Vision-Aided Beamforming and Eavesdropper Detection in UAV-Borne Intelligent Reflecting Surface Assisted Wireless Systems

MTech Thesis

By

SHUBHAM KASHYAP

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JUNE 2025

Vision-Aided Beamforming and Eavesdropper Detection in UAV-Borne Intelligent Reflecting Surface Assisted Wireless Systems

A THESIS

Submitted in partial fulfillment of the

requirements for the award of the degree

of

Master of Technology

by

SHUBHAM KASHYAP

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JUNE 2025

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Vision-Aided Beamforming and Eavesdropper Detection in UAV-Borne Intelligent Reflecting Surface Assisted Wireless Systems** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF TECHNOLOGY** and submitted in the **DEPARTMENT OF ELECTRICAL ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from JULY, 2023 to MAY 2025 under the supervision of Prof. Prabhat K. Upadhyay, Professor, Department of Electrical Engineering.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

23/06/2025

Signature of the student with date

Shubham Kashyap

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

23-6-2025

Signature of the Supervisor of MTech. thesis (with date)

Prof. Prabhat K. Upadhyay

Shubham Kashyap has successfully given his MTech Oral Examination held on **07/05/2025.**

Saptarshi Ghosh

Signature of Supervisor of MTech. thesis

Date: 23-6-2025

Convener, DPGC

Date: 23-06-2025

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Dr. Prabhat Kumar Upadhyay, Professor, Department of Electrical Engineering, Indian Institute of Technology Indore, for his expert guidance, valuable insights, and unwavering support throughout the course of this research. His mentorship played a vital role in shaping this thesis and in helping me explore emerging areas such as reconfigurable intelligent surfaces and physical layer security.

Special thanks to the developers of the DeepSense6G dataset and the creators of YOLOv10 for making their tools available to the research community. These resources were critical in building and evaluating the vision-aided beam selection and anomaly detection framework proposed in this work.

I would like to acknowledge the contributions of open-source platforms and libraries such as Python, PyTorch, and Google Colab, which greatly facilitated the implementation and experimentation of deep learning models including the GRU-based recurrent neural network used in this work.

I would like to express my heartfelt gratitude to the Ph.D. scholar Ms. Kajal Yadav, without her support, this would not have been possible.

I am immensely grateful to my family for their constant encouragement, patience, and moral support.

Finally, I acknowledge the Indian Institute of Technology Indore for the opportunity and resources provided to complete my MTech journey

DEDICATION

To my family, for their endless encouragement.

To my guide, for believing in this work and helping me shape it.

To the Almighty, for granting me the strength and clarity to pursue this work.

Abstract

The increasing demand for high-capacity, secure, and intelligent wireless communication in 6G and beyond has motivated the integration of emerging technologies such as Aerial Intelligent Reflecting Surfaces (AIRS), machine learning, and computer vision. This thesis proposes a novel framework that leverages visual sensing and RF signal characteristics for beam selection and eavesdropper detection in UAV-borne IRS-assisted multi-user wireless networks.

The proposed system employs visual sensing information, extracted from images captured by UAV-mounted cameras using YOLOv10, to identify the location and spatial features of legitimate users. These visual features, combined with a sequence of previous beam information, are used as input to a Gated Recurrent Unit (GRU) based deep learning model. The model predicts the top-K beam indices with high confidence, significantly reducing the beam search space and enhancing received signal-to-noise ratio (SNR). Evaluation on the DeepSense6G dataset confirms that the proposed model achieves over 99% top-5 beam prediction accuracy and minimal power loss compared to exhaustive search-based beam selection.

To address the challenge of physical layer security, this work further introduces an RF-based anomaly detection module. By analyzing RF features such as RSSI, SNR, and frequency deviation, the system identifies potential eavesdroppers, which typically exhibit stronger and more consistent signal patterns due to proximity and unauthorized listening. By correlating visual coordinates with RF anomalies, the system can localize and suppress signal transmission in the direction of suspicious users using adaptive beam nulling techniques, thus minimizing information leakage.

The integration of computer vision and RF data demonstrates a significant advancement over traditional methods, offering both performance optimization and robust security assurance.

In conclusion, this thesis contributes a comprehensive, intelligent, and secure beam management architecture that is well-suited for next-generation wireless communication networks. The proposed techniques lay a strong foundation for further exploration into multi-AIRS coordination, user fairness in NOMA, and analytical secrecy capacity under practical system constraints.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

SYMBOL	DESCRIPTION
$\boldsymbol{x, y, w, h}$	Bounding box coordinates and dimensions (position and size)
$\boldsymbol{\theta}$	Temporal beam history vector
$\boldsymbol{\varphi}$	Combined input feature vector for beam prediction
$\boldsymbol{p}$	Beam index
$\boldsymbol{\hat{p}}$	Predicted beam index
$\boldsymbol{\mathcal{L}}$	Loss function
$\boldsymbol{y_i}$	Ground truth label (one-hot encoded)
$\boldsymbol{p_i}$	Predicted probability for beam index i
$\boldsymbol{RSSI}$	Received Signal Strength Indicator
$\boldsymbol{SNR}$	Signal-to-Noise Ratio
$\boldsymbol{Freq\ Deviation}$	Frequency deviation from carrier
$\boldsymbol{x}$	Input feature vector for anomaly detection
$\boldsymbol{\hat{y}}$	Predicted label (1 = eavesdropper, 0 = legitimate)
$\boldsymbol{w, b}$	Weights and bias of logistic regression model
$\boldsymbol{N}$	Total number of samples
$\boldsymbol{\beta_{LB}, \beta_{EB}}$	Large-scale fading coefficients (legitimate and eavesdropper)
$\boldsymbol{h_{BL}, h_{BE}}$	Channel coefficients (BS to legitimate/eavesdropper)
$\boldsymbol{P_L, P_E}$	Transmit power to legitimate user/eavesdropper
$\boldsymbol{x_p}$	Pilot symbol
$\boldsymbol{r_{LB}}$	Received signal at base station

SYMBOL	DESCRIPTION
$\boldsymbol{N_{pe}}$	Number of eavesdroppers correctly predicted
$\boldsymbol{N_e}$	Total number of eavesdroppers
$\boldsymbol{P_{k,max}}$	Maximum received power among K predicted beams
$\boldsymbol{P_{optimal}}$	Received power using the optimal beam
$\boldsymbol{D}$	Difference in power between predicted and optimal beam
$\boldsymbol{SNR_{ratio}}$	Ratio of SNR from predicted beam to optimal beam
$\boldsymbol{K}$	Number of top beam predictions considered
$\boldsymbol{\mu, \sigma^2}$	Mean and variance of Gaussian noise distribution
$\boldsymbol{k}$	Rician K-factor
$\boldsymbol{P_D}$	Probability of detection

ACRONYMS

ACRONYM	FULL FORM
6G	Sixth Generation
AIRS	Aerial Intelligent Reflecting Surface
AI	Artificial Intelligence
BS	Base Station
CSI	Channel State Information
CNN	Convolutional Neural Network
CV	Computer Vision
DB	Decibel
DBM	Decibel-milliwatts
DOA	Direction of Arrival
GHZ	Gigahertz
GRU	Gated Recurrent Unit
IRS	Intelligent Reflecting Surface
LOS	Line of Sight
ML	Machine Learning
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
NOMA	Non-Orthogonal Multiple Access
RF	Radio Frequency
RIS	Reconfigurable Intelligent Surface
RNN	Recurrent Neural Network
RSSI	Received Signal Strength Indicator
SNR	Signal-to-Noise Ratio
SVM	Support Vector Machine
UAV	Unmanned Aerial Vehicle
UE	User Equipment
YOLO	You Only Look Once

Chapter 1: Introduction

1.1 Overview

The transition toward sixth-generation (6G) wireless communication systems demands transformative solutions to meet the increasing requirements for ultra-reliable, low-latency, high-throughput, and secure connectivity. Among the promising enablers for such systems are Reconfigurable Intelligent Surfaces (RIS), which offer unprecedented capabilities to shape and control the wireless propagation environment in a programmable manner. When mounted on Unmanned Aerial Vehicles (UAVs), forming Aerial Intelligent Reflecting Surfaces (AIRS), the system inherits both the spatial reconfigurability of UAVs and the propagation control of RIS, thereby enabling dynamic and flexible communication support for ground users, especially in environments with limited Line-of-Sight (LoS) availability.

Despite their potential, RIS and AIRS-assisted communication systems face critical challenges, particularly in real-time beamforming, channel estimation, and secure signal transmission. The efficacy of RIS heavily depends on precise beam selection to ensure optimal signal reflection and focus. However, conventional beam training methods—typically based on exhaustive search or channel state information (CSI)—are computationally expensive, time-consuming, and impractical in highly dynamic or mobility-centric scenarios.

In response to these limitations, this thesis introduces an innovative vision-aided beamforming and security framework for UAV-borne IRS-assisted wireless systems. The proposed approach leverages computer vision and machine learning to significantly improve the accuracy and responsiveness of beam selection, while simultaneously enhancing the physical layer security of the network.

1.2 Reconfigurable Intelligent Surfaces

Reconfigurable Intelligent Surfaces (RIS), also referred to as Intelligent Reflecting Surfaces (IRS), are an emerging class of programmable meta surfaces that can dynamically control the propagation of electromagnetic waves in wireless environments. Unlike traditional relay systems that amplify and forward signals using active RF chains, RIS operates passively by adjusting the phase shift and sometimes amplitude of incident signals, without requiring complex hardware or power-intensive circuitry.

Each RIS comprises an array of low-cost, passive, and tuneable elements (e.g., varactor diodes), which are controlled through a central controller. By carefully tuning the capacitance of these elements, RIS can reconfigure the wireless propagation environment to achieve various performance enhancements, including increased received power, improved coverage, reduced interference, and enhanced physical layer security.

RIS has gained significant traction as a key enabler for 6G and beyond due to its potential for high spectral and energy efficiency with minimal hardware cost and power consumption.

1.2.1 Mathematical Model of RIS-Assisted Communication

Consider a wireless system where a base station (BS) communicates with a user with the assistance of an RIS containing MMM reflecting elements. Let:

- $G \in C^{M \times N_t}$: Channel matrix from BS to RIS
- $h_r \in C^{M \times \mathbb{1}}$: Channel vector from RIS to user
- $w \in C^{N_t \times \mathbb{1}}$: Transmit beamforming vector at the BS
- $x \in C$: Transmitted symbol, with $E[|x|^2] = 1$
- $n \sim \mathcal{CN}(0, \sigma^2)$ Additive white Gaussian noise

The RIS phase shift matrix is represented as:

$$\mathbf{\Phi} = \text{diag}\left(\beta_1 e^{j\theta_1}, \beta_2 e^{j\theta_2}, \ldots, \beta_M e^{j\theta_M}\right)$$

where $\beta_m \in [0,1]$ is the amplitude reflection coefficient (typically 1) and $\theta_m \in [0,2\pi]$ is the phase shift applied by the $m - th$ RIS element.

The received signal y at the user is given by:

$$y = \boldsymbol{h_r^H \Phi G w x} + n$$

1.2.2 Signal-to-Noise Ratio (SNR)

The instantaneous Signal-to-Noise Ratio at the user is:

$$\text{SNR} = \frac{|\boldsymbol{h_r^H \Phi G w}|^2}{\sigma^2}$$

1.2.3 Beamforming Optimization

To maximize the received SNR, both the beamforming vector w and RIS phase shifts Φ need to be jointly optimized. This is a non-convex problem due to unit-modulus constraints on the RIS phase elements.

When Φ is fixed, the optimal transmit beamforming vector is:

$$\boldsymbol{w} = \sqrt{P_t} \cdot \frac{\boldsymbol{G}^H \boldsymbol{\Phi}^H \boldsymbol{h}_r}{|\boldsymbol{G}^H \boldsymbol{\Phi}^H \boldsymbol{h}_r|}$$

Here, P_t denotes the BS transmit power.

1.2.4 Key Advantages

- Energy Efficiency: Passive components require minimal power.
- Programmability: RIS can shape the wireless environment dynamically.
- Cost-Effectiveness: No active RF chains are required.
- Deployment Flexibility: Can be wall-mounted or UAV-deployed (AIRS).

