A

Major Project

On

MEMBERSHIP INFERENCE ATTACK AND DEFENCE FOR WIRELESS SIGNAL CLASSIFIERS WITH DEEP LEARNING

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

N. Vaishnavi	(217R1A05P4)
G. Jaswanth Kumar	(217R1A05M9)
B. Meenakshi	(217R1A05L5)

Under the Guidance of

Dr. V. Naresh Kumar

Associate Professor - HOD CSE - II

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi)

Recognized Under Section 2(f) & 12(B) of the UGCAct.1956,

Kandlakoya (V), Medchal Road, Hyderabad-501401.

April, 2025.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR
TECHNICAL CAMPUS
EXPLORE TO INVENT

CERTIFICATE

This is to certify that the project entitled “**MEMBERSHIP INFERENCE ATTACK AND DEFENCE FOR WIRELESS SIGNAL CLASSIFIERS WITH DEEP LEARNING**” being submitted by **N. Vaishnavi (217R1A05P4), G. Jaswanth Kumar (217R1A05M9) AND B. Meenakshi (217R1A05L5)** in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering to the Jawaharlal Nehru Technological University Hyderabad, during the year 2024-25.

The results embodied in this project have not been submitted to any other University or Institute for the award of any degree or diploma.

Dr. V. Naresh Kumar
Associate Professor - HOD CSE - II
INTERNAL GUIDE

Dr. Nuthanakanti Bhaskar
HOD

Dr. A. Raji Reddy
DIRECTOR

Signature of External Examiner

Submitted for viva voice Examination held on ______________________

ACKNOWLEDGEMENT

We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project, we also take this opportunity to express our profound gratitude and deep regard to our guide **Dr. V. Naresh Kumar,** Associate Professor(HOD CSE - II) for his exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by him shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to extend our heartfelt appreciation to the Project Review Committee (PRC) Coordinators—**Dr. K. Maheswari, Dr. J. Narasimharao, Ms. K. Shilpa, and Mr. K. Ranjith Reddy**—for their unwavering support, insightful guidance, and valuable inputs, which played a crucial role in steering this project through its various stages.

Our sincere appreciation also goes to **Dr. Nuthanakanti Bhaskar**, Head of the department, for his encouragement and continuous support in ensuring the successful completion of our project.

We are deeply grateful to **Dr. A. Raji Reddy**, Director, for his cooperation throughout the course of this project. Additionally, we extend our profound gratitude to Sri. **Ch. Gopal Reddy**, Chairman, Smt. **C. Vasantha Latha**, Secretary and Sri. **C. Abhinav Reddy**, Vice-Chairman, for fostering an excellent infrastructure and a conducive learning environment that greatly contributed to our progress.

We also acknowledge and appreciate the guidance and assistance provided by the faculty and staff of **CMR Technical Campus**, whose contributions have been invaluable in bringing this project to fruition.

Lastly, we sincerely thank our families for their unwavering support and encouragement. We also extend our gratitude to the teaching and non-teaching staff of CMR Technical Campus for their guidance and assistance. Their contributions, along with the support of everyone who helped directly or indirectly, have been invaluable in the successful completion of this project.

N. Vaishnavi (217R1A05P4)
G. Jaswanth Kumar (217R1A05M9)
B. Meenakshi (217R1A05L5)

VISION AND MISSION

INSTITUTE VISION:

To Impart quality education in serene atmosphere thus strive for excellence in Technology and Research.

INSTITUTE MISSION:

1. To create state of art facilities for effective Teaching- Learning Process.

2. Pursue and Disseminate Knowledge based research to meet the needs of Industry & Society.

3. Infuse Professional, Ethical and Societal values among Learning Community.

DEPARTMENT VISION:

To provide quality education and a conducive learning environment in computer engineering that foster critical thinking, creativity, and practical problem-solving skills.

DEPARTMENT MISSION:

1. To educate the students in fundamental principles of computing and induce the skills needed to solve practical problems.

2. To provide State-of-the-art computing laboratory facilities to promote industry institute interaction to enhance student’s practical knowledge.

3. To inculcate self-learning abilities, team spirit, and professional ethics among the students to serve society.

ABSTRACT

This project is titled as “Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning”. An over-the-air membership inference attack (MIA) is presented to leak private information from a wireless signal classifier. Machine learning (ML) provides powerful means to classify wireless signals, e.g., for PHY-layer authentication. As an adversarial machine learning attack, the MIA infers whether a signal of interest has been used in the training data of a target classifier. This private information incorporates waveform, channel, and device characteristics, and if leaked, can be exploited by an adversary to identify vulnerabilities of the underlying ML model (e.g., to infiltrate the PHY-layer authentication). One challenge for the over-the-air MIA is that the received signals and consequently the RF fingerprints at the adversary and the intended receiver differ due to the discrepancy in channel conditions.

Therefore, the adversary first builds a surrogate classifier by observing the spectrum and then launches the black box MIA on this classifier. The MIA results show that the adversary can reliably infer signals (and potentially the radio and channel information) used to build the target classifier. Therefore, a proactive defence is developed against the MIA by building a shadow MIA model and fooling the adversary. This defence can successfully reduce the MIA accuracy and prevent information leakage from the wireless signal classifier.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

1. INTRODUCTION

1. INTRODUCTION

The project, titled " Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning " With the rapid advancement of machine learning (ML) and deep learning (DL) in wireless communications, these technologies have become integral to various applications, including spectrum sensing, signal classification, and PHY-layer authentication. However, the integration of ML-based classifiers in wireless networks introduces new security and privacy risks. One such threat is the Membership Inference Attack (MIA), where an adversary attempts to determine whether specific data samples were used in training a target ML model. This can lead to privacy breaches by leaking sensitive information about waveform characteristics, radio devices, and channel conditions.

In this project, we explore the vulnerabilities of wireless signal classifiers to MIAs and propose a proactive defence mechanism. The adversary first builds a surrogate classifier using observed spectrum data and then launches a black-box MIA to infer training data membership. Our results demonstrate that the MIA can achieve high accuracy, posing significant privacy risks to wireless networks. This study highlights the importance of securing ML-based wireless systems against adversarial threats while maintaining classification accuracy.

