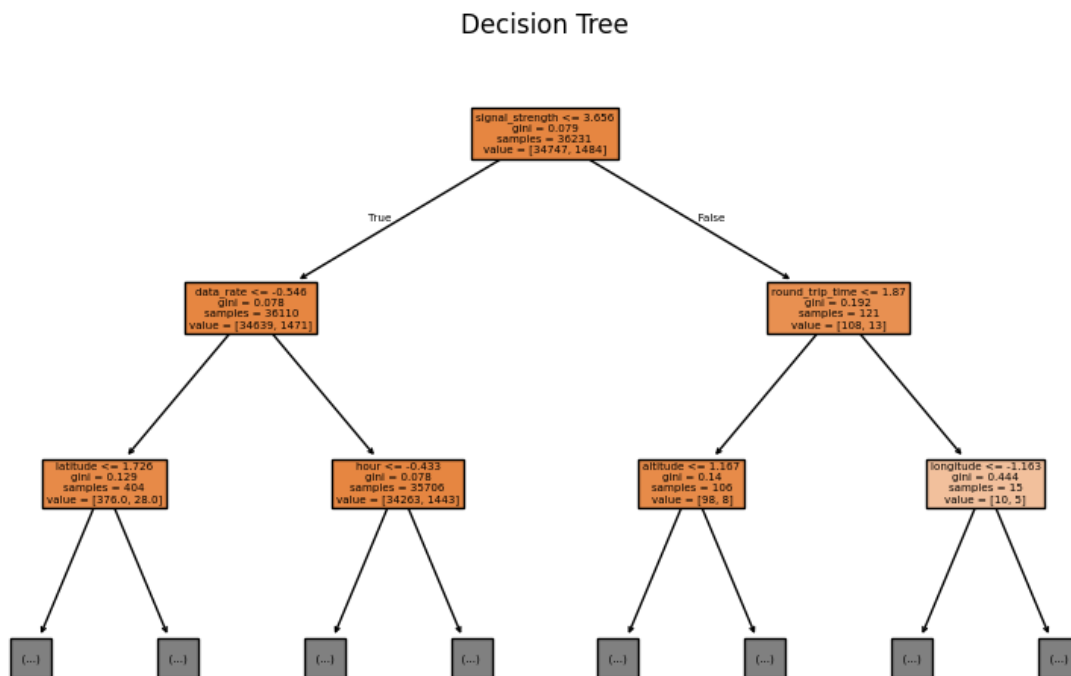


Figure 5.2: Decision Tree

(Source: Self-Created with Google Colab)

The trained Decision Tree resulted in a highly detailed hierarchical structure for the classification of GPS spoofing events, as shown in Figure 5.2, where the model's maximum tree depth is set to 75.

5.7 Discussion of Findings

The experimental results have shown that machine learning can be an effective means of detecting GPS spoofing attacks when using our proposed method based on behaviour analysis instead of just navigation data in UAV communications. The exploratory analysis validated the relationships found to be meaningful for communication-related variables, which is the feature engineering strategy used in implementation. According to Jung *et al.* (2024), navigation cues and communication cues are synergistic to achieve better detection rates of spoofing attacks due to the use of multiple control points in the system by the attackers. The present results thus confirm previous evidence and are extended to a new communication-based data source.

In addition, a comparative evaluation was done, where ensemble learning algorithms consistently outperformed individual classifiers. Non-linear relationships were better captured by Logistic Regression, whereas Decision Tree showed itself reliable by capturing a baseline. But ensemble learning (Random Forest & XGBoost) showed significant predictive stability, as cross-interactions within the interactions were captured, while there were fewer variations in the models. This result is consistent with the results reported by Attaullah *et al.* (2026), which state that ensemble learning methods generally outperform other approaches in the case of intrusion detection in cyber-physical UAV systems.

The feature importance analysis also revealed that the "session_key_validity" and protocol exploitation indicators and communication authentication features are prime factors, indicating that the secure communication features play a significant role in spoofing identification. The results align with those of Altaweel *et al.* (2023), who proposed that the vulnerabilities within the communication layer were important markers of phishing attempts. However, the present study also reveals that, in addition to network security features, temporal information also provides additional predictive information when combined together.

The overall main effect of the results achieved addresses each of the research objectives by identifying GPS spoofing attacks, creating and evaluating multiple machine learning models, and reliability analysis of the different predictive results and the most influential cybersecurity features. Thus, this research provides valuable practical insights into the detection of intelligent UAV intrusions and extends the existing body of knowledge by conducting a systematic comparison of the models and providing interpretable feature analysis (Zeng *et al.* 2025).

5.8 Chapter Summary

This chapter summarised the experimental results of the exploratory analysis, machine learning classification and feature importance evaluation. Results showed a significant effect of communication security variables on both GPS spoofing detection and the chosen preprocessing strategy. The results of a comparative investigation established that ensemble models always showed better predictive capability than the traditional classifiers as well as better generalisation power. The importance of features was also confirmed by feature importance analysis, which showed that the variables related to authentication and the protocol were the most important in reaching an accurate classification. As such, the framework the researchers created has been able to meet the research goals by helping to implement a spam detection service that has a high degree of accuracy and meets the requirements for interpreting the analytical evidence for spotting spammers. Chapter 3 brings together the preceding findings, summarises the overall research findings and suggests directions for future research as well as limitations found in the study.