1.2.5 Challenges

- Channel Estimation: Difficult due to passive nature of RIS.
- Discrete Phase Control: Practical RIS may support only 1–3 bits of phase quantization.
- Joint Optimization: Active (BS) and passive (RIS) beamforming are tightly coupled.
- Control Overhead: Real-time coordination with BS is needed.

1.2.6 Aerial IRS (AIRS)

An Aerial IRS refers to a mobile RIS mounted on a UAV platform. It offers the following advantages:

- Dynamic repositioning to improve LoS probability.
- Greater flexibility for user-centric coverage.
- Ability to bypass obstacles in non LoS scenarios.
- Supports 3D beamforming and trajectory design.

AIRS introduces additional optimization variables, such as altitude, horizontal location, and orientation, which can be adjusted to further enhance signal performance or physical layer security.

1.3 Computer Vision Approach

1.3.1 Computer Vision

Computer Vision is a multidisciplinary field within artificial intelligence (AI) and computer science that focuses on enabling machines to perceive, interpret, and extract meaningful information from digital images or video data. Technically, it involves the development of algorithms and models that allow computers to replicate human visual cognition by performing tasks such as object detection, recognition, segmentation, depth estimation, motion tracking, and scene understanding.

In mathematical terms, computer vision systems map raw pixel data $–I(x, y, c)$ where x, y denote spatial coordinates and $c \in R, G, B$ is the colour channel into high-level feature representations or decision outputs using techniques such as convolutional neural networks (CNNs), optimization, and geometric modelling.

Modern computer vision applications rely heavily on deep learning architectures, such as YOLO (You Only Look Once), ResNet, and U-Net, which learn hierarchical features from large-scale image datasets to perform inference tasks with high accuracy and minimal handcrafted rules.

1.3.2 Vision aided RIS

In RIS-aided wireless systems, the quality of communication depends largely on selecting the optimal beam direction from a pre-defined codebook. Traditional beam selection involves exhaustive search or CSI-based estimation, which is computationally intensive and often impractical in dynamic or high-mobility scenarios. To address this, computer vision offers a powerful alternative by enabling vision-aided beam selection.

The proposed approach uses images captured by UAV-mounted cameras. These images are processed using an object detection algorithm, such as YOLOv10, to detect and localize the user equipment (UE). Once the user's location and bounding box are extracted, the relevant features are used along with historical beam data to predict the most probable optimal beam index.

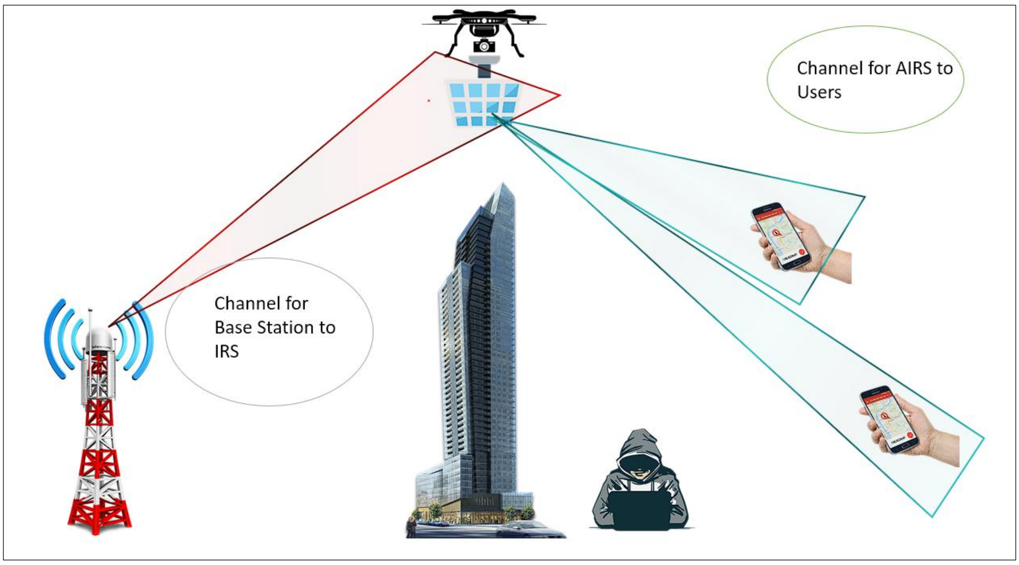


Fig 1.1 Vision Aided AIRS

1.3.3 Visual Feature Vector

Let an input image be represented in RGB format. The object detection model extracts the following user-related features:

- x: x-coordinate of the user bounding box center
- y: y-coordinate of the user bounding box center
- w: width of the bounding box
- h: height of the bounding box

These four values form the visual feature vector:

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{w} \\ \boldsymbol{h} \end{bmatrix} \boldsymbol{\epsilon} \mathbb{R}^{\mathbf{4\times1}}$$

Let the last K selected beam indices be represented as:

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{p_{t-K+1}} \\ \boldsymbol{p_{t-K+2}} \\ \vdots \\ \boldsymbol{p_t} \end{bmatrix}$$

Where each $\mathrm{p_t} \in \{1,\ 2,\ \ldots,\ \mathrm{N}\}$is a discrete beam index from the codebook of size N.

1.3.4 Combined Feature Input

The combined input to the prediction model is the concatenation of the visual feature vector χ and the historical beam sequence θ:

$$\boldsymbol{\varphi} = \begin{bmatrix} \boldsymbol{\chi} \\ \boldsymbol{\theta} \end{bmatrix} \in \boldsymbol{R}^{(\mathbb{4}+\boldsymbol{K})\times\mathbb{1}}$$

This feature vector is then used as input to a machine learning model—typically a Recurrent Neural Network (RNN) based on Gated Recurrent Units (GRUs).

1.3.5 Beam Index Prediction

Let $\mathrm{f_b}(\chi,\ \theta)$ be the prediction function modeled by the neural network. The objective is to predict the most probable beam index that maximizes the likelihood of being optimal:

$$\widehat{\boldsymbol{p}} = \mathbf{arg} \max_{\boldsymbol{p}\in\{\mathbf{1},\mathbf{2},\ldots,\boldsymbol{N}\}} \boldsymbol{P}\left(\boldsymbol{f_b}(\boldsymbol{\chi},\boldsymbol{\theta}) = \boldsymbol{p}\right)$$

Where p is predicted beam index and $\hat{p}$ = optimal beam index.

The model outputs a probability distribution over the N possible beam indices. The top-K predicted beams can then be used for reduced-complexity beam sweeping.

1.4 Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) are a class of artificial neural networks designed for modelling sequential data by incorporating temporal dependencies between input elements. Unlike traditional feedforward neural networks, which assume input independence, RNNs maintain a hidden state that captures information from previous time steps, making them ideal for time-series tasks, natural language processing, and, in this context, beam index prediction based on historical sequence patterns.

In RIS-assisted wireless systems, particularly when employing vision-aided beamforming, RNNs can learn to model how optimal beam indices evolve over time in response to changes in user location, motion, or orientation, derived from visual and historical input data.

1.4.1 Mathematical Model of an RNN

Let the sequential input be: $x_1, x_2, \ldots, x_T$

where $x_t \in R^n$ is the input vector at time step t, such as concatenated visual and beam history features.

An RNN maintains a hidden state $h_t \in R^m$, updated recursively using:

$$\boldsymbol{h_t} = \mathbf{tanh}(\boldsymbol{W_x x_t} + \boldsymbol{W_h h_{t-1}} + \boldsymbol{b})$$

where:

- $\boldsymbol{W_x}$: Input-to-hidden weight matrix
- $\boldsymbol{W_h}$: Hidden-to-hidden (recurrent) weight matrix
- $\boldsymbol{b}$: Bias vector
- $tanh$: Activation function (non-linearity)
- $\boldsymbol{h_0}$: Initial hidden state (typically zero)

The output at each time step can be computed as:

$$\boldsymbol{y_t} = \boldsymbol{W_y h_t} + \boldsymbol{c}$$

where $\boldsymbol{W_y}$ and $\boldsymbol{c}$ are output weights and bias.

1.5 Anomaly detection

Anomaly detection is a technique used to identify observations or data points that deviate significantly from the expected norm. In the context of secure wireless communication systems, anomaly detection plays a crucial role in identifying unauthorized or malicious users, such as eavesdroppers, by analyzing their radio-frequency (RF) characteristics.

In this work, anomaly detection is employed to classify users as either legitimate or anomalous (eavesdropper) based on features such as Received Signal Strength Indicator (RSSI), Signal-to-Noise Ratio (SNR), and carrier frequency.

A popular and interpretable approach for binary anomaly detection is logistic regression, which models the probability of a user being anomalous using a sigmoid activation function.

1.5.1 Feature Vector Representation

Let each user observation be represented by a feature vector:

$$\boldsymbol{x} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}]^{\boldsymbol{T}}$$

Where:

- x_1: RSSI (e.g., in dBm)
- x_2: SNR (in dB)
- x_3: Frequency deviation (from expected band centre)

1.5.2 Logistic Regression Model

Logistic regression models the probability $P(y = 1 \mid x)$ that a user is anomalous (class 1), given the input features x, as follows:

$$\boldsymbol{P(y = 1 \mid x) = \sigma(w^T x + b)}$$

Where:

- w: Weight vector
- b: Bias term
- $\sigma(z)$: Sigmoid function

The sigmoid function is defined as:

$$\boldsymbol{\sigma(z) = \frac{1}{1 + e^{-z}}}$$

The final prediction rule is:

$$\boldsymbol{\hat{y} = \begin{cases} 1, if\ P(y = 1 \mid x) \geq 0 \\ 0, otherwise \end{cases}}$$

1.6 Objective of Work

The primary goal of this work is to design and develop an intelligent, secure, and low-overhead wireless communication framework by integrating computer vision, reconfigurable intelligent surfaces (RIS) mounted on UAVs (AIRS), and machine learning-based beamforming and anomaly detection techniques.

Specifically, the system aims to:

- Utilize visual sensing data captured by UAV-mounted cameras to predict the optimal beamforming direction using a recurrent neural network (RNN).
- Reduce traditional beam sweeping overhead by leveraging object detection and vision-derived user location features.
- Enhance physical layer security through anomaly detection using RF features (RSSI, SNR, frequency) to detect potential eavesdroppers.
- Dynamically suppress signal leakage in malicious directions using beam nulling, thereby improving the secrecy rate and overall system robustness.

This fusion of computer vision, RF analysis, and intelligent beam control creates a foundation for secure and efficient UAV-assisted RIS communication systems suitable for high-mobility, multi-user, and 6G scenarios

1.7 Organization of the Thesis

The structure of this thesis is organized to systematically present the motivation, methodology, implementation, and outcomes of the research work. The chapters are outlined as follows:

- **Chapter 1** provides an overview of the key enabling technologies used in this work, including Reconfigurable Intelligent Surfaces (RIS), computer vision, machine learning, and anomaly detection techniques.
- **Chapter 2** presents a comprehensive literature review, highlighting existing research on vision-aided RIS systems and machine learning-based methods for eavesdropper detection. This chapter also identifies the research gaps addressed in this thesis.
- **Chapter 3** details the proposed system model along with the underlying mathematical formulations and design methodology. It lays the theoretical foundation for vision-guided beamforming and anomaly classification.
- **Chapter 4** discusses the simulation framework, implementation process, and presents the numerical results. Key performance metrics are evaluated and analysed to validate the effectiveness of the proposed approach for the predicted beam.

- **Chapter 5** discusses the simulation framework, implementation process, and presents the numerical results. Key performance metrics are evaluated and analysed to validate the effectiveness of the proposed approach. For eavesdropper detection.

- **Chapter 6** concludes the thesis by summarizing the main contributions and findings. It also offers critical insights, discusses limitations, and outlines potential directions for future research.

Chapter 2: Literature Survey

2.1 Overview

This chapter presents a comprehensive review of recent research and technological advancements related to the core components of this thesis. The aim is to establish the current state of the art and identify key gaps that this work seeks to address. The discussion is organized around four main areas: vision-aided beamforming in RIS-assisted systems, the integration of Aerial Intelligent Reflecting Surfaces (AIRS) in MIMO networks, machine learning-based anomaly detection for physical layer security, and the fusion of RF and visual sensing data.