1.1 PROJECT PURPOSE

The primary purpose of the project "Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning" is to enhance the security and privacy of deep learning-based wireless signal classifiers, which are increasingly being deployed in critical applications such as communication networks, IoT devices, and wireless surveillance systems. These classifiers, while powerful, are vulnerable to membership inference attacks (MIAs), where an adversary can deduce whether a specific data point (e.g., a signal or communication) was part of the model’s training dataset. This can lead to serious privacy risks, particularly in sensitive environments where the details of the data used for training must remain confidential.

1.2 PROJECT FEATURES

This project incorporates several key features to improve the accuracy and efficiency of Attack Prediction and classification.

Membership Inference Attack (MIA): Implements an over-the-air MIA to determine if a wireless signal was used in training a classifier, posing a privacy risk.

Two MIA Attack Settings: Identifies membership using signals from the same radio device.

Differentiates between signals from different radio devices.

Deep Learning-Based Wireless Classification: Uses a Deep Neural Network (DNN) to classify wireless signals, supporting PHY-layer authentication and authorized user detection.

Implementation and Evaluation: Built using Python with TensorFlow for deep learning modelling. Uses Django-ORM for backend operations and MySQL for data storage. Provides web-based monitoring and analysis using HTML, CSS, and JavaScript.

This project demonstrates the privacy risks of Membership Inference Attacks (MIA) on wireless signal classifiers. It shows how adversaries can exploit deep learning models to infer training data and leak sensitive information. To counter this, a proactive defence using a shadow MIA model is implemented, effectively reducing attack accuracy. The results highlight the need for robust security measures to protect ML-based wireless systems from adversarial threats.

2. LITERATURE SURVEY

2.LITERATURE SURVEY

Machine learning (ML) and deep learning (DL) have revolutionized wireless communication systems by enabling efficient signal classification, spectrum sensing, and PHY-layer authentication. These advancements have significantly improved wireless network security and performance by allowing automated detection of unauthorized signals, optimizing resource allocation, and enhancing modulation recognition. However, the integration of ML in wireless networks has introduced new security and privacy challenges, particularly from adversarial attacks that exploit vulnerabilities in deep learning models. One such attack is the **Membership Inference Attack (MIA)**, which threatens the privacy of ML-based classifiers by determining whether a specific signal was used in the training dataset. This attack can lead to sensitive information leakage about waveform characteristics, radio devices, and channel environments, making wireless networks susceptible to further security threats. This literature survey explores previous research on ML-based wireless systems, adversarial attacks, MIAs, and existing defense mechanisms to highlight the significance of securing wireless classifiers against privacy breaches.

Machine Learning in Wireless Communications

The adoption of machine learning in wireless communication has enabled intelligent decision-making, real-time signal processing, and enhanced security mechanisms. DL models, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have been widely employed for tasks such as **modulation classification, RF fingerprinting, and spectrum sensing**. These ML-based classifiers are used for **PHY-layer authentication**, allowing networks to distinguish between authorized and unauthorized devices based on signal characteristics. Research has shown that ML-driven classifiers can achieve high accuracy in identifying modulation types and detecting adversarial interference. However, despite their effectiveness, these classifiers are vulnerable to adversarial machine learning attacks, which exploit model weaknesses to manipulate classification results.

One major security challenge in ML-based wireless systems is the overfitting of models to training data, making them susceptible to inference attacks.

Overfitting occurs when a model learns specific patterns from the training data rather than generalizing well to unseen data. This vulnerability can be exploited by an adversary to launch MIAs, where the attacker attempts to determine whether a given signal was part of the training set. Such attacks pose serious **privacy risks**, as they can reveal information about the transmitter, receiver, and channel conditions used during model training.

Membership Inference Attacks (MIA)

Membership Inference Attacks have been extensively studied in various domains, including computer vision, healthcare, and natural language processing (NLP). The primary goal of an MIA is to infer whether a particular data sample was used to train a target ML model. This attack is particularly dangerous in scenarios where training data includes private or sensitive information, such as medical records, financial transactions, or biometric data.

In the context of wireless networks, MIAs represent a novel attack vector that has not been widely explored. Unlike traditional data domains where the adversary can directly query the ML model, wireless MIAs involve over-the-air attacks, where the adversary passively listens to transmitted signals and attempts to infer training membership. This attack can be performed using a black-box approach, where the attacker has no direct access to the model, or a white-box approach, where the attacker has partial knowledge of the model architecture.

Prior studies on MIAs in other fields have demonstrated that these attacks are highly effective when models are overfitted or poorly regularized. Research has shown that MIAs can achieve high attack success rates by exploiting differences in model confidence scores for training and non-training samples. This concept is applicable to wireless networks, where an adversary can build a surrogate classifier to replicate the behavior of the target classifier and launch MIAs to infer signal membership.

Adversarial Machine Learning in Wireless Networks

Wireless networks face various **adversarial machine learning (AML) threats**, including **evasion attacks, data poisoning, model inversion, and spoofing attacks**. These attacks leverage ML vulnerabilities to manipulate model predictions, degrade system performance, or extract private information.

1. Evasion Attacks: In evasion attacks, adversaries introduce adversarial perturbations into signals to fool ML-based classifiers. For example, adversarial noise can be added to a signal to mislead modulation classifiers, causing incorrect predictions. Previous studies have shown that channel-aware adversarial attacks can significantly reduce the accuracy of DL-based modulation classifiers.

2. Poisoning Attacks: These attacks involve injecting malicious samples into the training dataset, causing the model to learn incorrect patterns. Poisoning attacks can degrade the performance of wireless classifiers by introducing biased or misleading training samples.

3. Model Inversion Attacks: In model inversion attacks, an adversary attempts to reconstruct training data from the model's output. This attack is particularly concerning in wireless networks, where private information about waveforms, radio devices, and channel conditions can be extracted.

4. **Spoofing Attacks:** In spoofing attacks, adversaries generate signals that mimic those of authorized users to gain unauthorized access. MIAs can be used as a preliminary step to gather information about legitimate users before launching spoofing attacks.

Recent research has highlighted that ML-based wireless classifiers are vulnerable to multiple adversarial threats, including MIAs. Given that wireless channels are dynamic and unpredictable, defending against these attacks requires robust security mechanisms that account for signal variations, noise, and adversarial manipulations.