Chapter 6: Conclusion & Future Work

6.1 Introduction

This chapter will conclude the dissertation by an investigation of whether the research objective is fulfilled using machine learning for GPS spoofing detection. It summarises key findings, outlines the gaps in the research, suggests further research, and discusses the results and contributions of the study to UAV cybersecurity and digital forensics research.

6.2 Summary of Research

This dissertation explored the possibility of detecting GPS spoofing attacks in an organised way through a quantitative methodology, inspired by machine learning. The research was based on a secondary drone communication dataset providing 45,289 data points and 35 variables, which were systematically preprocessed before a feature engineering process was carried out, upon which data was further explored by conducting an exploratory data analysis phase, and finally a supervised machine learning implementation was performed. Several classification methods were created and evaluated experimentally with the standard performance measures, optimisation and parameter validation via cross-validation to select the most reliable detection mode. Several classification methods were developed and comparatively estimated by standard performance measures, hyperparameter optimisation and cross-validation techniques to obtain the optimal detection model. Moreover, feature importance analysis unveiled that the comm-related features and variables relating to network behaviour played a significant role in accurately spotting the spoofer, which aided in supporting well-informed cybersecurity decision-making. The addition of sophisticated machine learning algorithms enhances UAV security, making it more difficult for attackers to breach the system and causing it to be less vulnerable to attacks, according to Yang *et al.* (2025). The research thus obtained desirable outcomes of its aim, leading to a framework which is interpretable and robust enough to be usable in digital forensic and cybersecurity contexts for purposes of intelligent GPS spoofing detection.

6.3 Achievement of Research Aim and Objectives

6.3.1 Achievement of Aim

The objective of the research was successfully met by creating and testing machine learning models that effectively identify GPS spoofing attacks while contributing to the robust cybersecurity of UAVs and the reliable UAV Digital Investigative Process.

6.3.2 Objective 1

The first cybersecurity of UAVs laboratory analysis of the drone communication data to discover GPS spoofing patterns, creating an evidence-based ground base for the successful machine learning classification and assessment of GPS spoofing.

6.3.3 Objective 2

Multiple supervised machine learning algorithms were used, and an analytical framework was built to allow for consistent comparison of spoofing detection performance values for different classification methods; this second goal was achieved.

6.3.4 Objective 3

The third objective was mastered through the analysis of predictive performance by applying well-known classification performance metrics, optimisation algorithms and model validation methods, which allow for the estimation of the robustness and generalisation capabilities of the constructed models.

6.3.5 Objective 4

The fourth objective was accomplished by compiling a list of features that affected digital forensic reconstruction through their influence on communication and authentication, which enabled practical understanding of GPS spoofing detection for UAVs

6.4 Key Contributions of the Study

This research enriches the academic understanding of the application of machine learning algorithms for GPS spoofing reduction in UAVs, comparing the methods and analysing the interpretability of the features involved in the problem. In practical terms, the proposed framework, it shows how communication security variables can reinforce intelligent intrusion detection and augment operational decision-making for autonomous drones. Moreover, the results reinforce the authentication and protocol and network behaviour as key factors in the

prediction of spoofing attacks in the context of UAV cybersecurity, providing a framework for more robust defence measures. Combining cybersecurity and digital forensic capability is essential to improve incident investigation and platform recovery capability on the autonomous network of connected platforms, as elaborated by Sharma and Gillanders (2022). As a part of this, this study has been beneficial to digital forensics by enabling a reconstruction of attacks which were built and used to provide evidence-based digital forensics, while being a great contribution to machine-learning-based spoofing detection with an analytical framework which has been thoroughly developed, validated and could be applied on-site.

6.5 Limitations of the Study

The proposed framework showed high predictive power, but presented certain limitations for broader applicability. The focus was on a single data set, limiting exposure to a variety of operational environments and attack characteristics. Because the experiments were conducted offline, the assessment was done without the influence of dynamic scenarios like time variation of network and latency from UAV communication during experiments, which may affect detection effectiveness. Moreover, only a selection of machine learning algorithms were taken into consideration - excluding sophisticated deep learning and hybrid methods that could result in even more effective adaptability. This presents a limitation in generalisability from one diverse type of UAV platform to another and from current cyber threats to future ones, so deployment potential must be interpreted with caution (Alqahtani and Kumar, 2024).

6.6 Recommendations for Future Work

To enhance transparent decision-making, future research should include deep learning architectures, along with explainable AI. Larger and more diversified data would be positive for real-time UAV deployment for external validity. Moreover, collaborative cybersecurity services combined with federated learning may further advance the UAV defence capabilities in a state of resilience, while maintaining federated data privacy and security (Yang *et al.* 2025).

6.7 Final Conclusion

Machine learning-based GPS spoofing detection offers practical application and a groundbreaking approach to enhance UAV cybersecurity, proving that data-driven predictive

tools can enhance resilient, accurate, and scalable protection against the rising and changing nature of cyber threats.