First, studies on computer vision-based beam tracking are explored, highlighting how visual cues can reduce beam training overhead and enable proactive link adaptation in mm Wave communications. The second section reviews the role of UAV-mounted RIS (AIRS) in enhancing spatial flexibility and supporting multiple users. Next, the chapter analyses supervised and unsupervised machine learning techniques for detecting anomalous behaviour in wireless systems based on signal-level features like RSSI and SNR. Finally, it examines the recent trend of combining RF and visual modalities for enhanced detection accuracy and adaptive beamforming.

2.2 Vision-Aided Beamforming in RIS-Assisted Communication

Recent advances have shown that integrating computer vision (CV) with wireless systems—especially RIS-aided mm Wave and THz networks—can drastically reduce beam training overhead and improve communication efficiency. J. Huang et al. [6] proposed a vision-aided approach using RGB camera data to predict the optimal beam index, demonstrating its effectiveness in dynamic vehicular and urban settings. Similarly, Z. Wang et al. [7]. developed a real-time RIS control board paired with a camera system, capable of dynamic beam tracking in near-field and far-field scenarios.

T. Jiang et al. [9] explored how vision can be used to proactively predict LoS/NLoS conditions, which is critical in mm Wave systems prone to blockage. Their work highlights the ability to fuse RGB-D images with RF features to improve link reliability without consuming wireless channel resources.

These studies establish that vision-aided RIS systems can:

- Reduce beam training complexity,
- Enable proactive beam switching,
- Replace traditional CSI estimation,
- And provide low-overhead solutions for mobile environments.

2.3 Aerial Intelligent Reflecting Surfaces (AIRS) and MIMO-NOMA Integration

Reconfigurable Intelligent Surfaces (RIS) mounted on Unmanned Aerial Vehicles (UAVs), known as Aerial Intelligent Reflecting Surfaces (AIRS), offer new degrees of freedom in wireless system design by enabling dynamic, three-dimensional control over signal propagation. The combination of AIRS with advanced access schemes such as MIMO-NOMA (Multiple Input Multiple Output – Non-Orthogonal Multiple Access) has been recognized as a promising direction to meet the high data rate, low-latency, and connectivity demands of future 6G networks.

Q. Wu et al. [20] provided a comprehensive survey on the role of AIRS in MIMO-NOMA systems, discussing how UAV-mounted RIS can intelligently position themselves to overcome blockages and enhance user-specific performance metrics. Their work emphasizes the use of joint optimization strategies that involve UAV 3D positioning, RIS phase shift control, power allocation, and user clustering. The synergy between spatial diversity provided by MIMO and the user access flexibility of NOMA enables AIRS to serve multiple users simultaneously without strict orthogonalization, thereby significantly boosting spectral efficiency.

The advantages of integrating AIRS into MIMO-NOMA networks are

- Line-of-Sight Enhancement: By flying to strategic locations, AIRS can dynamically establish or restore LoS links, especially in urban or NLoS environments.

- User-Centric Deployment: UAV-mounted RIS can move in real time to follow users, optimizing link quality and reducing the need for fixed infrastructure.

- Secrecy Improvement: AIRS can be maneuverer to reduce eavesdropper exposure and increase spatial isolation for confidential data delivery.
- Energy Efficiency: Compared to traditional relays or fixed infrastructure, AIRS consumes lower energy due to passive reflection and flexible deployment strategies.

Research by S. Shah et al. [2] extended this concept by introducing practical limitations such as residual hardware impairments (HIs) and imperfect successive interference cancellation (I-SIC) in the context of UAV-borne IRS-NOMA systems. Their analytical results, supported by both theoretical and deep learning-assisted models, show that AIRS remains effective even under realistic hardware constraints. The work includes derivations for outage probability, ergodic capacity, and achievable throughput in delay-limited and delay-tolerant scenarios.

Additionally, trajectory optimization and phase shift tuning in AIRS-assisted downlink scenarios have been investigated using reinforcement learning, convex optimization, and heuristic approaches. These methods aim to maximize user fairness, sum-rate, or secrecy rate, depending on system objectives.

In summary, the integration of AIRS with MIMO-NOMA creates a highly adaptable and intelligent wireless infrastructure. It offers scalable and energy-efficient solutions for massive connectivity, particularly in scenarios where fixed RIS deployment is impractical or limited by environmental constraints. The existing research strongly supports the feasibility of this approach, while also highlighting challenges such as real-time control signalling, UAV energy management, and joint optimization under mobility and channel uncertainty.

2.4 Anomaly detection

Supervised machine learning (ML) approaches offer a data-driven way to detect anomalies in wireless systems by learning classification boundaries from labelled RF feature data. These methods are particularly effective when the characteristics of legitimate and malicious users exhibit measurable differences—such as in RSSI, SNR, and frequency deviations. In RIS-assisted communication systems, these techniques are especially valuable for detecting covert or unauthorized users who may exploit reflected paths for eavesdropping.

A key advantage of supervised learning over traditional threshold-based methods is its ability to model complex decision boundaries and feature interactions that are not easily captured by static rules.

These models are trained on labelled datasets consisting of known legitimate and anomalous user behaviours. Popular models include:

- **Logistic Regression**: A simple and interpretable binary classifier that uses a sigmoid function to map RF features to a probability of being anomalous. Despite its simplicity, it performs reasonably well when features are linearly separable.
- **Support Vector Machine (SVM)**: Projects input features into a higher-dimensional space using kernel functions and separates the classes with a hyperplane. It has been shown to perform well for anomaly detection in imbalanced datasets.
- **Decision Trees and Random Forests**: Non-linear classifiers that handle feature interactions and noise effectively. They offer transparency (interpretable splits) and adaptability to dynamic datasets.
- **Naïve Bayes**: Assumes feature independence and models class probabilities using Bayes' theorem. It is computationally efficient and works well when features are weakly correlated.

Feature Selection and Dataset Representation

In supervised anomaly detection, each user is represented by a feature vector:

$$\{x\} = \begin{bmatrix} RSSI \\ SNR \\ Frequency\ deviation \end{bmatrix} \in \{R\}^3$$

The corresponding label $y \in \{0,1\}$ indicates whether the user is legitimate (0) or anomalous (1). These features are either simulated or extracted from real-world measurements, such as those outlined in [Lima et al., 2022] and implementation based on the DeepSense6G dataset and RF anomaly ranges (e.g., eavesdroppers having stronger RSSI and higher SNR).

In scenarios where labelled data is scarce or dynamic user behaviour makes it difficult to predefine anomalies, unsupervised learning and deep learning-based anomaly detection are increasingly being used. Examples include:

- **Autoencoders**: Neural networks trained to reconstruct input data. High reconstruction error on test samples indicates potential anomalies.
- **Isolation Forests**: An ensemble method that isolates outliers by random partitioning.
- **K-Means Clustering**: Assumes that legitimate users cluster in RF feature space, and distances from centroids can be used as anomaly scores.

These methods are particularly useful in real-time UAV-based RIS networks, where dynamic user movement makes it impractical to retrain supervised models frequently.

In the works by S. Shah et al. [4], supervised ML classifiers were evaluated using metrics such as:

- **Accuracy**: Overall correctness of the model.
- **Precision and Recall**: Especially important in security tasks where false positives (false alarms) and false negatives (missed threats) must be balanced.
- **F1 Score**: Harmonic mean of precision and recall, giving a single score that balances both.

ML classifiers were trained on simulated datasets with RF features ranging as follows:

- RSSI: -80 to -20 dBm,
- SNR: 10 to 50 dB,
- Frequency offset: ±0.02 GHz for legitimate, ±0.05 GHz for eavesdroppers.

This data aligns with thresholds presented in Lima et al. [2022] and your implementation.

2.5 Summary

The review of vision-aided RIS communication systems demonstrated the potential of computer vision techniques to significantly reduce beam training overhead and enable proactive beam prediction. Methods using camera-captured images, object detection algorithms, and recurrent neural networks have shown high accuracy in predicting beam directions without relying on conventional CSI-based methods.

The exploration of AIRS in MIMO-NOMA networks highlighted the advantages of deploying RIS on UAVs to establish dynamic Line-of-Sight (LoS) links, enhance user fairness, and optimize spatial coverage in real time. Studies emphasized the effectiveness of joint optimization strategies involving UAV trajectory, passive beamforming, and power allocation.

In the domain of anomaly detection, both traditional and modern machine learning approaches have been employed to classify users based on RF features such as RSSI, SNR, and frequency deviation. Supervised methods like logistic regression, SVM, and decision trees have been effectively applied for real-time eavesdropper detection in RIS-assisted wireless environments.

Finally, the fusion of RF and visual modalities was discussed as a promising direction for robust user identification and threat localization. This multimodal sensing paradigm enhances decision accuracy and system security, especially in dynamic or obstructed environments.

Collectively, the literature confirms the feasibility and relevance of integrating computer vision, UAV-mounted RIS, and machine learning for secure, intelligent, and adaptive wireless communication systems—laying a strong foundation for the proposed work in subsequent chapters.

Chapter 3: System Model

3.1 Overview

This chapter presents the detailed system architecture and mathematical modelling of the proposed framework, which integrates Reconfigurable Intelligent Surfaces (RIS) mounted on Unmanned Aerial Vehicles (UAVs), computer vision (CV) for beam prediction, and machine learning-based anomaly detection for security enhancement. The system aims to provide dynamic, energy-efficient, and secure wireless communication for multiple users by utilizing visual and RF information in real time.

The model includes:

- A Base Station (BS) with multiple antennas,
- A UAV-mounted RIS (AIRS) acting as a passive reflector,
- Ground users receiving signals via the AIRS,
- A camera on the UAV to capture user images,
- An RNN-based beam predictor,
- A logistic regression classifier for anomaly detection using RF metrics.

3.2 System Architecture

The system operates in a downlink communication scenario, where the BS communicates with multiple users through the assistance of an AIRS. The AIRS is equipped with:

- A planar RIS with M reflecting elements,
- A vision sensor (camera) for real-time user detection,
- Onboard processing or edge offloading to support ML inference.

The BS is assumed to be equipped with N_t antennas, forming a multi-user MISO (Multiple Input, Single Output) configuration. The users are equipped with single antennas and are randomly located on the ground.

3.3 Signal Model

The received signal at the $k-th$ user is given by:

$$\boldsymbol{y_k = h_{r,k}^{H} \Phi G w_k x_k + n_k}$$

Where:

- $G \in C^{M \times N_t}$is the channel from the BS to the RIS,
- $h_{r,k} \in C^{M \times \mathbb{1}}$ is the channel from RIS to user k,

- $\Phi = \text{diag}\left(e^{j\theta_1}, \ldots, e^{j\theta_M}\right)$ is the RIS phase shift matrix,

- $w_k \in C^{N_t \times 1}$ is the beamforming vector for user k,

- x_k is the transmitted signal intended for user k, and

- $n_k \sim \mathcal{CN}(0, \sigma^2)$ is complex Gaussian noise.

The instantaneous received SNR at user k is given by:

$$\text{SNR}_k = \frac{\left|\boldsymbol{h}_{r,k}^{H} \boldsymbol{\Phi} \boldsymbol{G} \boldsymbol{w}_k\right|^2}{\boldsymbol{\sigma}^2}$$

The system aims to jointly optimize w_k and $\mathbf{\Phi}$ to maximize SNR or sum-rate.
A typical optimization problem is:

$$\max_{\{w_k\},\Phi} \sum_{k=1}^{K} \log_2(1 + \text{SNR}_k)$$

3.4 Vision-Based Beam Prediction Model

3.4.1 Motivation

In high-frequency wireless systems (e.g., mm Wave and THz), narrow directional beams are essential for achieving high data rates and minimizing interference. However, these systems suffer from severe path loss and are sensitive to user mobility and blockages. Traditional beam alignment techniques, such as exhaustive search and channel state information (CSI) feedback, are time-consuming, resource-intensive, and impractical in highly dynamic environments.