Defensive Strategies Against Membership Inference Attacks

Several defence mechanisms have been proposed to protect ML models from MIAs, including differential privacy, adversarial regularization, and randomized smoothing. These techniques aim to reduce overfitting, obscure model confidence scores, or introduce noise to prevent information leakage.

1. Differential Privacy: Differential privacy introduces controlled noise into training data or model outputs to prevent an adversary from distinguishing between training and non-training samples. This method has been widely used to protect sensitive datasets in fields such as healthcare and finance.

2. **Adversarial Regularization:** Regularization techniques, such as dropout, weight decay, and adversarial training, help prevent overfitting and improve model generalization. Studies have shown that regularized models are less susceptible to MIAs.

3. **Randomized Smoothing:** This approach adds small random noise to model predictions, making it difficult for an adversary to exploit confidence score differences. Randomized smoothing has been successfully applied in computer vision and NLP domains to defend against inference attacks.

3. **Shadow MIA Defense:** One effective approach is to build a shadow MIA model that mimics the adversary's attack strategy. This model is used to identify and mitigate attack vulnerabilities before deploying the classifier in a real-world wireless environment. The defense mechanism involves introducing controlled perturbations in classification scores to mislead the adversary without affecting the classifier's overall accuracy.

The integration of ML in wireless networks offers numerous benefits but also introduces new security and privacy risks. Membership Inference Attacks (MIAs) pose a significant threat to wireless signal classifiers by enabling adversaries to infer training data membership, leading to potential privacy breaches. Previous research on adversarial machine learning has demonstrated that ML models are vulnerable to evasion, poisoning, and inference attacks, highlighting the need for robust defensive mechanisms.

To address these challenges, existing defensive strategies such as differential privacy, adversarial regularization, and randomized smoothing have been explored. However, in the wireless domain, where adversaries rely on over-the-air attacks, a more proactive defense mechanism is required. The use of a shadow MIA model provides an effective solution by simulating attack conditions and introducing perturbations that reduce MIA success rates.

This literature survey underscores the importance of securing ML-based wireless systems against MIAs and other adversarial threats. Future research should focus on enhancing model robustness, developing real-time attack detection systems, and integrating ML security measures to ensure the privacy and security of wireless communication networks

2.1 REVIEW OF RELATED WORK

The study of Membership Inference Attacks (MIA) and defenses in machine learning (ML) has gained significant attention, particularly in the fields of computer vision, healthcare, and data privacy. However, their application in wireless communication systems remains relatively unexplored. Wireless signal classifiers, which utilize deep learning models to identify and authenticate signals, are vulnerable to inference attacks that can leak sensitive information about training data. This review discusses previous research on ML-based wireless signal classification, adversarial machine learning, and privacy threats posed by MIAs, highlighting their strengths and limitations.

1. <u>Traditional Wireless Security Approaches</u>

Early security measures in wireless networks primarily focused on cryptographic techniques and signal authentication mechanisms. Encryption protocols such as AES (Advanced Encryption Standard) and RSA (Rivest-Shamir-Adleman) were widely used to ensure secure communication. Additionally, PHY-layer authentication was introduced to verify the legitimacy of signals by analysing unique radio frequency (RF) fingerprints.

While these traditional approaches provided strong security, they were computationally expensive and often infeasible for resource-constrained IoT and 5G devices. Furthermore, encryption alone could not prevent inference attacks, where adversaries extract sensitive information from ML-based classifiers.

2. <u>Machine Learning-Based Approaches</u>

With advancements in ML, researchers explored feature-based classifiers such as Support Vector Machines (SVM), Random Forest, and k-Nearest Neighbours (KNN) for modulation classification and signal identification. These models relied on handcrafted features like spectral characteristics, waveform properties, and statistical features.

Despite their effectiveness, traditional ML models suffered from poor generalization when applied to dynamic wireless environments with varying channel conditions. Additionally, manually extracted features were often insufficient to capture complex signal variations.

3. Deep Learning-Based Approaches

Recent advancements in deep learning have significantly improved wireless signal classification. Convolutional Neural Networks (CNNs) have been widely used for modulation classification, leveraging their ability to extract high-level spatial features from in-phase (I) and quadrature (Q) signal components. Additionally, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have been employed to capture temporal dependencies in wireless signals.

One of the most effective architectures for wireless signal classification is the CNN-LSTM hybrid, where CNNs extract spatial features, and LSTMs analyse temporal relationships. Studies have demonstrated the effectiveness of pre-trained CNN models like VGG-16, Rest Net, and Efficient Net for feature extraction, followed by LSTM layers for sequence modelling. Despite their success, these models often suffer from overfitting, making them susceptible to membership inference attacks that exploit learned data distributions.

4. Membership Inference Attacks in Machine Learning

Membership Inference Attacks (MIA) have been extensively studied in image classification, medical data analysis, and financial security, where attackers aim to determine whether a specific data sample was used to train an ML model. Previous studies have shown that overfitted models are particularly susceptible to MIAs, as they memorize training data characteristics. Research has demonstrated that MIAs can be executed using confidence scores, entropy values, or query-based access to the model’s output.

Traditional MIAs assume that an attacker can query a target model and observe its output. However, in wireless communication, an adversary must rely on over-the-air signal observations, making MIAs more challenging due to channel variations, and interference.

5. Recent Advances: Over-the-Air Membership Inference Attacks (MIA) in Wireless Systems

While previous studies have focused on MIAs in static datasets and structured data domains, wireless networks present unique challenges due to dynamic channel conditions and real-time signal variations.

The concept of membership inference in wireless classifiers has not been extensively explored. This study introduces the first over-the-air MIA, demonstrating how an adversary can infer training data membership despite variations in received signals.This review highlights the evolution of content moderation techniques, emphasizing the shift from rule-based and machine learning approaches to deep learning-based models. The proposed solution for inappropriate content detection in YouTube videos.

6. Comparison with the Proposed Approach

While existing research has explored various adversarial threats to wireless classifiers, MIAs specifically targeting training data privacy in real-world wireless environments have not been extensively studied. The proposed approach improves upon previous work by:

Introducing the first over-the-air MIA for wireless signal classifiers.