Reflection

Reflection on Research Skills

To further develop my literature search, academic writing, and critical analysis skills throughout this dissertation, I used recent research to evaluate cybersecurity literature for UAVs, GPS spoofing and machine learning literature. I was able to develop an argument and compare other perspectives instead of using summarisation. I also felt a greater understanding of making claims to guide methodological choices based on evidence from peer-reviewed journals. This process deepened my understanding of determining research gaps and connecting theory to practice in an academic research environment.

Reflection on Technical Skills

Through implementing the full machine learning process using Python and Google Colab, I am able to enhance my technical skills greatly. I implemented, tested, and deployed data preprocessing, feature engineering, exploratory data analysis, and model development techniques like Logistic Regression, Decision Tree, Random Forest and XGBoost. I also improved at applying feature scaling, hyperparameter tuning and model evaluation techniques. These experiences reinforced my understanding of the potential of machine learning to solve real world cyber security challenges that UAV communication systems have.

Reflection on Project Management Skills

This project was great to work on to boost project management abilities with effective planning, scheduling and frequent progress monitoring. I segmented the dissertation into smaller steps and segments like Literature review, data preparation, implementation of the model, result evaluation and report preparation. I prioritised tasks and adapted my schedule when unexpected obstacles came up in order to keep the progress going. This systematic approach helped me to better manage my time and develop my problem-solving skills and showed me the significance of maintaining my goals and academic deadlines for a large research project.

Challenges Faced and Solutions

However, the project faced multiple challenges, especially in preprocessing the data and appropriately applying machine learning strategies. Tools for converting timestamps, extracting GPS coordinates and encoding categorical variables had to be carefully tested to ensure data quality. The model optimisation also included trial-and-error with various combinations of the hyperparameters until the training was successful. Analytical thinking continued to take a significant boost in debugging Python code, using Google Colab. So,

solving the issues one at a time helped my ability to solve more complex technical issues and also helped to find a more reliable system for the proposed system.

Personal and Professional Development

Through writing this dissertation, I gained a lot in my personal and professional growth by deeply engaging with and expanding my knowledge of machine learning, cybersecurity, and UAV communication systems. I became more confident in researching independently, with qualitative and quantitative observations, making inferences and coming up with conclusions from the data I collected. The project itself was also an opportunity to develop my own technical communication and analytical abilities, which are key attributes of professional practice. In the future, I feel that these experiences will assist in cybersecurity, data analytics and/or intelligent system development using artificial intelligence and autonomous technology.

Reference List

- Aftatah, M., Khalil, A. and Zebbara, K., 2026. Secure navigation through gps/ins integration: Comparative analysis of supervised deep learning and kalman filtering for precision and spoofing detection. *IEEE Access*. <https://ieeexplore.ieee.org/abstract/document/11367623/>
- Agnew, D., Del Aguila, A. and McNair, J., 2024. Enhanced network metric prediction for machine learning-based cyber security of a software-defined UAV relay network. *IEEE Access*, 12, pp.54202-54219. <https://ieeexplore.ieee.org/abstract/document/10497092/>
- Ajayi, V.O., 2023. A review on primary sources of data and secondary sources of data. Available at SSRN 5378785. <https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=5378785>
- Akintade, S., Roy, K. and Kim, S., 2025. Hybrid Deep Machine Learning Feature Selection for High-Dimensional Cybersecurity Data. *IEEE Access*. <https://ieeexplore.ieee.org/abstract/document/11184578/>
- Aldaej, A., Ahanger, T.A., Atiquzzaman, M., Ullah, I. and Yousufudin, M., 2022. Smart cybersecurity framework for IoT-empowered drones: Machine learning perspective. *Sensors*, 22(7), p.2630. <https://www.mdpi.com/1424-8220/22/7/2630>
- Alford, S. and Teater, B., 2025. Quantitative research. In *Handbook of research methods in social work* (pp. 156-171). Edward Elgar Publishing. https://www.researchgate.net/profile/Sabretta-Alford/publication/388139110_Quantitative_research/links/6888b8d94eccfb3f29c6091d/Quantitative-research.pdf#page=177
- Alharahsheh, H.H. and Pius, A., 2020. A review of key paradigms: Positivism VS interpretivism. *Global academic journal of humanities and social sciences*, 2(3), pp.39-43. https://gajrc.com/media/articles/GAJHSS_23_39-43_VMGJbOK.pdf
- Alhoraibi, L., Alghazzawi, D. and Alhebshi, R., 2024. Detection of GPS spoofing attacks in UAVs based on adversarial machine learning model. *Sensors*, 24(18), p.6156. <https://www.mdpi.com/1424-8220/24/18/6156>
- Ali, I.M., 2024. A guide for positivist research paradigm: From philosophy to methodology. *Ideology Journal*, 9(2). <http://www.ideologyjournal.com/ojs/index.php/ideology/article/download/596/380>
- Alotaibi, F., Al-Dhaqm, A. and Al-Otaibi, Y.D., 2023. A conceptual digital forensic investigation model applicable to the drone forensics field. *Engineering, Technology & Applied*

<https://mail.etasr.com/index.php/ETASR/article/download/6195/3238>

Alqahtani, H. and Kumar, G., 2024. Cybersecurity in electric and flying vehicles: threats, challenges, AI solutions & future directions. *ACM Computing Surveys*, 57(4), pp.1-34.