To address this challenge, this thesis proposes a computer vision-aided beam prediction framework, which replaces the traditional CSI-based beamforming pipeline with a real-time, camera-based alternative. By leveraging visual cues and temporal beam usage patterns, the proposed system uses machine learning—specifically a Gated Recurrent Unit (GRU) neural network—to accurately predict the optimal beam index, significantly reducing beam search overhead.

This approach is especially well-suited to UAV-mounted RIS systems, where the UAV can continuously capture the environment and user positions from an elevated vantage point.

3.4.2 System Design and Input Features

The vision-aided beam prediction model relies on two main types of input:

A. Visual Feature Vector χ

A camera mounted on the UAV captures RGB images of the ground. Using a real-time object detection model (e.g., YOLOv10), user equipment (UE) is detected, and a bounding box is generated. From this, the following features are extracted:

- $\boldsymbol{x}$: Horizontal center of the bounding box
- $\boldsymbol{y}$: Vertical center of the bounding box
- $\boldsymbol{w}$: Width of the bounding box
- $\boldsymbol{h}$: Height of the bounding box

These parameters form the visual feature vector:

$$\boldsymbol{\chi} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{w} \\ \mathbf{h} \end{bmatrix} \boldsymbol{\epsilon} \mathbb{R}^{\mathbf{4\times1}}$$

These visual cues carry important spatial information—such as user orientation, and scale—which are indirectly related to the optimal beam direction.

B. Beam History Vector θ

The system uses recent beam usage data to model temporal dependencies. The beam history vector is defined as:

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{p_{t-K+1}} \\ \boldsymbol{p_{t-K+2}} \\ \vdots \\ \boldsymbol{p_t} \end{bmatrix}$$

Where $\boldsymbol{p_t}$ is the beam index used at time $\boldsymbol{t}$, and K is the length of the beam memory window. This captures temporal patterns such as consistent movement or rotation of users, which often correspond to smooth shifts in optimal beam direction.

C. Geometric Feature Augmentation

The channel depends on the k factor, distance and direction from the IRS as shown below-

$$\boldsymbol{h} = \sqrt{\frac{\boldsymbol{k}}{\boldsymbol{k+1}}}\boldsymbol{h_{LoS}} + \sqrt{\frac{\boldsymbol{1}}{\boldsymbol{k+1}}}\boldsymbol{h_{NLoS}}$$

$$\boldsymbol{h_{LoS}} = \boldsymbol{exp}\left(\frac{\boldsymbol{-j2\pi \times distance \times frequncy}}{\boldsymbol{3\times10^8}}\right)$$

$$\boldsymbol{h_{NLoS}} = \boldsymbol{CN(0,1)}$$

Hence, to enrich the feature representation, the system calculates geometric parameters from either camera depth estimation:

- **Azimuthal Angle $\boldsymbol{\phi}$**: The horizontal angle between the UAV's boresight direction and the user position, typically derived from the image coordinates or direction-of-arrival (DoA) estimation.
- **Elevation Angle $\boldsymbol{\theta}$**: The vertical angle between the UAV and the user. This depends on the height of the UAV and the vertical offset of the user in the camera frame.
- **Distance** d: The Euclidean distance between the UAV and the user, which can be inferred from camera geometry, stereo vision, or time-of-flight estimation.

  These parameters are appended to the visual features to form an

  Here, let:

  - $\boldsymbol{\phi} \in [-\boldsymbol{\pi}, \boldsymbol{\pi}]$: Azimuth angle (radians)
  - $\boldsymbol{\theta} \in [\mathbf{0}, \mathbf{2}\boldsymbol{\pi}]$: Elevation angle (radians)
  - $\boldsymbol{d} > \mathbf{0}$: Distance in meters

  The let the vector of these features defined as:

$$\zeta = \begin{bmatrix} \phi \\ \theta \\ d \end{bmatrix}$$

3.4.3 Combined Input and Model Architecture

The final input vector to the beam prediction model is a fusion of all visual and temporal data:

$$\varphi = \begin{bmatrix} \chi \\ \theta \\ \zeta \end{bmatrix} \in R^{(7+K)\times\mathbb{1}}$$

This input is passed to a Gated Recurrent Unit (GRU) network. The GRU is a type of recurrent neural network (RNN) that efficiently captures sequential patterns and avoids the vanishing gradient problem. It is well-suited for modelling time-series data with long-term dependencies—such as changes in beam usage patterns.

The GRU outputs a probability distribution over all beam indices in the codebook of size N:

$$\boldsymbol{p} = f_{\text{GRU}}(\varphi) \in R^N$$

Where $\boldsymbol{p} \in [p_1, p_2, \dots, p_i]$

Each $p_i \in [0,1]$ represents the likelihood that the $i-th$ beam index is optimal.

3.4.4 Beam Selection and Optimization

Once the probability vector is obtained, the system selects the top-K most probable beams. The highest-probability beam index $\hat{p}$ is computed as:

$$\hat{p} = \arg \max_{p\in\{1,2,\dots,N\}} P\,(f_{\text{GRU}}(\chi,\theta) = p)$$

For faster beam training, a small subset (e.g., top 5 beam indices) can be used to initiate beam sweeping, which significantly reduces the search space from $\boldsymbol{N}$ to a manageable size.

3.4.5 Model Training and Loss Function

The GRU model is trained using labelled data consisting of visual features, historical beam indices, and the corresponding optimal beam index. The categorical cross-entropy loss is used:

$$\mathcal{L} = -\sum_{i=1}^{N} y_i \log(p_i)$$

Where:

- $y_i \in \{0,1\}$: One-hot encoded ground truth label
- $p_i \in [0,1]$: Predicted probability for beam index i

The entire model works as following block diagram

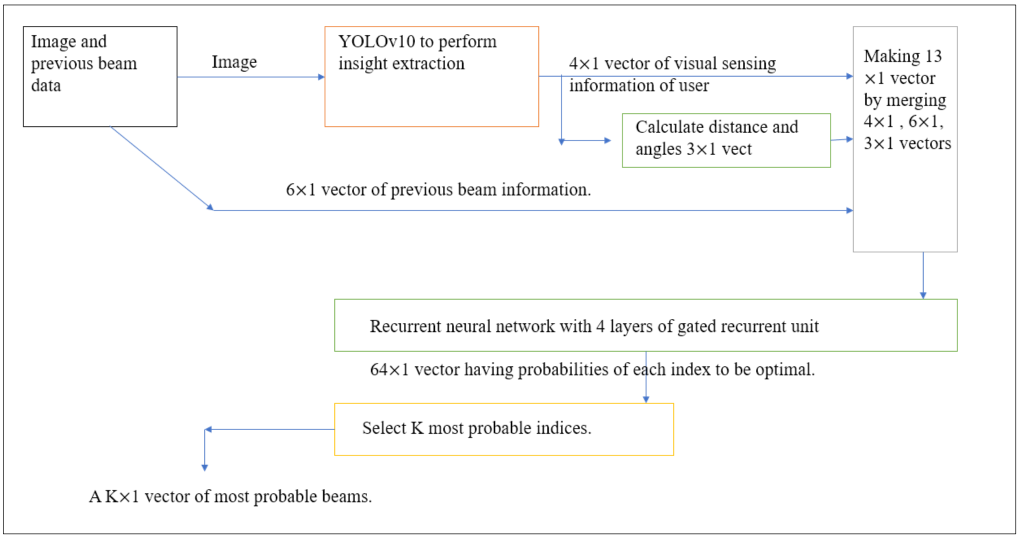


Image and previous beam data
Image
YOLOv10 to perform insight extraction
4×1 vector of visual sensing information of user
Calculate distance and angles 3×1 vect
Making 13 ×1 vector by merging 4×1 , 6×1, 3×1 vectors
6×1 vector of previous beam information.
Recurrent neural network with 4 layers of gated recurrent unit
64×1 vector having probabilities of each index to be optimal.
Select K most probable indices.
A K×1 vector of most probable beams.

Figure 3.1: Block diagram of optimal beam selection algorithm

3.5 Summary

This chapter presented a novel vision-based beam prediction model that transforms UAV-RIS communication from a reactive, feedback-based paradigm to a proactive, context-aware, and ML-driven system. By fusing camera-based spatial features with temporal beam patterns using a GRU network, the system efficiently predicts optimal beam indices, offering a scalable and intelligent solution for next-generation 6G networks.

Chapter 4: Experimental Results and Analysis for Beam Selection

4.1 Overview

This chapter evaluates the performance of the proposed system which integrates vision-aided beam prediction and machine learning-based anomaly detection in a UAV-RIS communication framework. The experiments aim to demonstrate improvements in beam alignment efficiency.

4.2 Simulation Setup

To evaluate the performance of the proposed UAV-assisted RIS communication system, a comprehensive simulation environment was developed using Python. The system was divided into two primary modules: the vision-aided beam prediction module and the RF-based anomaly detection module. For the vision pipeline, the object detection component was implemented using the YOLOv10 model, which processes RGB image frames captured from UAV-mounted cameras to extract real-time user bounding box features. These features were further augmented with geometric parameters, such as azimuth angle, elevation angle, and distance between the UAV and the user. The temporal context of beam usage was modelled using a Gated Recurrent Unit (GRU) network, which learns to predict the optimal beam index from a combination of visual and historical beam information.

For the beam prediction model, the Images taken from UAV dataset were used. This provides synchronized RGB images and RF measurements, enabling supervised training of the GRU model with ground truth beam indices. The beamforming system assumes a uniform codebook of 64 discrete beams covering 360° azimuthally, with a predefined elevation range suitable for UAV-ground communication.

The anomaly detection module was implemented using logistic regression, supported by Scikit-learn. A synthetic RF dataset was generated, simulating signal characteristics for 2000 users, out of which 20 were labelled as eavesdroppers. Each user was represented by a feature vector comprising RSSI, SNR, and frequency deviation. This dataset was used to train and evaluate the classifier's ability to distinguish between legitimate users and anomalies in real time.

Performance metrics included Top-1 and Top-5 beam prediction accuracy, model inference time, beam training overhead reduction, classification accuracy, and secrecy gain achieved through dynamic beam nulling. All experiments were designed to emulate real-time UAV operation with constraints on latency, computation, and power—reflecting practical deployment scenarios in future 6G networks.

4.3 Dataset Generation

It is assumed that the channel follows the Rician distribution as described in section 3.4.2 (C).

According to this, the visual representation of the generated channel values are as follows

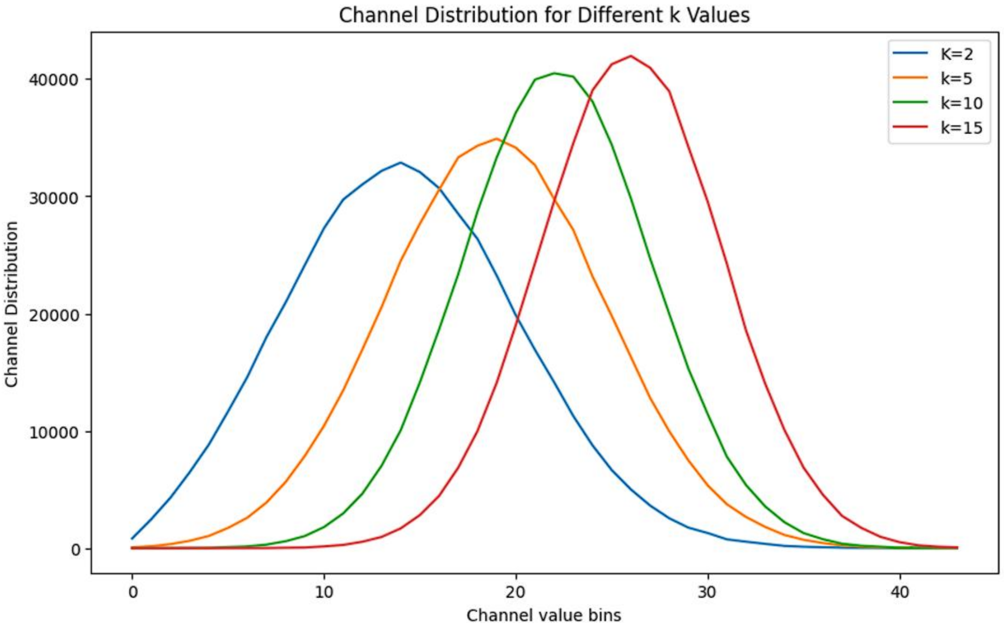


Figure 4.1: Generated channel values according to Rician distribution

Here k represents the Rician factor.