Developing a surrogate classifier to mimic the target classifier's behaviour based on observed spectrum data.

Evaluating MIA performance in different attack scenarios, including variations in signal-to-noise ratio (SNR) and channel effects.

Implementing a defence mechanism using a shadow MIA model, which reduces attack success by introducing controlled perturbations in classification outputs.

2.2 DEFINITION OF PROBLEM STATEMENT

The growing use of deep learning models in wireless signal classification has raised significant privacy concerns, particularly with the vulnerability of these models to membership inference attacks (MIAs). In a membership inference attack, an adversary can determine whether a specific data point was part of the model's training dataset, which poses serious risks to privacy, especially in sensitive applications such as communication networks, healthcare, and IoT systems. This project addresses the problem of safeguarding deep learning-based wireless signal classifiers against such attacks, exploring the vulnerabilities inherent in these models while developing and evaluating defence mechanisms that protect sensitive training data without compromising classifier performance. The goal is to provide a secure framework that ensures privacy in wireless systems while maintaining high model accuracy.

2.3 EXISTING SYSTEM

An existing system presents channel-aware adversarial attacks against deep learning-based wireless signal classifiers. There is a transmitter that transmits signals with different modulation types. A deep neural network is used at each receiver to classify its over-the-air received signals to modulation types. In the meantime, an adversary transmits an adversarial perturbation (subject to a power budget) to fool receivers into making errors in classifying signals that are received as superpositions of transmitted signals and adversarial perturbations. First, these evasion attacks are shown to fail when channels are not considered in designing adversarial perturbations. Then, realistic attacks are presented by considering channel effects from the adversary to each receiver. After showing that a channel-aware attack is selective (i.e., it affects only the receiver whose channel is considered in the perturbation design), a broadcast adversarial attack is presented by crafting a common adversarial perturbation to simultaneously fool classifiers at different receivers. The major vulnerability of modulation classifiers to over-the-air adversarial attacks is shown by accounting for different levels of information available about the channel, the transmitter input, and the classifier model. Finally, a certified defence based on randomized smoothing that augments training data with noise is introduced to make the modulation classifier robust to adversarial perturbations.

Limitations of Existing System

1. **Vulnerability to Adversarial Attacks:** The wireless signal classifier can be easily fooled by carefully crafted adversarial signals, leading to misclassification of signals.
2. **Potential for Attack Adaptation:** Attackers can adapt their strategies by learning how the defense works, making the system continuously vulnerable to evolving attacks.
3. **Dependency on Channel Knowledge for Attacks**: If the attacker does not have accurate knowledge of the wireless channel conditions, their attack fails**.**
4. **Limited Attack Scope Without Channel Awareness:** This means attackers must put extra effort into understanding the environment before launching a successful attack.
5. **Increased Complexity in Defense Mechanisms:** The proposed randomized smoothing defense makes the classifier more robust, but it adds computational overhead to the system.

2.4 PROPOSED SYSTEM

In this project, we present the first MIA that is launched against a wireless classifier over the air to infer about training data and leak private information on waveform, device, and channel characteristics. We consider two settings for the MIA: (i) the MIA should be able to identify signals from the same radio device as member and non-member, and (ii) non member signals are generated by different radio devices. We extend the MIA such that it is launched by using not only received signals but also their noisy variations by accounting for channel variations. We show through detailed numerical results that the success of the MIA is high, i.e., the MIA can infer the training data membership of the wireless signal classifier with high accuracy. We present a defence scheme to protect wireless signal classifiers from the MIA and show that this defence can reduce the accuracy of the MIA significantly.

Advantages of the Proposed System:

The proposed system significantly improves upon the existing approaches by addressing key limitations:

1. Identifies Privacy Risks in Wireless Signal Classification: This project highlights a new privacy threat in deep learning-based wireless classifiers. It shows that an attacker can steal private information about waveforms, devices, and channel conditions.

2. Simulates Real-World Attack Scenarios: The project tests the MIA attack in two realistic scenarios: Attacking signals from the same device. Attacking signals from different devices. These tests provide a comprehensive understanding of how attackers can exploit wireless systems.

3. Improves Attack Accuracy with Noisy Variations: Unlike traditional attacks, this method considers channel variations and noisy signals, making it more practical. This approach better reflects real-world conditions, improving the accuracy and reliability of the attack analysis.

4. Provides a Defence Mechanism Against MIA: A proactive defence is introduced to protect wireless classifiers from MIA attacks. The defence method adds controlled noise to classification outputs, making it harder for the attacker to steal information**.**

5. Enhances Security in Wireless Communication Systems: This research is important for applications like 5G, IoT, and network authentication, where security is critical. By identifying and mitigating threats, this project helps make wireless networks more robust against cyberattacks.

6. Contributes to the Field of Adversarial Machine Learning: This project extends adversarial machine learning to the wireless domain, which is still a growing research area. The findings can help in future security research for deep learning-based wireless systems.

2.5 OBJECTIVES

1. 3. Develop an Over-the-Air Membership Inference Attack (MIA): Implement the first over-the-air MIA to analyse privacy vulnerabilities in wireless signal classifiers.
Demonstrate how an adversary can infer whether a signal was used in training a machine learning (ML) model.

2. Investigate MIA in Different Wireless Scenarios: Examine two attack settings:
(i) Identifying signals from the same device as members or non-members.
(ii) Differentiating between member and non-member signals from different devices.
Analyze the impact of channel variations and noise on attack effectiveness.

3. Evaluate Attack Success and Impact on Privacy: Conduct experiments to measure MIA accuracy across different Signal-to-Noise Ratios (SNRs).
Assess the risk of private data leakage, including waveform, device, and channel characteristics.

4. Optimize Machine Learning-Based Wireless Security: Ensure the defense mechanism preserves classification accuracy while preventing inference attacks.

5. Contribute to Wireless Security Research: Bridge the gap between adversarial machine learning and wireless security. Offer insights into future research directions for enhancing privacy in ML-based wireless communication systems.

2.6 HARDWARE & SOFTWARE REQUIREMENTS

2.6.1 HARDWARE REQUIREMENTS:

Hardware interfaces specifies the logical characteristics of each interface between the software product and the hardware components of the system. The following are some hardware requirements,

• Processor	:	Pentium IV or higher
• Hard disk	:	20GB or above
• RAM	:	8GB or above.