<https://dl.acm.org/doi/abs/10.1145/3697830>

Altaweel, A., Mukkath, H. and Kamel, I., 2023. Gps spoofing attacks in fanets: A systematic literature review. *IEEE Access*, 11, pp.55233-55280.

<https://ieeexplore.ieee.org/abstract/document/10139317/>

Attaullah, H.M., Harris, M., Khan, I.U., Alam, M.M. and Su'ud, M.M., 2026. Comparative analysis of ml-based intrusion detection system for cyber-physical uav system. *Journal of Computational and Cognitive Engineering*, 5(1), pp.44-52.

<https://ojs.bonviewpress.com/index.php/JCCE/article/view/5886>

Bose, S.C., 2021. GPS spoofing detection by neural network machine learning. *IEEE Aerospace and Electronic Systems Magazine*, 37(6), pp.18-31.

<https://ieeexplore.ieee.org/abstract/document/9656691/>

Boumahdi, A., Azmi, M., Zegrari, M., Eddermoug, N., Tazili, S. and Ettalibi, A., 2025. Feature Selection in Cybersecurity: A Comparative Study of Machine Learning Models. *Procedia Computer Science*, 265, pp.140-148.

<https://www.sciencedirect.com/science/article/pii/S1877050925022161>

Choi, W., Choi, T. and Heo, S., 2023. A comparative study of automated machine learning platforms for exercise anthropometry-based typology analysis: performance evaluation of AWS SageMaker, GCP VertexAI, and MS Azure. *Bioengineering*, 10(8), p.891.

<https://www.mdpi.com/2306-5354/10/8/891>

de Puisseau, C.W., Peters, J., Doerpelkus, C., Tercan, H. and Meisen, T., 2023. schlably: A Python framework for deep reinforcement learning based scheduling experiments. *SoftwareX*, 22, p.101383. <https://www.sciencedirect.com/science/article/pii/S2352711023000791>

Eshmawi, A.A., Umer, M., Ashraf, I. and Park, Y., 2024. Enhanced machine learning ensemble approach for securing small unmanned aerial vehicles from GPS spoofing attacks. *IEEE Access*, 12, pp.27344-27355.

<https://ieeexplore.ieee.org/iel7/6287639/10380310/10416854.pdf>

Ferrão, I.G., Da Silva, S.A., Pigatto, D.F. and Branco, K.R., 2020, November. GpGPSpoofing: Detecting gps fraud in unmanned aerial vehicles. In *2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in*

Education (WRE) (pp. 1-6). IEEE. https://www.researchgate.net/profile/Isadora-Ferrao/publication/348336032_GPS_Spoofing_Detecting_GPS_Fraud_in_Unmanned_Aerial_Vehicles/links/6076f968a020ab5f2175a2c6/GPS-Spoofing-Detecting-GPS-Fraud-in-Unmanned-Aerial-Vehicles.pdf

Ghanad, A., 2023. An overview of quantitative research methods. *International journal of multidisciplinary research and analysis*, 6(08), pp.3794-3803. https://www.researchgate.net/profile/Anahita-Ghanad/publication/373370007_An_Overview_of_Quantitative_Research_Methods/links/67b28ffc645ef274a48341a1/An-Overview-of-Quantitative-Research-Methods.pdf

Hassler, S.C., Mughal, U.A. and Ismail, M., 2023. Cyber-physical intrusion detection system for unmanned aerial vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 25(6), pp.6106-6117. <https://ieeexplore.ieee.org/abstract/document/10368002/>

Jan, M., Spangaro, A., Lenartowicz, M. and Mattiazzi Usaj, M., 2024. From pixels to insights: Machine learning and deep learning for bioimage analysis. *BioEssays*, 46(2), p.2300114. <https://onlinelibrary.wiley.com/doi/abs/10.1002/bies.202300114>

Joshi, A., Spilbergs, A. and Mīkelsone, E., 2024, June. AI-enabled drone autonomous navigation and decision making for defence security. In *ENVIRONMENT. TECHNOLOGY. RESOURCES. Proceedings of the International Scientific and Practical Conference* (Vol. 4, pp. 138-143). <https://archive-journals.rtu.lv/etr/article/download/5071/5011>

Jung, J.H., Hong, M.Y. and Yoon, J.W., 2025. GPS Spoofing Attacks on Autonomous Navigation Systems for UAVs. *IEEE Access*. <https://ieeexplore.ieee.org/iel8/6287639/10820123/11202405.pdf>

Jung, J.H., Hong, M.Y., Choi, H. and Yoon, J.W., 2024. An analysis of GPS spoofing attack and efficient approach to spoofing detection in PX4. *IEEE Access*, 12, pp.46668-46677. <https://ieeexplore.ieee.org/abstract/document/10480421/>