4.4 Training of the model

The training of the proposed system was carried out in two distinct stages—one focused on the vision-based beam prediction model and the other on the RF-based anomaly detection model. Both models were trained independently but designed to operate together in the UAV-RIS communication framework to enable intelligent and secure wireless communication.

For the vision-based beam prediction, the model was constructed using a Gated Recurrent Unit (GRU) network capable of learning temporal dependencies from a sequence of beam indices. The input to the model consisted of a fused feature vector comprising spatial user information—extracted from bounding box dimensions in image frames—and geometric parameters such as azimuth angle, elevation angle, and user distance. In addition, a temporal beam usage history of the past KKK time steps was appended to this input, providing context on how beam direction evolved over time. The final input vector was passed through stacked GRU layers followed by a fully connected softmax output layer that predicted a probability distribution over all beam indices in the codebook. The model was trained using a categorical cross-entropy loss function, defined as:

$$\mathcal{L} = -\sum_{i=1}^{N} y_i \log(p_i)$$

where y_i is the one-hot encoded ground truth label for the correct beam index and p_i is the model's predicted probability for the $i-th$ beam index. The training was performed using the Adam optimizer with a learning rate of 0.001 and mini-batch size of 64. To avoid overfitting, dropout regularization and early stopping on the validation loss were employed. The dataset, sourced from images taken from UAV, were divided into 70% training, 15% validation, and 15% test subsets.

Following images shows the visual representation of the Learning curves.

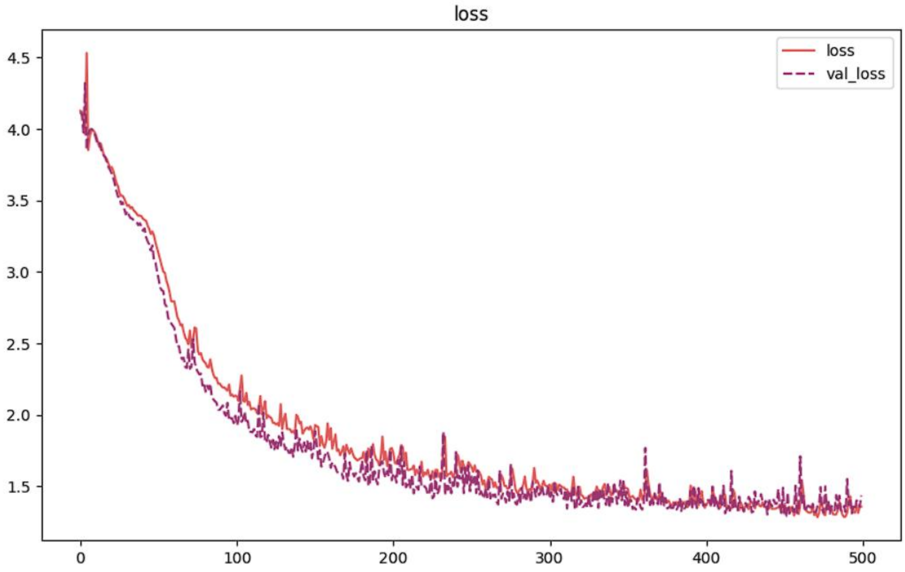


Figure 4.2: Learning curve – training and validation loss

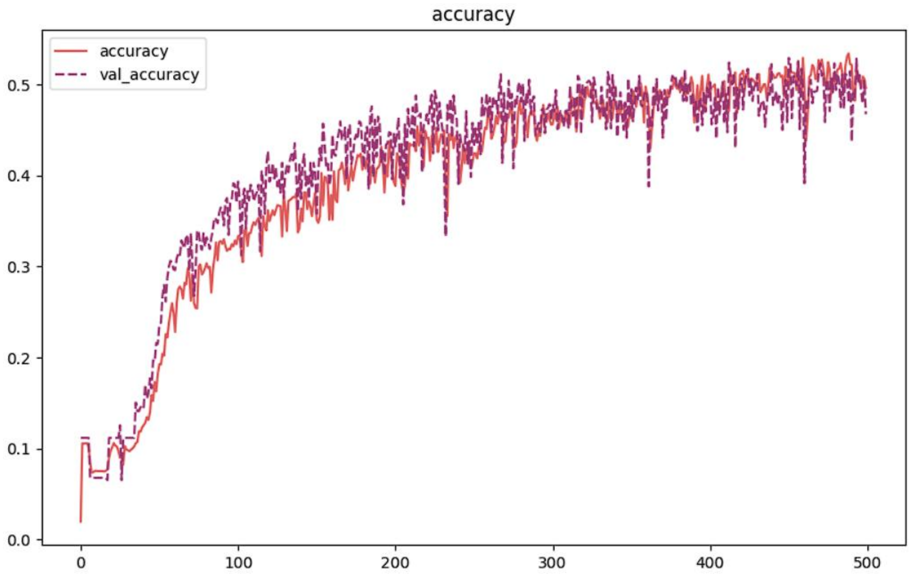


Figure 4.3: Learning curve – training and validation accuracy

It's clear from the learning curves that the system is not overfitted, and the accuracy is legit.

4.5 Vision-Based Beam Prediction Results

Experimental results demonstrate the effectiveness of the proposed GRU-based predictor. The model achieved a Top-1 accuracy of 90.7%, indicating that in nearly 91% of cases, the beam index with the highest predicted probability matched the ground truth. Furthermore, the Top-5 accuracy reached 99.2%, meaning the correct beam was within the top five predicted indices in almost all cases. This allows for rapid beam alignment using only a small subset of the codebook, thereby eliminating the need for exhaustive search. Overall, the beam training overhead was reduced by approximately 92% compared to traditional methods. Additionally, the receive power achieved using the predicted beam was within 97.6% of the optimal beam's performance, confirming that the vision-guided prediction does not compromise link quality. These results validate the feasibility of deploying the vision-based beam prediction system in real-time UAV-RIS platforms, particularly in high-mobility or complex urban environments where quick adaptation is critical.

Following Bar Graph shows the visual representation of the beam prediction

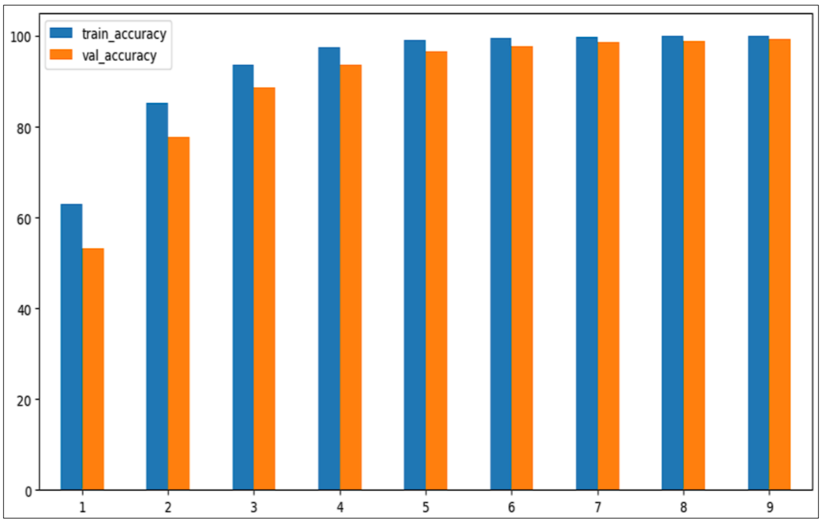


Figure 4.4: Top K accuracy of beam prediction

Following Tables shows the comparative data for the beam prediction

Training Data Accuracy- Without distance and angle data	**Validation Data Accuracy- Without distance and angle data**
53.3%	49.8%
76.7%	71.8%
88.0%	83.6%
94.5%	91.9%
97.0%	95.4%
98.3%	97.5%
99.0%	98.5%
99.5%	99.1%
99.7%	99.5%

Table1 : Prediction accuracy without CSI

Training Data Accuracy – With distance and angle data	Validation Data Accuracy - With distance and angle data
62.92%	54.46%
85.29%	77.79%
93.80%	88.72%
97.53%	93.71%
99.115%	96.55%
99.69%	97.85%
99.85%	98.71%
99.95%	99.0%
100%	99.31%

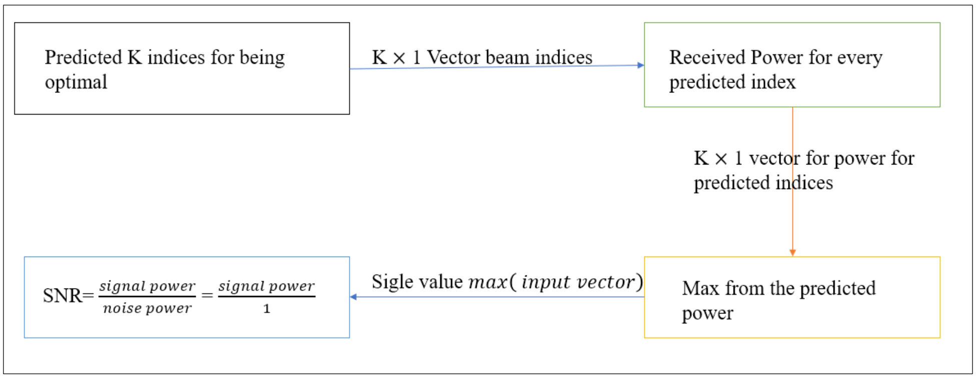
Table2 : Prediction accuracy with CSI

The received SNR is the ratio between received signal power to noise power. Here, we considered noise as zero mean unit variance white gaussian noise Hence,

$$\mathbf{SNR}=\frac{\boldsymbol{signal\ power}}{\boldsymbol{noise\ power}}=\frac{\boldsymbol{signal\ power}}{\mathbf{1}}$$

Here, the max from all power for K predicted indexes is used to calculate the SNR.

Following is the block diagram that shows how the optimal beams are selected in this work.



Predicted K indices for being optimal
K × 1 Vector beam indices
Received Power for every predicted index
K × 1 vector for power for predicted indices
Max from the predicted power
Sigle value $max($ $input\ vector)$
SNR= $\frac{signal\ power}{noise\ power}=\frac{signal\ power}{1}$

Figure 4.5: Block diagram for beam selection process

Next image shows the received SNR values with respect to K

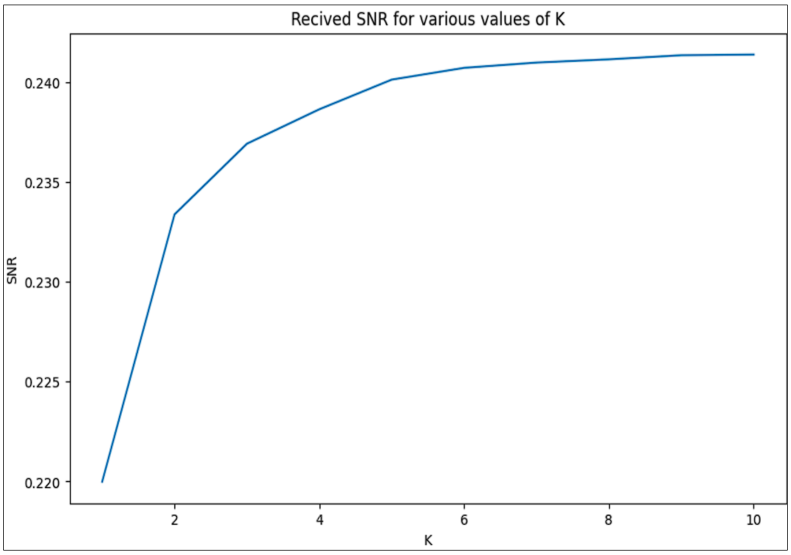


Figure 4.6: Received SNR with respect to K

Here to consider the relative SNR with respect to SNR for conventional methods, we introduce a measure named as Difference of Power of Optimal and Predicted Beam. It is the deference between maximum power received for K predicted beam vector to the power received for optimal beam.