2.6.2 SOFTWARE REQUIREMENTS:

Software Requirements specifies the logical characteristics of each interface and software components of the system. The following are some software requirements,

• Operating system	:	Windows 7 or above.
• Language	:	Python
• Back-End	:	Django-ORM
• Frame Work	:	Tkinter

3. SYSTEM ARCHITECTURE AND DESIGN

3. SYSTEM ARCHITECTURE AND DESIGN

Project architecture refers to the structural framework and design of a project, encompassing its components, interactions, and overall organization. It provides a clear blueprint for development, ensuring efficiency, scalability, and alignment with project goals. Effective architecture guides the project's lifecycle, from planning to execution, enhancing collaboration and reducing complexity.

3.1 PROJECT ARCHITECTURE

The system architecture for Membership Inference Attack Prediction consists of four key components: the Web Server, Web Database, Service Provider, and Remote User. The Web Server processes user queries, stores dataset results, and interacts with the Web Database for data access. The Service Provider handles dataset training, accuracy analysis, attack prediction, and user management. Remote Users can register, log in, predict attack types, and view their profiles. This structured approach ensures efficient data processing, secure storage, and accurate attack prediction.

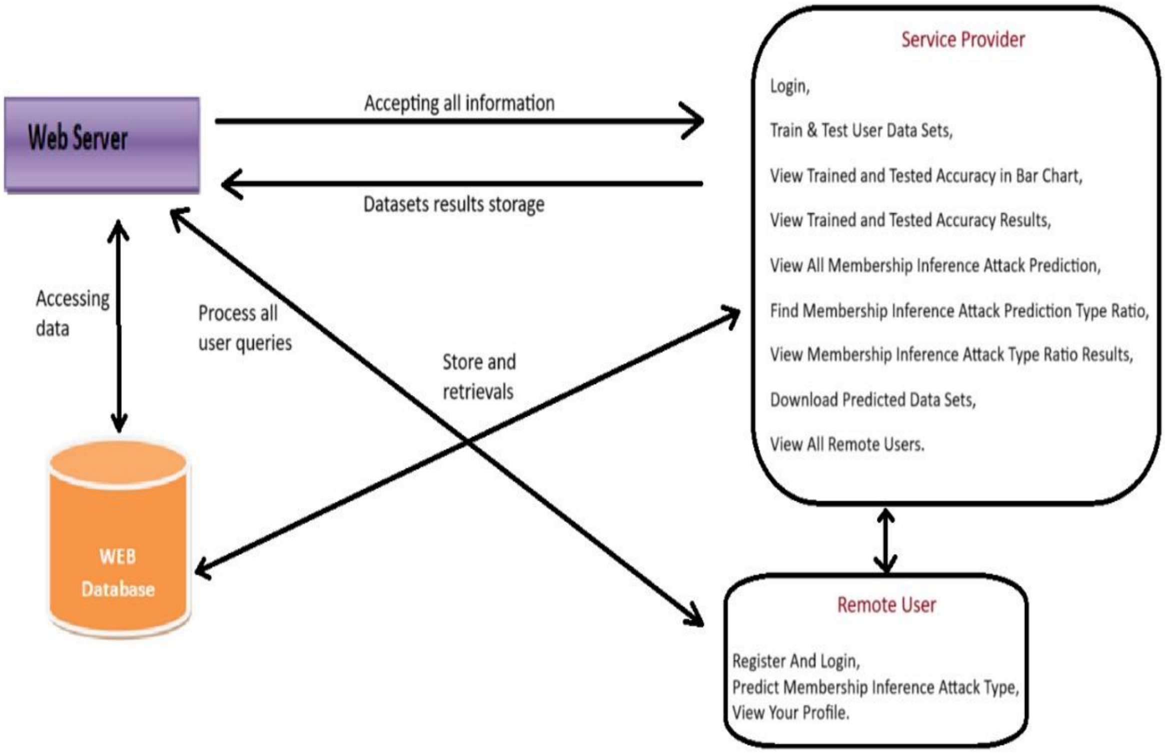

Web Server
Accepting all information
Datasets results storage
Accessing data
Process all user queries
Store and retrievals
WEB Database
Service Provider
Login,
Train & Test User Data Sets,
View Trained and Tested Accuracy in Bar Chart,
View Trained and Tested Accuracy Results,
View All Membership Inference Attack Prediction,
Find Membership Inference Attack Prediction Type Ratio,
View Membership Inference Attack Type Ratio Results,
Download Predicted Data Sets,
View All Remote Users.
Remote User
Register And Login,
Predict Membership Inference Attack Type,
View Your Profile.

Figure 3.1: Project Architecture of Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning

3.2 DESCRIPTION

Input Data : The project uses **wireless signal data** collected from various radio devices for training a deep learning-based wireless signal classifier. The dataset includes signals with different waveform, device, and channel characteristics.

Reading Data: The collected wireless signals are processed using signal processing techniques and DL frameworks to extract meaningful information for classification.

Feature Extraction : A Deep Neural Network (DNN) extracts high-dimensional features from the wireless signal data, capturing key characteristics such as modulation type.

Membership Inference Attack (MIA) Simulation: The extracted signal features are used to simulate a Membership Inference Attack (MIA), where an adversary attempts to determine if a particular signal was part of the training dataset.

Adversary Model Training: The adversary builds a surrogate model by analyzing observed wireless signals.

Defence Mechanism: To counteract MIAs, a defensive model is introduced using a shadow MIA approach, which helps fool the adversary by adding noise to the signal data thereby reducing attack accuracy.

Evaluation and Results: The system is tested against various MIA scenarios, demonstrating: High success rate of MIA in unprotected models. Significant reduction in attack accuracy when defence mechanisms are applied, proving the effectiveness of the proposed solution.

Feedback : The system continuously refines its defence mechanisms based on attack performance metrics and real-time user feedback, ensuring improved security against evolving adversarial threats.

3.3 DATA FLOW DIAGRAM

A Data Flow Diagram visually represents the flow of data within the Membership Inference Attack (MIA) and Defence System for wireless signal classifiers. Based on the project’s introduction, the DFD will illustrate how external signals interact with the system, how an adversary launches an attack, and how the system applies defence mechanisms.