Junzhi, L., Haitao, W., Jinfeng, G., Fang, L., Yu, Z., Gangqiang, L. and Yu, H., 2025. Performance testing and analysis of a new GNSS spoofing detection method in different spoofing scenarios. *IEEE Access*. <https://ieeexplore.ieee.org/abstract/document/10937204/>

Kamiri, J. and Mariga, G., 2021. Research methods in machine learning: A content analysis. *International Journal of Computer and Information Technology* (2279-0764), 10(2). <https://pdfs.semanticscholar.org/6c30/e1e3606ad9a8c244618c0ce0f5b613e7807c.pdf>

Khan, A.A., Khan, M.M., Khan, K.M., Arshad, J. and Ahmad, F., 2021. A blockchain-based decentralized machine learning framework for collaborative intrusion detection within UAVs. *Computer Networks*, 196, p.108217.

<https://pureportal.coventry.ac.uk/files/42737190/Binder8.pdf>

Kumar, S. and Ujire, D.K., 2024. Inductive and deductive approaches to qualitative research. *International Journal of Multidisciplinary Educational Research*, 13(1), pp.58-63. [http://ijmer.s3.amazonaws.com/pdf/volume13/volume13-issue1\(4\)/9.pdf](http://ijmer.s3.amazonaws.com/pdf/volume13/volume13-issue1(4)/9.pdf)

Lee, S., Seo, H. and Kim, D., 2023. Digital forensic research for analyzing drone pilot: Focusing on dji remote controller. *Sensors*, 23(21), p.8934. <https://www.mdpi.com/1424-8220/23/21/8934>

Ly, B. and Ly, R., 2021. Cybersecurity in unmanned aerial vehicles (UAVs). *Journal of cyber security technology*, 5(2), pp.120-137. <https://www.academia.edu/download/99432481/23742917.2020.pdf>

Markoulidakis, I. and Markoulidakis, G., 2024. Probabilistic confusion matrix: a novel method for machine learning algorithm generalized performance analysis. *Technologies*, 12(7), p.113. <https://www.mdpi.com/2227-7080/12/7/113>

Masadeh, A.E., Alhafnawi, M., Salameh, H.A.B., Musa, A. and Jararweh, Y., 2022. Reinforcement learning-based security/safety uav system for intrusion detection under dynamic and uncertain target movement. *IEEE Transactions on Engineering Management*, 71, pp.12498-12508. <https://ieeexplore.ieee.org/abstract/document/9763080/>

Mijwil, M.M. and Aljanabi, M., 2024. A comparative analysis of machine learning algorithms for classification of diabetes utilizing confusion matrix analysis. *Baghdad Science Journal*, 21(5), p.24. <https://bsj.uobaghdad.edu.iq/home/vol21/iss5/24/>

Mohammed, U.M., Omolara, A.E., Abiodun, O.I., Rasheed, J., Osman, O., Lar, P.M., Adeyinka, P.O. and Olugbenga, A.G., 2025. Cyber threat in drone systems: bridging real-time security, legal admissibility, and digital forensic solution readiness. *Frontiers in Communications and Networks*, 6, p.1661928. <https://www.frontiersin.org/journals/communications-and-networks/articles/10.3389/frcmn.2025.1661928/pdf>

Namagytyn, P., Brzozowski, M., Dyka, Z. and Langendoerfer, P., 2023. GPS-spoofing attack detection mechanism for UAV swarms. *arXiv preprint arXiv:2301.12766*.
<https://arxiv.org/pdf/2301.12766>

Namagembe, C.O., Ibrahim, M., Rahman, M.A. and Pillai, P., 2026. Machine Learning-Based GPS Spoofing Detection and Mitigation for UAVs. *Computers, Materials, & Continua*, 86(2), p.1. <https://search.proquest.com/openview/d30b7b28c095d97ae86f985a2b29cafa/1?pq-origsite=gscholar&cbl=2048737>

Nayfeh, M., Li, Y., Al Shamaileh, K., Devabhaktuni, V. and Kaabouch, N., 2023. Machine learning modeling of GPS features with applications to UAV location spoofing detection and classification. *Computers & Security*, 126, p.103085.
<https://www.sciencedirect.com/science/article/am/pii/S0167404822004771>

Ogab, M., Zaidi, S., Bourouis, A. and Calafate, C.T., 2025. Machine learning-based intrusion detection systems for the Internet of Drones: A systematic literature review. *IEEE Access*.
<https://ieeexplore.ieee.org/abstract/document/11018757/>

Okoli, C., 2023. Inductive, abductive and deductive theorising. *International Journal of Management Concepts and Philosophy*, 16(3), pp.302-316.
<https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=3774317>

Prakash, V., Williams, A., Garg, L., Barik, P. and Dhanaraj, R.K., 2022. Cloud-based framework for performing digital forensic investigations. *International Journal of Wireless Information Networks*, 29(4), pp.419-441. <https://link.springer.com/article/10.1007/s10776-022-00560-z>

Rahman, K., Aziz, M.A., Kashif, A.U. and Cheema, T.A., 2022. Detection of security attacks using intrusion detection system for UAV networks: A survey. In *Big Data analytics and computational intelligence for cybersecurity* (pp. 109-123). Cham: Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-031-05752-6_7