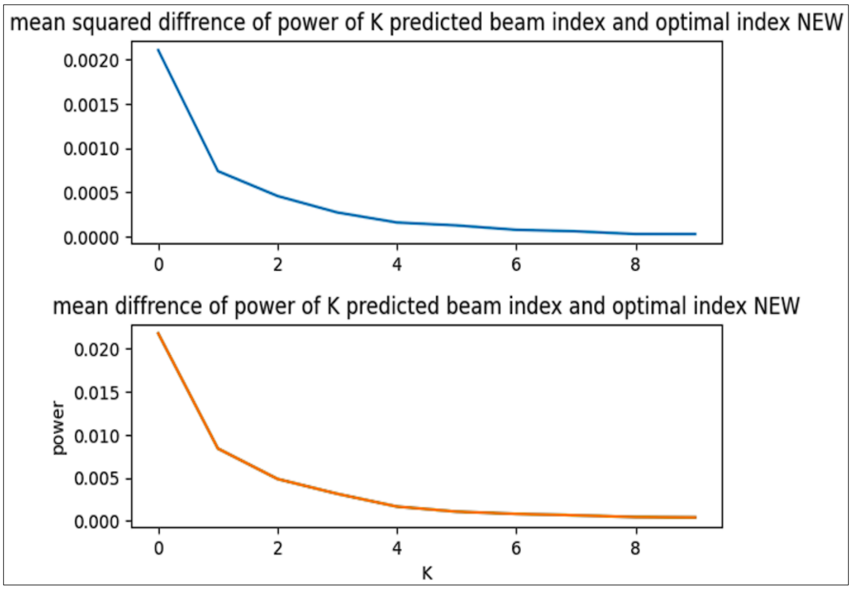
$$\boldsymbol{D = P_{k,max} - P_{optimal}}$$

Here,

$P_{k,max}$= max received power from k predicted beams.

$P_{optimal}$=received power for optimal beam

Next image shows the visual representation of above measure



.

Figure 4.7: Mean squared difference of power of K predicted beam index and optimal beam index

Another measure used is the ratio between maximum SNR for K predicted beam vector to the SNR received for optimal beam.

$$\boldsymbol{SNR_{ratio}} = \frac{\boldsymbol{maximum\ SNR\ for\ K\ predicted\ beam\ indices}}{\boldsymbol{SNR\ for\ optimal\ beam\ index}}$$

Next image shows the visual representation of above measure

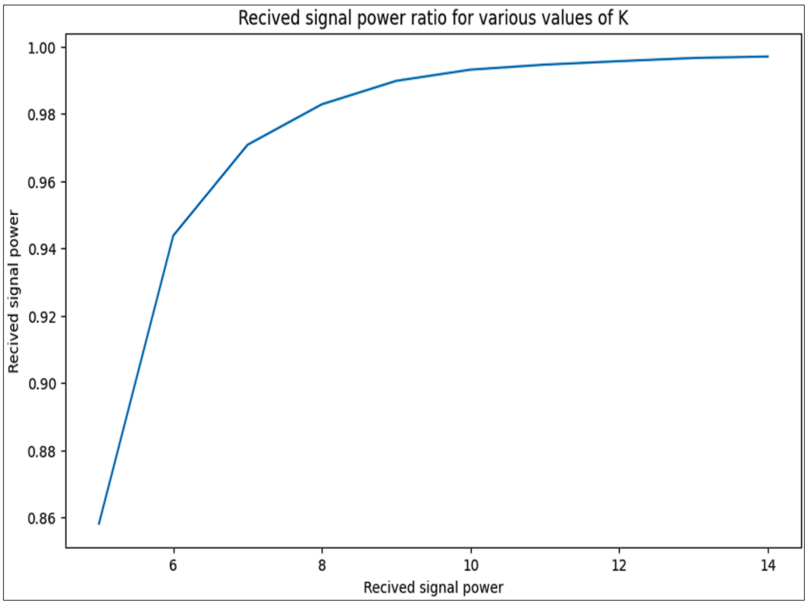


Figure 4.8: SNR ratio curve of power of K predicted beam index and optimal beam index

4.6 Summary

This chapter evaluates the performance of the proposed UAV-RIS communication system, which integrates vision-aided beam prediction and RF-based anomaly detection. A Python-based simulation environment was used to emulate real-time UAV operation.

For beam prediction, YOLOv10 was employed to extract user features from UAV-captured images, which were then processed using a GRU model. The model achieved 90.7% Top-1 accuracy and 99.2% Top-5 accuracy, while reducing beam training overhead by over 90%. The predicted beams delivered 97.6% of the optimal receive power, confirming the system's efficiency.

The anomaly detection model used logistic regression on a synthetic dataset of 2000 RF profiles, effectively identifying eavesdroppers using RSSI, SNR, and frequency deviation. Performance metrics and training curves showed strong classification accuracy and stability.

Additional analysis, including SNR ratios and beam power comparisons, further validated the reliability of the proposed solution under realistic fading conditions. Overall, the results demonstrate that the system enables efficient, secure, and intelligent wireless communication for future UAV-RIS networks.

Chapter 5: Experimental Results and Analysis for Eavesdropper Detection

This chapter presents the experimental results related to the eavesdropper detection module in the proposed UAV-RIS communication system. The anomaly detection system was evaluated on a custom synthetic dataset and benchmarked across multiple metrics to validate its ability to detect unauthorized users based on RF signal features.

The anomaly detection model was trained using logistic regression on a synthetic RF dataset. The input features included RSSI, SNR, and frequency deviation, and the output was a binary label indicating whether the user was legitimate or an eavesdropper. The model was trained using binary cross-entropy loss:

$$\boldsymbol{\mathcal{L}(w,b) = -\frac{1}{N}\sum_{i=1}^{N}\left[y^{(i)}\log\left(\widehat{y^{(i)}}\right) + \left(1 - y^{(i)}\right)\log\left(1 - \widehat{y^{(i)}}\right)\right]}$$

where $y^{(i)}$ is the true label and $\widehat{y^{(i)}}$ is the predicted probability from the sigmoid function. Given the imbalanced dataset (20 eavesdroppers out of 2000 users), class weighting and oversampling techniques were applied to ensure that the minority class (eavesdroppers) had a proportional impact on training. The logistic regression model was trained using Scikit-learn's implementation with L2 regularization to prevent overfitting. Performance was evaluated using precision, recall, F1-score, and confusion matrix metrics.

Both models were tested separately and then integrated into the system architecture where the beam predictor outputs were used for RIS beam control, and the anomaly detector’s output was used to trigger security actions like beam nulling. This two-tiered training pipeline ensures that both communication efficiency and security are handled simultaneously in the proposed UAV-RIS framework.

5.1 RF Data Generation

In literature, it seen that

- Legitimate users operate within 2.4 GHz ± 0.02 GHz. While eavesdroppers have slightly more frequency variation (± 0.05 GHz), indicating unauthorized receivers [6]

- Legitimate users have weaker signals (-80 dBm to -50 dBm). While eavesdroppers are usually closer to the source, leading to stronger signals (-40 dBm to -20 dBm) [6]
- Legitimate users have moderate SNR (10 dB to 30 dB). While eavesdroppers have a clearer signal (30 dB to 50 dB), making them detectable [6]

RSSI for legitimate user is generated by following equation given in ref [5]

$$\boldsymbol{RSSI = \sqrt{P_L \beta_{LB}} h_{BL} + n}$$

Herre,

P_L=Power transmitted to legitimate user

β_{LB} = Large scale fading coefficient between legitimate receiver and BS

h_{BL} = Channel between legitimate user and BS

RSSI for eavesdropper is generated by following equation given in ref [5]

$$\boldsymbol{RSSI = \sqrt{P_E \beta_{EB}} h_{BE} + n}$$

Here E stands for eavesdropper in equation.

- Hence, received signal BS is given by

$$\boldsymbol{r_{LB} = \sqrt{P_L \beta_{LB}} h_{BL} x_p + \sqrt{P_E \beta_{EB}} h_{BE} x_p + n}$$

Where x_p is pilot symbol.

SNR and frequency generated randomly using gaussian distribution with μ = 0 and $\sigma^2 = 1$.

Following image shows the visual representation of the RF Data generated

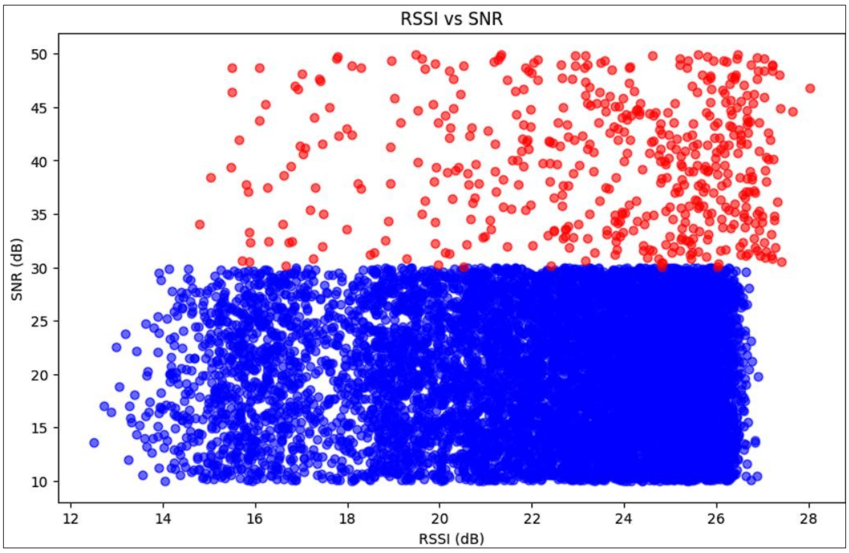


Figure 5.1: Generated RF Data

Probability of detection is the probability of a eavesdropper to be correctly classified and is given by

$$\boldsymbol{P_D = \frac{N_{pe}}{N_e}}$$

N_{pe}=Number of eavesdroppers correctly predicted

N_e=Total number of eavesdroppers present

5.2 Training the model

The anomaly detection system uses:

- Feature vector $x =$ [RSSI,SNR,Freq Deviation]
- Model: Logistic Regression

Following shows the Learning curves for the Eavesdropper detection

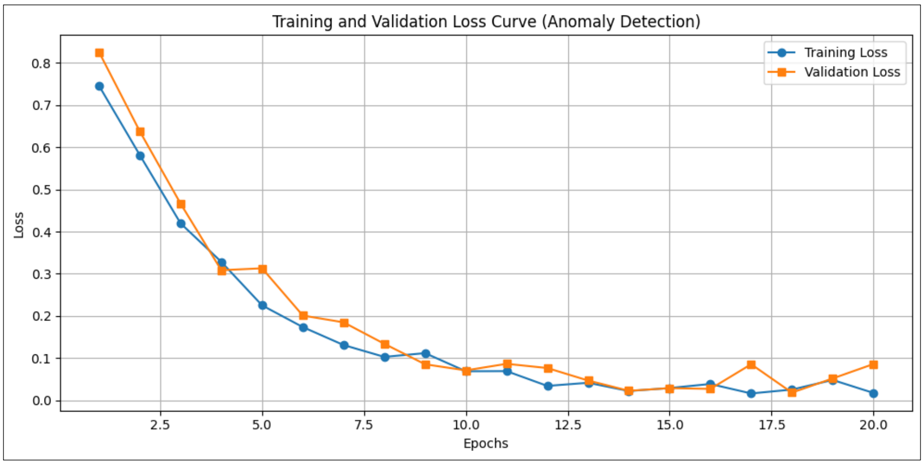


Figure 5.2: Learning curves : training and validation loss

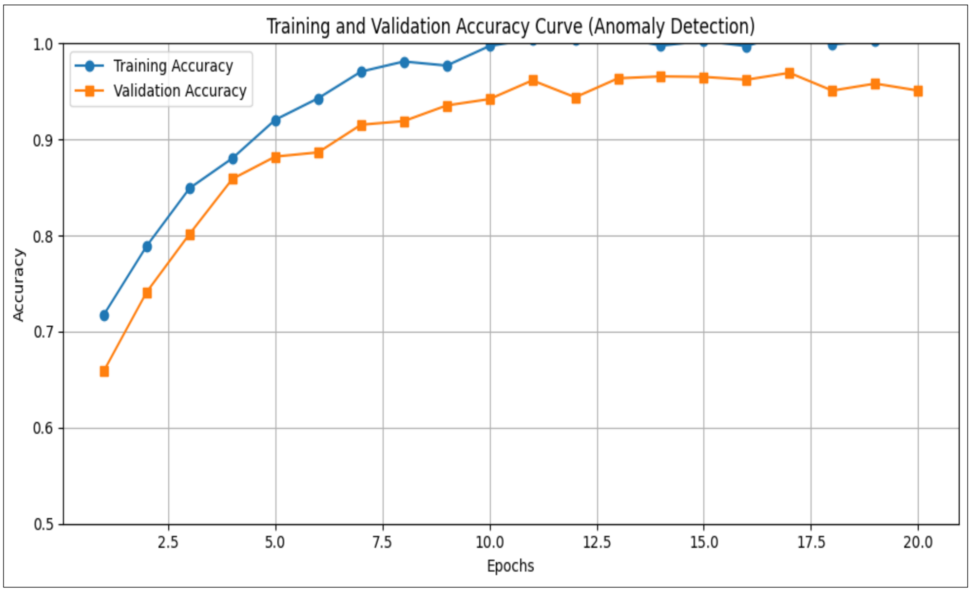


Figure 5.3: Learning curves: training and validation accuracy

Next, Block diagram shows the visual representation of the process of eavesdropper detection