A Data Flow Diagram comprises Four primary elements:

- External Entities: Observes wireless signals and launches MIA. Runs a deep learning classifier for authentication.
- Processes: Users transmit RF signals, which the service provider receives and processes. The received signals are preprocessed to extract features.
- Data Flows: RF signals are transmitted between authorized users and the service provider.
- Data Stores: Stores labeled wireless signals used for training.

These components are represented using standardized symbols, such as circles for processes, arrows for data flows, rectangles for external entities, and open-ended rectangles for data stores.

Benefits:

- Provides a structured understanding of attack and defense mechanisms.
- Helps identify vulnerabilities and improve system security.
- Ensures efficient data handling and classification performance.

Applications:

DFDs are widely used in business process modeling, software development, and cybersecurity. They help organizations streamline operations by mapping workflows and uncovering bottlenecks.

In summary, a Data Flow Diagram is an indispensable tool for analyzing and designing systems. Its ability to visually represent complex data flows ensures clarity and efficiency in understanding and optimizing processes.

Levels of DFD:

DFDs are structured hierarchically:

- Level 0 (Context Diagram): Shows interactions between the adversary, service provider, and authorized users, along with the flow of data in the attack-defense scenario.
- Level 1: Expands Level 0 by detailing key processes, such as signal processing, classification, MIA execution, and defense mechanisms.
- Level 2+: Provides further breakdown of each attack and defense mechanism, including shadow model training, perturbation application, and evaluation metrics computation.

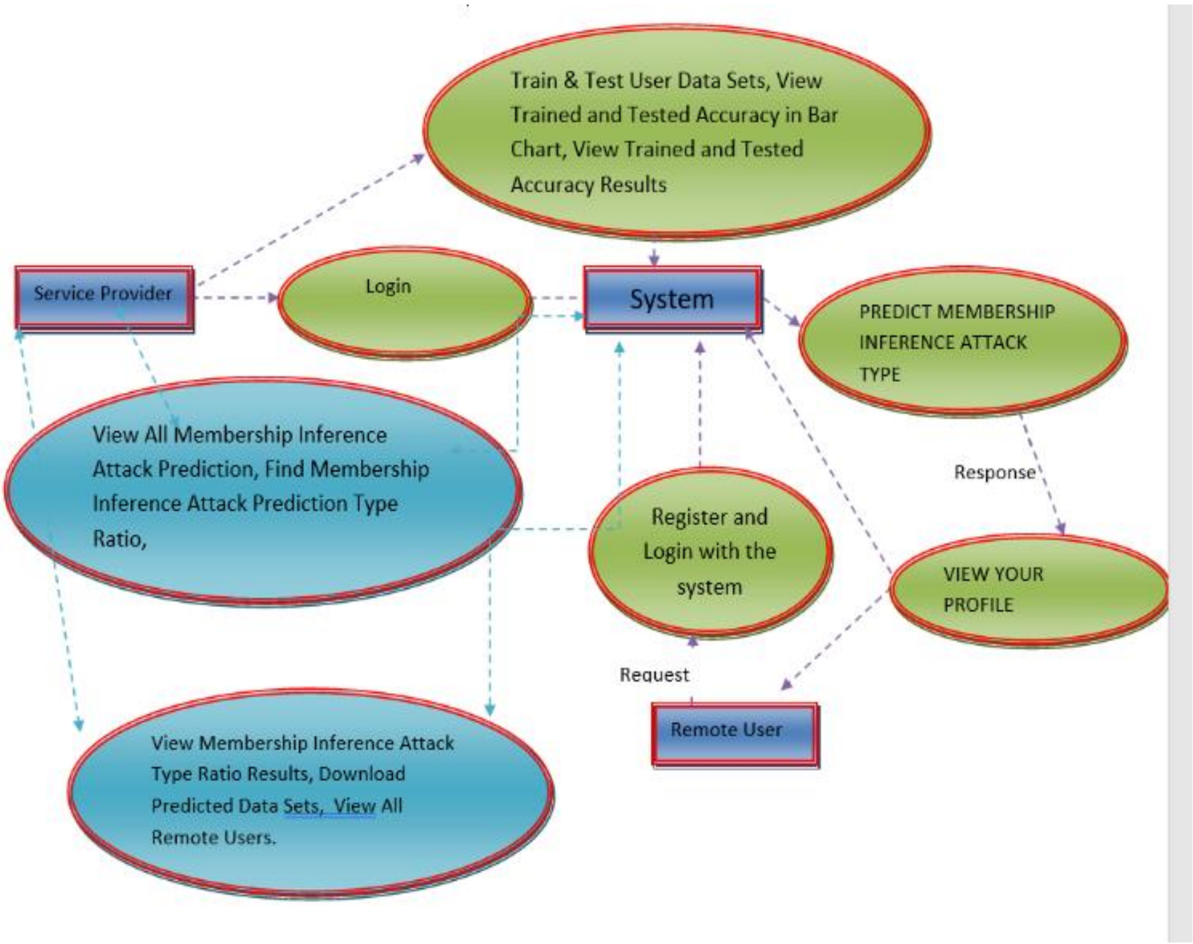

Train & Test User Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results
Service Provider
Login
System
PREDICT MEMBERSHIP INFERENCE ATTACK TYPE
View All Membership Inference Attack Prediction, Find Membership Inference Attack Prediction Type Ratio,
Response
Register and Login with the system
VIEW YOUR PROFILE
Request
Remote User
View Membership Inference Attack Type Ratio Results, Download Predicted Data Sets, View All Remote Users.

Figure 3.2: Dataflow Diagram of a Membership Inference Attack And Defence For Wireless Signal Classifiers With Deep Learning.

4. IMPLEMENTATION

4. IMPLEMENTATION

The implementation phase of a project involves executing the planned strategies and tasks. It requires meticulous coordination, resource allocation, and monitoring to ensure that objectives are met efficiently. Effective implementation is crucial for achieving project goals and delivering expected outcomes within the set timeline and budget constraints.

4.1 ALGORITHMS USED

1. Naive Bayes Classifier

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature . Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias). Relatively efficient compared to traditional machine learning approaches.