Sathyanarayanan, S. and Tantri, B.R., 2024. Confusion matrix-based performance evaluation metrics. *African Journal of Biomedical Research*, 27(4S), pp.4023-4031.
https://www.researchgate.net/profile/Sathyanarayanan-Swaminathan/publication/386347454_Confusion_Matrix-Based_Performance_Evaluation_Metrics/links/674e9e453d17281c7df6b440/Confusion-Matrix-Based-Performance-Evaluation-Metrics.pdf

Schäfer, G., Schirl, M., Rehrl, J., Huber, S. and Hirlaender, S., 2024, August. Python-based reinforcement learning on simulink models. In *International Conference on Soft Methods in Probability and Statistics* (pp. 449-456). Cham: Springer Nature Switzerland.
https://link.springer.com/chapter/10.1007/978-3-031-65993-5_55

Shafik, W., Matinkhah, S.M. and Shokoor, F., 2023. Cybersecurity in unmanned aerial vehicles: A review. *International Journal on Smart Sensing and Intelligent Systems*, 16(1).
https://www.researchgate.net/profile/Wasswa-Shafik-2/publication/374807059_Cybersecurity_in_Unmanned_Aerial_Vehicles_a_Review/links/65300c635d51a8012b530cd3/Cybersecurity-in-Unmanned-Aerial-Vehicles-a-Review.pdf

Shafique, A., Mehmood, A. and Elhadeif, M., 2021. Detecting signal spoofing attack in UAVs using machine learning models. *IEEE access*, 9, pp.93803-93815.
<https://ieeexplore.ieee.org/iel7/6287639/6514899/09456965.pdf>

Sharifani, K. and Amini, M., 2023. Machine learning and deep learning: A review of methods and applications. *World Information Technology and Engineering Journal*, 10(07), pp.3897-3904. <https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=4458723>

Sharma, P. and Gillanders, J., 2022. Cybersecurity and forensics in connected autonomous vehicles: A review of the state-of-the-art. *IEEE Access*, 10, pp.108979-108996.
<https://ieeexplore.ieee.org/abstract/document/9916257/>

Shrestha, R., Omidkar, A., Roudi, S.A., Abbas, R. and Kim, S., 2021. Machine-learning-enabled intrusion detection system for cellular connected UAV networks. *Electronics*, 10(13), p.1549. <https://www.mdpi.com/2079-9292/10/13/1549>

Sibe, R.T. and Bekom, D., 2025. Digital Forensic Investigation of an Unmanned Aerial Vehicle (UAV): A Technical Case Study of a DJI Phantom III Professional Drone. *Journal of Cybersecurity & Information Management*, 15(1).
https://www.researchgate.net/profile/Robinson_Sibe/publication/383043176_Digital_Forensic_Investigation_of_an_Unmanned_Aerial_Vehicle_UAV_A_Technical_Case_Study_of_a_DJI_Phantom_III_Professional_Drone/links/66b9fff951aa0775f27aa231/Digital-Forensic-Investigation-of-an-Unmanned-Aerial-Vehicle-UAV-A-Technical-Case-Study-of-a-DJI-Phantom-III-Professional-Drone.pdf

- Sihag, V., Choudhary, G., Choudhary, P. and Dragoni, N., 2023. Cyber4Drone: A systematic review of cyber security and forensics in next-generation drones. *Drones*, 7(7), p.430. <https://www.mdpi.com/2504-446X/7/7/430>
- Squatrito, A., 2024. Machine Learning-based GPS Jamming and Spoofing Detection. <https://commons.erau.edu/cgi/viewcontent.cgi?article=1842&context=edt>
- Stoykova, R. and Franke, K., 2023. Reliability validation enabling framework (RVEF) for digital forensics in criminal investigations. *Forensic Science International: Digital Investigation*, 45, p.301554. <https://www.sciencedirect.com/science/article/pii/S266628172300063X>
- Studiawan, H., Grispos, G. and Choo, K.K.R., 2023. Unmanned aerial vehicle (UAV) forensics: the good, the bad, and the unaddressed. *Computers & Security*, 132, p.103340. https://www.researchgate.net/profile/George-Grispos-2/publication/371683832_Unmanned_Aerial_Vehicle_UAV_Forensics_The_Good_The_Bad_and_the_Unaddressed/links/674658ed3d17281c7de19f8b/Unmanned-Aerial-Vehicle-UAV-Forensics-The-Good-The-Bad-and-the-Unaddressed.pdf
- Suliman, M.H., Liu, M., Gursay, M.C. and Kong, F., 2025. Path planning for UAVs under GPS permanent faults. *ACM Transactions on Cyber-Physical Systems*, 8(3), pp.1-23. <https://dl.acm.org/doi/pdf/10.1145/3653074>
- Sun, C., Fontanesi, G., Canberk, B., Mohajerzadeh, A., Chatzinotas, S., Grace, D. and Ahmadi, H., 2024. Advancing uav communications: A comprehensive survey of cutting-edge machine learning techniques. *IEEE Open Journal of Vehicular Technology*, 5, pp.825-854. <https://ieeexplore.ieee.org/abstract/document/10531095/>
- Sun, Y., Yu, M., Wang, L., Li, T. and Dong, M., 2023. A deep-learning-based GPS signal spoofing detection method for small UAVs. *Drones*, 7(6), p.370. <https://www.mdpi.com/2504-446X/7/6/370>
- Taherdoost, H., 2021. Data collection methods and tools for research; a step-by-step guide to choose data collection technique for academic and business research projects. *International journal of academic research in management (IJARM)*, 10(1), pp.10-38. <https://hal.science/hal-03741847v1/file/Data%20Collection%20Methods%20and%20Tools%20for%20Research%20.pdf>