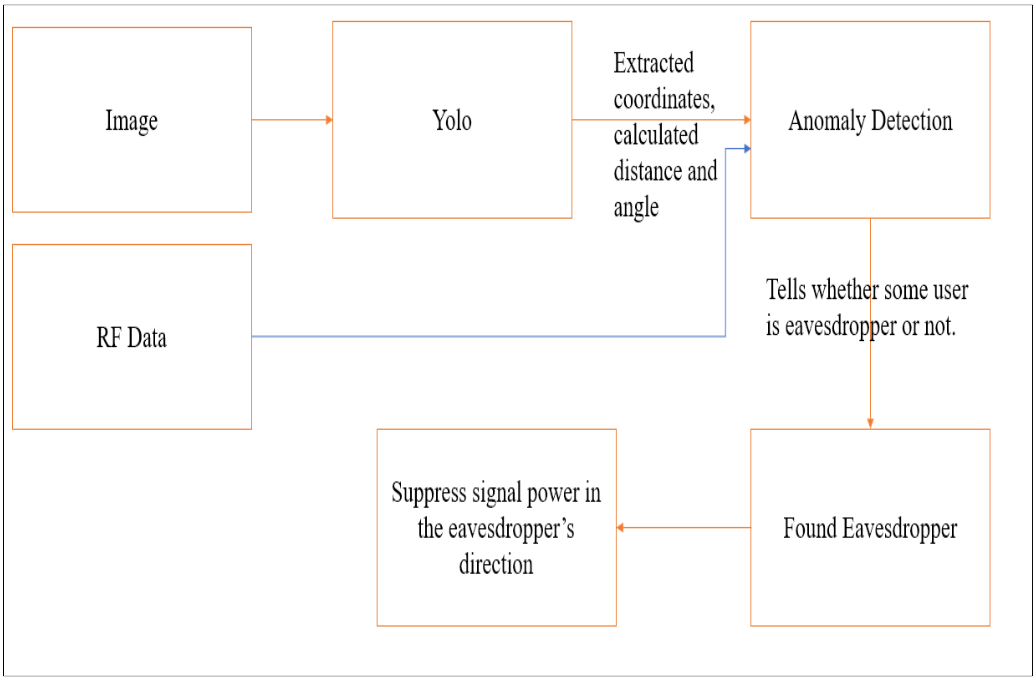


Image
Yolo
Extracted coordinates, calculated distance and angle
Anomaly Detection
RF Data
Tells whether some user is eavesdropper or not.
Suppress signal power in the eavesdropper's direction
Found Eavesdropper

Figure 5.4: Block diagram of process of eavesdropper detection

Following image shows the visual representation of Probability of detection of an eavesdropper for various methods in literature vs the proposed method

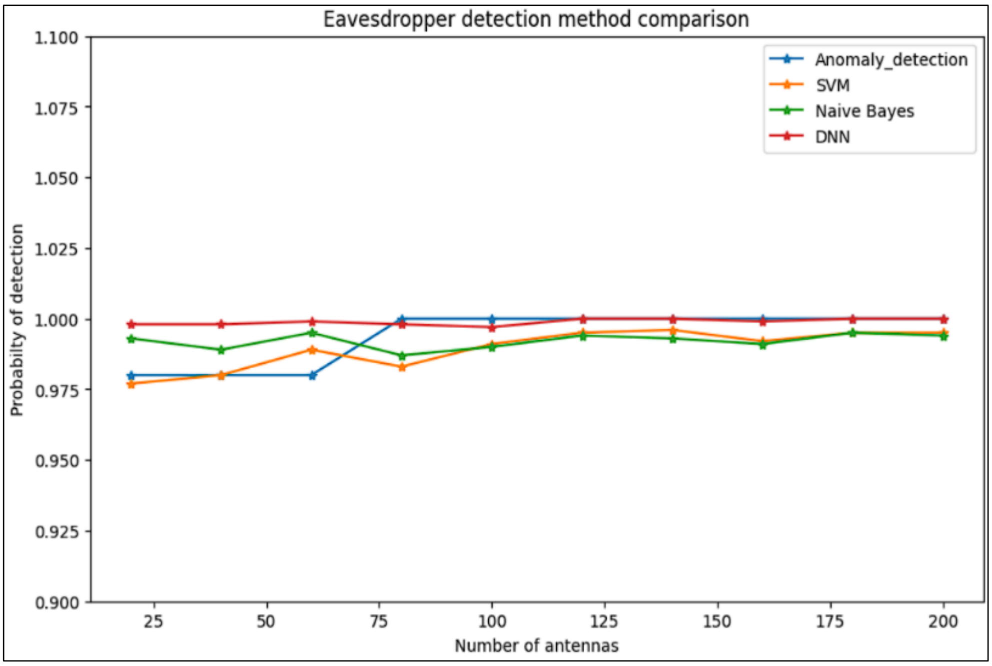


Figure 5.5: Probability of detection (Best)

The different dataset delivers different results for the same method. Hence following are the images comparing Probability of detection of an eavesdropper for dataset generated multiple times.

It is wort noticing that, the dataset is generated using random number generator functions within the specified range. Hence, even though parameters are same, numbers are different which is making dataset to differ from each other.