<u>Advantages of Naïve Bayes Classifier:</u>

- Simple and computationally efficient for large-scale classification.
- Works well with noisy data, which is common in wireless communication.

<u>Disadvantages of Naïve Bayes Classifier:</u>

- Assumes feature independence, which may not hold true for wireless signals.
- May not perform well with complex signal characteristics.

2. KNN- K Nearest Neighbors

The **K-Nearest Neighbours (KNN)** algorithm is a simple yet powerful machine learning classification technique that relies on a similarity-based approach to classify data points. Unlike traditional machine learning models that require extensive training, KNN follows a lazy learning approach, meaning it does not build an explicit model during the training phase. Instead, it stores the entire dataset and only performs calculations when a new data point needs to be classified.

KNN is a non-parametric algorithm, meaning it does not make any assumptions about the underlying data distribution. This makes it highly flexible and applicable to various classification and regression tasks. The core idea behind KNN is that similar data points exist in close proximity.

In the context of wireless signal classification, KNN can be used to identify modulation schemes, detect anomalies, and classify signals based on their waveform characteristics. Its ability to adapt to different types of data without prior assumptions makes it a valuable tool in machine learning-based wireless communication systems.

However, its performance can be sensitive to the choice of K (the number of neighbours) and the distance metric used, making parameter tuning an important step in achieving optimal results.

Advantages of KNN-Based Models:

- Simple and effective for small-scale signal classification.
- No training phase required, making it easy to implement.

Disadvantages of RNN-Based Models:

- Slow for large datasets as it requires distance calculations for all samples.
- Sensitive to noise and irrelevant features, which can impact accuracy in wireless environments.

3. Support Vector Machine(SVM):

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point *x* and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space.

SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptrons*, both of which are widely used for classification in machine learning.

Advantages of SVM:

- Performs well in high-dimensional spaces, making it suitable for complex signal data.
- Can use different kernel functions to improve classification performance.

Disadvantages of SVM:

- Computationally expensive, especially with large datasets.
- Requires careful parameter tuning to avoid overfitting.

4. Gradient Boosting

Gradient Boosting is a powerful and widely used machine learning technique for regression and classification tasks. It is based on the principle of boosting, where multiple weak models (also called base learners) are combined to form a strong predictive model. The most commonly used base learner in Gradient Boosting is the Decision Tree, and when decision trees are used, the method is referred to as Gradient-Boosted Trees (GBT). This approach generally outperforms Random Forests in many scenarios due to its ability to minimize errors more effectively. Unlike traditional ensemble methods such as bagging (e.g., Random Forests), Gradient Boosting builds models in a stage-wise fashion, meaning that each new model is trained to correct the errors of the previous models. It does this by minimizing a differentiable loss function**.**

Gradient Boosting is widely used in various fields, including financial modeling, healthcare, image recognition, and wireless communication systems. It is particularly effective in handling complex relationships in data and can be fine-tuned using hyperparameters such as the learning rate, number of trees, and tree depth. In the context of wireless signal classification, Gradient Boosting can be applied to identify modulation types, detect intrusions, and enhance security mechanisms. Its ability to optimize different loss functions makes it an excellent choice for high-accuracy classification models.

Advantages of Gradient Boosting:

- Highly accurate and adaptable to different signal types.
- Reduces bias and variance, improving MIA effectiveness.

Disadvantages of Gradient Boosting:

- Computationally intensive due to sequential model training.
- Prone to overfitting if not properly tuned.

In this project, multiple machine learning algorithms, including Support Vector Machine (SVM), Gradient Boosting, K-Nearest Neighbours (KNN), and Naïve Bayes, are utilized for wireless signal classification and membership inference attack (MIA) detection. Each algorithm plays a crucial role in analysing and classifying wireless signals based on waveform, device, and channel characteristics. Such inappropriate content may put bad influence on growing kids and need a technique to prevent such content before showing to kids.

Support Vector Machine (SVM) is used for its ability to effectively separate complex signal patterns using hyperplanes. It is particularly useful in distinguishing between member and non-member signals in the MIA attack, as it can handle high-dimensional data and detect boundary differences between classified and non-classified signals.

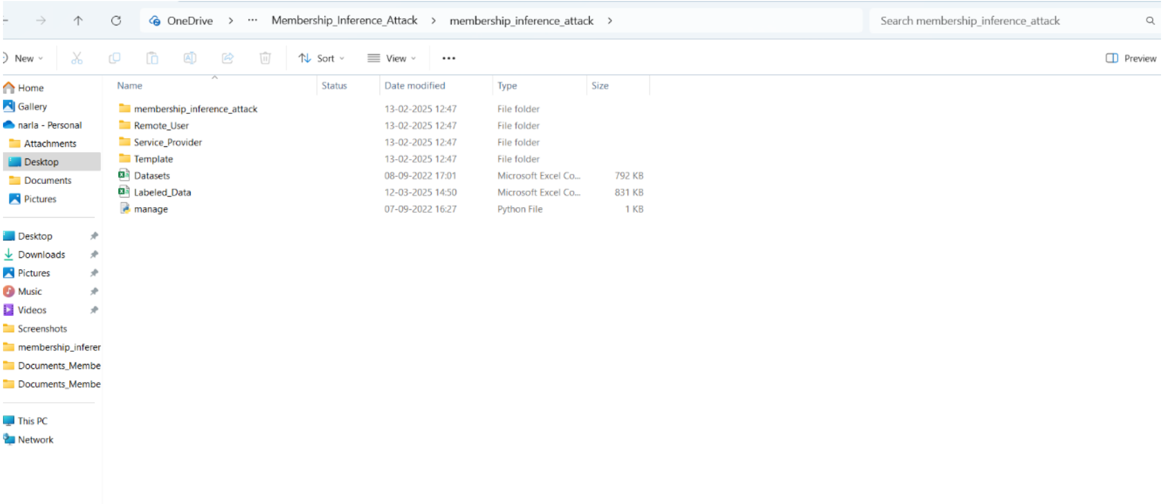
Gradient Boosting, specifically Gradient-Boosted Trees (GBT), is applied for high-accuracy classification by sequentially improving predictions through an ensemble of decision trees. It is effective in minimizing classification errors in wireless signal data, allowing the model to adaptively improve by learning from past mistakes, making it highly useful in detecting MIA threats with enhanced precision.