Tang, A.C., 2021. A review on cybersecurity vulnerabilities for urban air mobility. In *AIAA Scitech 2021 Forum* (p. 0773).
<https://ntrs.nasa.gov/api/citations/20205011115/downloads/A%20Review%20of%20Cybersecurity%20Vulnerabilities%20for%20UAM%20Final%20Draft.pdf>

Wang, J., Nie, L., Gu, Z., Wang, J., Tan, R. and Kumari, S., 2025. Real-time GPS spoofing detection in consumer drones through multi-sensor data fusion and TimesNet. *IEEE Transactions on Consumer Electronics*.
<https://ieeexplore.ieee.org/abstract/document/10964312/>

Yang, Z., Zhang, Y., Zeng, J., Yang, Y., Jia, Y., Song, H., Lv, T., Sun, Q. and An, J., 2025. AI-driven safety and security for UAVs: From machine learning to large language models. *Drones*, 9(6), p.392. <https://www.mdpi.com/2504-446X/9/6/392>

Zeng, Q., Bashir, A. and Nait-Abdesselam, F., 2025, September. Uavids-2025: A benchmark dataset for intrusion detection in uav networks using machine learning techniques. In *2025 IEEE Conference on Communications and Network Security (CNS)* (pp. 1-9). IEEE.
<https://ieeexplore.ieee.org/abstract/document/11194990/>

Image References

Alhoraibi, L., Alghazzawi, D. and Alhebshi, R., 2024. Detection of GPS spoofing attacks in UAVs based on adversarial machine learning model. *Sensors*, 24(18), p.6156.
<https://www.mdpi.com/1424-8220/24/18/6156>

Basan, E., Basan, A., Nekrasov, A., Fidge, C., Sushkin, N. and Peskova, O., 2021. GPS-spoofing attack detection technology for UAVs based on Kullback–Leibler divergence. *Drones*, 6(1), p.8. <https://www.mdpi.com/2504-446X/6/1/8>

Singh, V., Sahana, S.K. and Bhattacharjee, V., 2025. A novel CNN-GRU-LSTM based deep learning model for accurate traffic prediction. *Discover Computing*, 28(1), p.38.
<https://link.springer.com/content/pdf/10.1007/s10791-025-09526-0.pdf>

Researchmethodscommunity, 2024. *Sage Research Methods Community*. [online] *Sage Research Methods Community*. Available at:
<https://researchmethodscommunity.sagepub.com/blog/topics-problems-methodologies>.

Appendices

Appendix 1: Dataset Link

<https://www.kaggle.com/datasets/datasetengineer/drone-communication-dataset>

Appendix 2: Google Colab Link

<https://colab.research.google.com/drive/1wvlzOd2FPU1u9yrK6WvCBJPRhhrewrkU?usp=sharing>