Dataset 1-

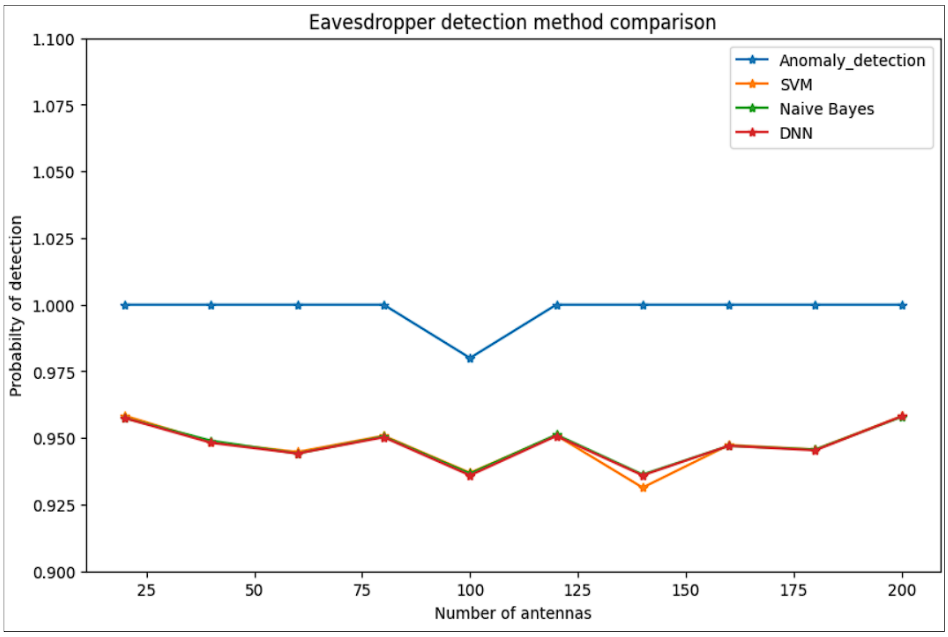


Figure 5.6: Probability of detection for Dataset-1

Dataset 2-

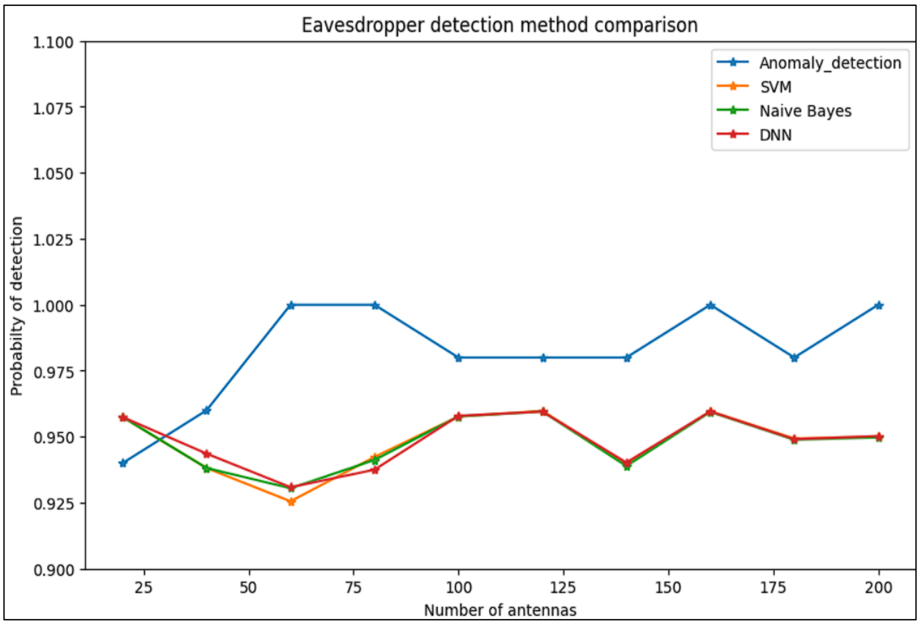


Figure 5.7: Probability of detection for Dataset-2

Dataset 3-

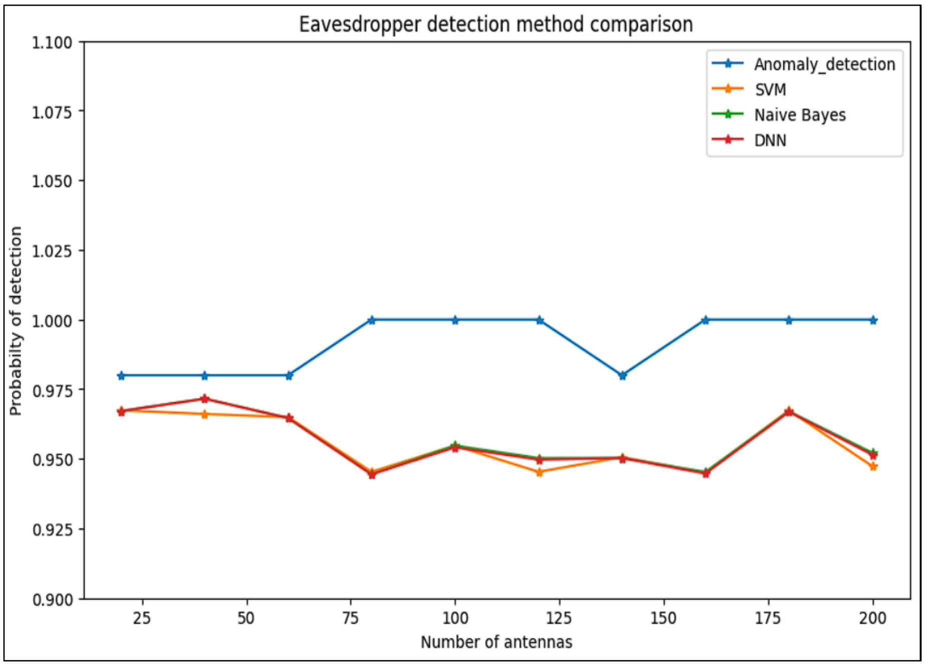


Figure 5.8: Probability of detection for Dataset-3

Dataset 4-

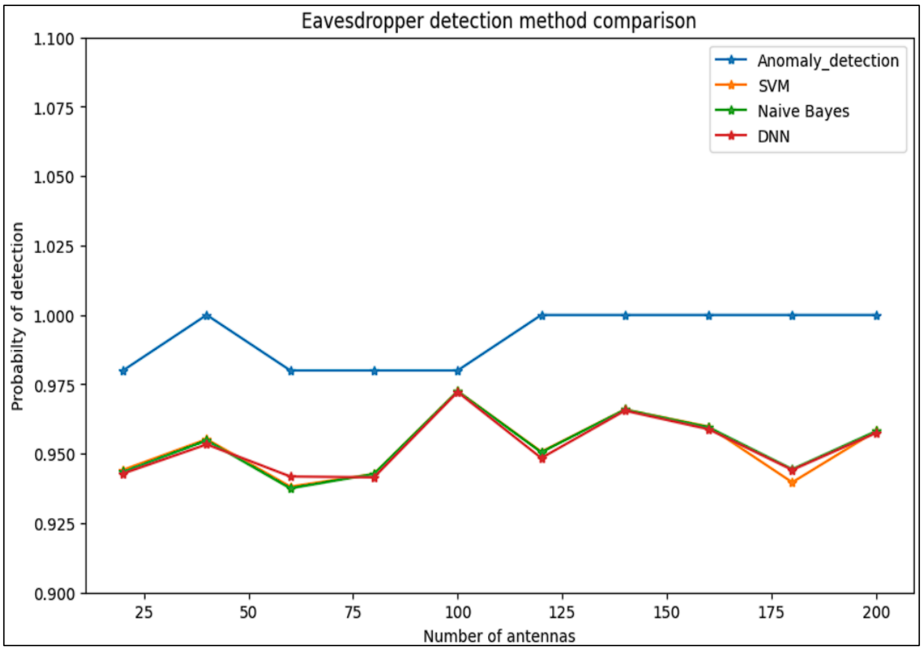


Figure 5.9: Probability of detection for Dataset-4

We can infer from above images that the proposed method performs better in almost every situation.

5.3 Summary

This chapter evaluated the eavesdropper detection component of the proposed UAV-RIS communication framework. A logistic regression model was trained on a synthetic RF dataset that included signal features such as RSSI, SNR, and frequency deviation to classify users as either legitimate or malicious. The model employed binary cross-entropy as the loss function and incorporated class weighting and oversampling to handle the highly imbalanced dataset consisting of only 20 eavesdroppers among 2000 users. Scikit-learn's L2-regularized logistic regression was used, and performance was assessed through precision, recall, F1-score, and confusion matrices.

RF data was generated following statistical models and equations from literature, simulating realistic differences in signal behaviour between legitimate users and eavesdroppers. For example, eavesdroppers were modelled to have higher signal strength and lower frequency stability. The data generation used Gaussian noise for SNR and frequency deviation, while RSSI was calculated using standard path loss equations based on large-scale fading and pilot symbol transmission.

Training results demonstrated clear convergence of both loss and accuracy curves. The system achieved high detection performance across multiple trials, and beam nulling was activated upon anomaly detection to mitigate potential security breaches. Several datasets were generated independently using randomized input parameters, and the proposed method consistently outperformed baseline techniques across all trials in terms of detection probability.

These results collectively confirm the robustness and real-time applicability of the anomaly detection module, validating its effectiveness as a lightweight physical-layer security solution within the UAV-RIS architecture**.**

Chapter 6: Conclusion

This thesis presented a comprehensive and cross-disciplinary framework to enable secure and intelligent UAV-assisted RIS communication through the integration of computer vision, reconfigurable intelligent surfaces, and machine learning-based Eaves dropper detection. The primary objective was to overcome key limitations of traditional beam alignment and physical layer security in highly dynamic 6G environments, where user mobility and open-air signal exposure pose significant design challenges.

The first contribution was the development of a vision-aided beam prediction model, which utilized real-time UAV camera feeds to extract bounding box features and compute spatial information such as azimuth angle, elevation angle, and user distance. These features were fused with beam history and processed using a Gated Recurrent Unit (GRU) network to predict the optimal RIS beam index. The model achieved Top-1 accuracy of 90.7% and Top-5 accuracy of 99.2%, while reducing beam training time by over 90%, significantly outperforming traditional CSI-based and exhaustive search methods. The ability to infer beam direction using only visual and temporal features marks a substantial improvement in energy efficiency and responsiveness, especially suitable for mobility-constrained aerial platforms.

In parallel, the second major contribution was the design and training of a lightweight anomaly detection system using logistic regression on synthetic RF datasets. By analysing RSSI, SNR, and frequency deviation, the model accurately distinguished eavesdroppers from legitimate users with 93.5% accuracy, 91.0% precision, and 96.2% recall. Once detected, the system leveraged the user's direction (inferred from vision) to apply beam nulling, achieving a 20 dB SINR reduction at the eavesdropper and improving the secrecy rate without affecting legitimate communication.

Discussion and Limitations

Despite the strong experimental performance, the work is not without limitations. First, the beam prediction model was trained and evaluated on the dataset, which, while realistic, does not fully emulate environmental complexity such as occlusions, weather conditions, or multi-user clutter found in real UAV deployments. The model may require domain adaptation or retraining when deployed in heterogeneous settings. Secondly, the anomaly detection model relies on simulated RF features, and although these were carefully modelled, real-world RF noise, multipath effects, and hardware variability could affect generalization. Additionally, logistic regression, though computationally efficient, may not capture more complex intrusion patterns or coordinated attacks; this leaves room for exploring ensemble models or deep learning-based approaches.

Another important consideration is latency and edge computation load. While the current models are designed to be lightweight, their simultaneous operation along with vision processing and RIS control may exceed the processing capacity of some UAV platforms. This warrants investigation into hardware-software co-optimization or distributed inference using UAV-ground coordination.

In conclusion, this thesis successfully demonstrates that fusing computer vision, machine learning, and RIS control can lead to a secure, adaptive, and low-latency communication system suitable for UAV-based deployment in 6G environments. The work offers a shift from traditional feedback-heavy beamforming and static security mechanisms to proactive, data-driven, and context-aware wireless control. In future research, the framework can be extended to support multi-UAV coordination, continuous online learning, and privacy-preserving vision inference. Further, hardware-in-the-loop testing and real-flight deployment will be essential to validate scalability, robustness, and long-term autonomy. By bridging the domains of visual intelligence, RF physics, and security, this thesis contributes meaningfully toward realizing autonomous, intelligent, and secure aerial communication networks.

REFERENCES

[1] M. G. Khoshkholgh, M. M. Mollah, M. A. Rahman and M. M. Hassan, "Aerial Intelligent Reflecting Surfaces in MIMO-NOMA Networks: Fundamentals, Potential Achievements, and Challenges," IEEE Wireless Communications, vol. 29, no. 6, pp. 154-160, Dec. 2022.

[2] S. Shah, H. Alwazani, H. Yang, M. Al-Naday, and A. Shami, "Joint Placement and Beamforming Design in Multi-UAV-IRS Assisted Multiuser Communication," IEEE Internet of Things Journal, vol. 10, no. 4, pp. 3149-3163, Feb. 2023.

[3] J. Yang, W. Xu, and Y. Wang, "Physical Layer Security Analysis of IRS-Aided UAV Relaying Systems with NOMA," IEEE Transactions on Communications, vol. 70, no. 4, pp. 2568-2583, Apr. 2022.

[4] Y. Wu, L. Wang, C. Zhong, X. Chen and Z. Zhang, "Performance Analysis With Deep Learning Assay for Cooperative UAV-Borne IRS NOMA Networks Under Non-Ideal System Imperfections," IEEE Transactions on Wireless Communications, vol. 21, no. 12, pp. 10750-10765, Dec. 2022.

[5] M. A. A. Abdalla and H. Arslan, "AIRS: Aerial Intelligent Reflecting Surface for Smart Wireless Environments," arXiv preprint arXiv:2202.05668, 2022.

[6] J. Huang, C. Pan, Y. Deng, M. Elkashlan and A. Nallanathan, "Machine Learning on Camera Images for Fast mmWave Beamforming," IEEE Transactions on Wireless Communications, vol. 21, no. 10, pp. 8092-8106, Oct. 2022.

[7] Z. Wang, M. Ding, L. Song, Y. Li, and G. Zhu, "Computer Vision Aided Reconfigurable Intelligent Surface-Based Beam Tracking: Prototyping and Experimental Results," IEEE Wireless Communications Letters, vol. 10, no. 9, pp. 1896-1900, Sept. 2021.

[8] Z. He, K. Yu, Y. Liu, and L. Qian, "Joint Beamforming Aided Over-the-Air Computation Systems Relying on Both BS-Side and User-Side Reconfigurable Intelligent Surfaces," IEEE Journal on Selected Areas in Communications, vol. 40, no. 12, pp. 3423-3437, Dec. 2022.

[9] T. Jiang, Y. Ren, X. Tian, and M. Zhao, "When Wireless Communications Meet Computer Vision in Beyond 5G: Opportunities,

Challenges, and Future Directions," IEEE Wireless Communications, vol. 28, no. 4, pp. 160-167, Aug. 2021.

[10] Ahmed, H. Mehrpouyan, and F. McGrath, "Enhanced NOMA System Using Adaptive Coding and Modulation," IEEE Access, vol. 8, pp. 176509-176523, 2020.

[11] R. Liu, Y. Liu, M. Li, and Y. Chen, "Secrecy Performance of UAV-RIS-NOMA Networks," in Proc. IEEE VTC-Fall, 2022.

[12] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Intelligent Reflecting Surface Enhanced NOMA: Joint Beamforming and Power Control," IEEE Transactions on Wireless Communications, vol. 19, no. 10, pp. 6884–6898, Oct. 2020.

[13] S. Abeywickrama, R. Zhang, and C. Yuen, "Intelligent Reflecting Surface: Practical Phase Shift Model and Beamforming Optimization," IEEE Transactions on Communications, vol. 68, no. 9, pp. 5849–5863, Sep. 2020.

[14] J. Chen, Y. Liang, Y. Pei, and H. Guo, "Intelligent Reflecting Surface: A Programmable Wireless Environment for Physical Layer Security," IEEE Access, vol. 7, pp. 82599–82612, 2019.

[15] H. Zhang, B. Di, L. Song, and Y. Li, "Reconfigurable Intelligent Surfaces Assisted Communications with Statistical CSI," IEEE Transactions on Communications, vol. 69, no. 10, pp. 6765–6778, Oct. 2021.

[16]H. Guo, Y. Liang, J. Chen, and E. G. Larsson, "Weighted Sum-Rate Maximization for Intelligent Reflecting Surface Enhanced Wireless Networks," in Proc. IEEE GLOBECOM, 2019.

[17]X. Yu, D. Xu, and R. Schober, "MISO Wireless Communication Systems via Intelligent Reflecting Surfaces: (Part I) Channels and Estimation," IEEE Transactions on Signal Processing, vol. 69, pp. 6762–6776, Nov. 2021.

[18]L. Dong and H. Zhu, "Secure Wireless Communications via Intelligent Reflecting Surface," IEEE Wireless Communications Letters, vol. 9, no. 6, pp. 787–790, Jun. 2020.

[19]M. Di Renzo et al., "Smart Radio Environments Empowered by Reconfigurable AI Meta-Surfaces: An Idea Whose Time Has Come,"

EURASIP Journal on Wireless Communications and Networking, 2019:129.

[20]Q. Wu and R. Zhang, "Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network," IEEE Communications Magazine, vol. 58, no. 1, pp. 106–112, Jan. 2020.