Appendix 3: Full Python Code

```
# PART 1: DATA LOADING, EXPLORATION, PREPROCESSING & EDA
# Dissertation: GPS Spoofing Detection in UAV Communication

# 1. Import Libraries

import warnings
warnings.filterwarnings("ignore")

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.preprocessing import LabelEncoder, StandardScaler

# 2. Load Dataset

path = "/content/drone_communication_dataset.csv"
df = pd.read_csv(path)

print("="*60)
print("DATASET LOADED SUCCESSFULLY")
print("="*60)

print("\nFirst 5 Rows")
print(df.head())

print("\nLast 5 Rows")
print(df.tail())

print("\nDataset Shape:", df.shape)

# 3. Dataset Information

print("\nDataset Information")
print(df.info())

print("\nSummary Statistics")
print(df.describe())

print("\nMissing Values")
print(df.isnull().sum())

print("\nDuplicate Rows:", df.duplicated().sum())

print("\nData Types")
print(df.dtypes)

# 4. Feature Engineering

# Convert timestamp
df['timestamp'] = pd.to_datetime(df['timestamp'])

df['hour'] = df['timestamp'].dt.hour
df['day'] = df['timestamp'].dt.day
df['month'] = df['timestamp'].dt.month
```



```
df['timestamp'] = df['timestamp'].dt.strftime('%Y-%m-%d %H:%M:%S')

# Split GPS Coordinates
gps = df['drone_gps_coordinates'].str.strip("(").str.split(",", expand=True)

df['latitude'] = gps[0].astype(float)
df['longitude'] = gps[1].astype(float)

# Drop Original Columns
df.drop(['timestamp', 'drone_gps_coordinates'], axis=1, inplace=True)

# Encode Categorical Columns
le = LabelEncoder()

for col in ['communication_protocol', 'encryption_type']:
    df[col] = le.fit_transform(df[col])

print("\nFeature Engineering Completed")

# 5. Feature Scaling

target = "label_gps_spoofing"

features = [c for c in df.columns if c != target]

scaler = StandardScaler()

df[features] = scaler.fit_transform(df[features])

print("Feature Scaling Completed")

# 6. Exploratory Data Analysis

# ----- Graph 1 -----
plt.figure(figsize=(6,4))
counts = df[target].value_counts()
counts.plot(kind="bar")

plt.title("GPS Spoofing Class Distribution")
plt.xlabel("Class")
plt.ylabel("Count")

for i,v in enumerate(counts):
    plt.text(i,v,str(v),ha='center')

plt.show()

print("\nClass Distribution")
print(counts)

# ----- Graph 2 -----
plt.figure(figsize=(8,6))
corr = df.corr(numeric_only=True)
plt.imshow(corr,cmap="coolwarm")
plt.colorbar()
plt.title("Correlation Heatmap")
plt.xticks([])
plt.yticks([])
plt.show()

print("\nHighest Correlation")
```



```
print("\nHighest Correlation")
print(corr[target].sort_values(ascending=False).head(10))

# ----- Graph 3 -----
plt.figure(figsize=(6,4))
plt.hist(df["signal_strength"],bins=30)

plt.title("Signal Strength Distribution")
plt.xlabel("Signal Strength")
plt.ylabel("Frequency")

plt.show()

print("\nSignal Strength Statistics")
print(df["signal_strength"].describe())

# ----- Graph 4 -----
protocol_counts = pd.read_csv(path)["communication_protocol"].value_counts()

plt.figure(figsize=(6,4))
protocol_counts.plot(kind="bar")

plt.title("Communication Protocol")
plt.xlabel("Protocol")
plt.ylabel("Count")

for i,v in enumerate(protocol_counts):
    plt.text(i,v,str(v),ha='center')

plt.show()

print("\nProtocol Counts")
print(protocol_counts)

# ----- Graph 5 -----
plt.figure(figsize=(8,4))
plt.boxplot([
    df["signal_strength"],
    df["round_trip_time"],
    df["packet_loss_rate"]
],
labels=["Signal", "RTT", "Packet Loss"])

plt.title("Boxplot of Important Features")
plt.show()

print("\nMedian Values")
print(df[["signal_strength",
    "round_trip_time",
    "packet_loss_rate"]].median())

# 7. Save Cleaned Dataset

df.to_csv("/content/cleaned_drone_dataset.csv",index=False)

print("\nCleaned Dataset Saved Successfully")

print("\nFinal Dataset Shape:",df.shape)
```



PART 2 : MACHINE LEARNING MODELS

```
import pandas as pd
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split, cross_val_score, GridSearchCV
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.ensemble import RandomForestClassifier
from xgboost import XGBClassifier

# Load Cleaned Dataset

df = pd.read_csv("/content/cleaned_drone_dataset.csv")

X = df.drop("label_gps_spoofing", axis=1)
y = df["label_gps_spoofing"]

X_train,X_test,y_train,y_test=train_test_split(
    X,y,test_size=0.2,random_state=42,stratify=y)

# Train Models

models={
    "Logistic Regression":LogisticRegression(max_iter=1000),
    "Decision Tree":DecisionTreeClassifier(random_state=42),
    "Random Forest":RandomForestClassifier(random_state=42),
    "XGBoost":XGBClassifier(eval_metric="logloss",random_state=42)
}

results={}

for name,model in models.items():
    model.fit(X_train,y_train)
    acc=cross_val_score(model,X,y,cv=5,scoring="accuracy").mean()
    results[name]=round(acc,4)

print("\nCross Validation Accuracy")
print(pd.DataFrame(results.items(),columns=["Model","Accuracy"]))

# Hyperparameter Tuning

params={
    "max_depth":[3,5],
    "n_estimators":[100,200]
}

grid=GridSearchCV(
    XGBClassifier(eval_metric="logloss"),
    params,
    cv=3,
    scoring="accuracy")

grid.fit(X_train,y_train)

best_model=grid.best_estimator_

print("\nBest Parameters")
print(grid.best_params_)
```



```
print("Best Accuracy:", round(grid.best_score_, 4))

# Graph 1 : Decision Tree

plt.figure(figsize=(10, 6))
plot_tree(
    models["Decision Tree"],
    max_depth=2,
    filled=True,
    feature_names=X.columns)

plt.title("Decision Tree")
plt.show()

print("Decision Tree Depth:",
      models["Decision Tree"].get_depth())

# Graph 2 : Feature Importance

importance = pd.Series(
    best_model.feature_importances_,
    index=X.columns).sort_values(ascending=False).head(10)

plt.figure(figsize=(8, 5))
importance.plot(kind="bar")

plt.title("Top 10 Important Features")
plt.ylabel("Importance")
plt.show()

print("\nTop 10 Features")
print(importance)
```