K-Nearest Neighbours (KNN) is employed for its instance-based learning capability, making it well-suited for real-time classification of wireless signals. By comparing new signals with stored training data, KNN helps in determining whether a signal has been used in training or not, aiding in membership inference analysis.

Naïve Bayes is utilized due to its probabilistic classification approach, making it useful for modulation classification and predicting training data membership based on probability distributions. Its efficiency in handling large datasets with conditional independence assumptions makes it a fast and reliable choice for initial filtering and classification tasks in the project.

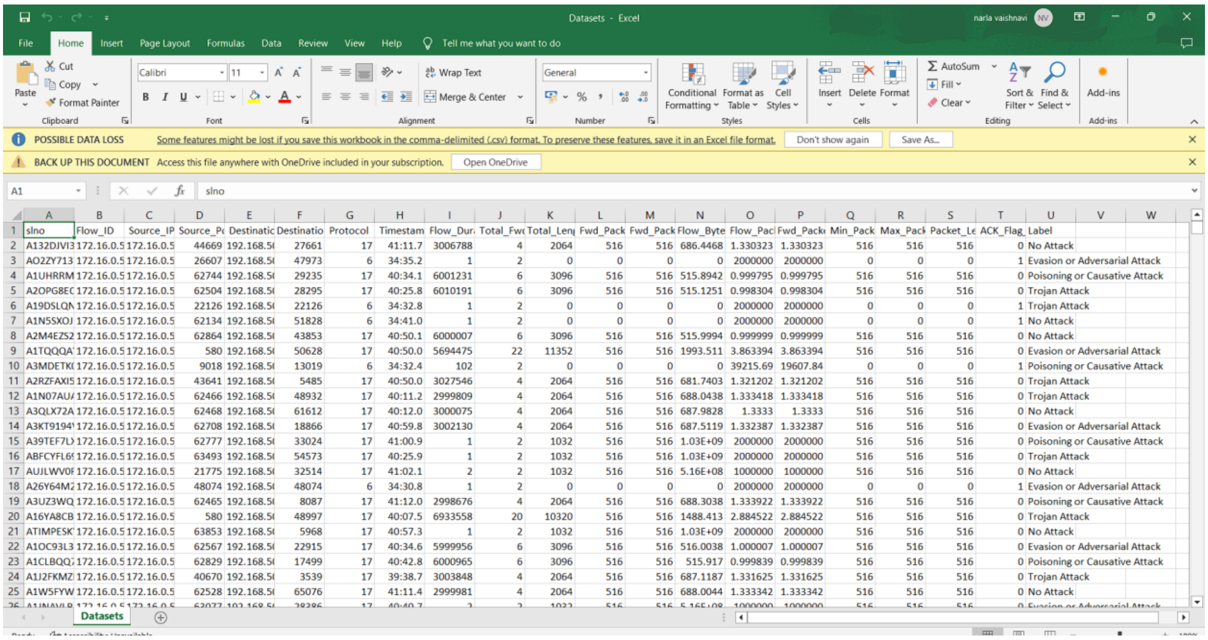
By combining these algorithms, the project ensures robust classification and defence mechanisms against MIAs in wireless networks, enhancing the privacy and security of ML-based wireless signal classifiers.

To train all algorithm we have used below Dataset folders and below screen showing dataset details



Name	Status	Date modified	Type	Size
membership_inference_attack		13-02-2025 12:47	File folder	
Remote_User		13-02-2025 12:47	File folder	
Service_Provider		13-02-2025 12:47	File folder	
Template		13-02-2025 12:47	File folder	
Datasets		08-09-2022 17:01	Microsoft Excel Co...	792 KB
Labeled_Data		12-03-2025 14:50	Microsoft Excel Co...	831 KB
manage		07-09-2022 16:27	Python File	1 KB

Figure 4.1: Dataset Directory Structure of Membership Inference Attack Project, Showing Key Folders and Data Files



POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format. Don't show again Save As...

BACK UP THIS DOCUMENT Access this file anywhere with OneDrive included in your subscription. Open OneDrive

A1 slno

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	slno	Flow_ID	Source_IP	Source_Pc	Destinatic	Destinatio	Protocol	Timestam	Flow_Dur:	Total_Fwc	Total_Len	Fwd_Pack	Fwd_Pack	Flow_Byte	Flow_Pac	Fwd_Pack	Min_Pack	Max_Pack	Packet_Le	ACK_Flag_	Label
2	A132DJVI3	172.16.0.5	172.16.0.5	44669	192.168.5	27661	17	41:11.7	3006788	4	2064	516	516	686.4468	1.330323	1.330323	516	516	516	0	No Attack
3	AO2ZY713	172.16.0.5	172.16.0.5	26607	192.168.5	47973	6	34:35.2	1	2	0	0	0	0	2000000	2000000	0	0	0	1	Evasion or Adversarial Attack
4	A1UHRRM	172.16.0.5	172.16.0.5	62744	192.168.5	29235	17	40:34.1	6001231	6	3096	516	516	515.8942	0.999795	0.999795	516	516	516	0	Poisoning or Causative Attack
5	A2OPG8EC	172.16.0.5	172.16.0.5	62504	192.168.5	28295	17	40:25.8	6010191	6	3096	516	516	515.1251	0.998304	0.998304	516	516	516	0	Trojan Attack
6	A19DSLQN	172.16.0.5	172.16.0.5	22126	192.168.5	22126	6	34:32.8	1	2	0	0	0	0	2000000	2000000	0	0	0	1	Trojan Attack
7	A1N5SXOJ	172.16.0.5	172.16.0.5	62134	192.168.5	51828	6	34:41.0	1	2	0	0	0	0	2000000	2000000	0	0	0	1	No Attack
8	A2M4EZS2	172.16.0.5	172.16.0.5	62864	192.168.5	43853	17	40:50.1	6000007	6	3096	516	516	515.9994	0.999999	0.999999	516	516	516	0	No Attack
9	A1TQQQA'	172.16.0.5	172.16.0.5	580	192.168.5	50628	17	40:50.0	5694475	22	11352	516	516	1993.511	3.863394	3.863394	516	516	516	0	Evasion or Adversarial Attack
10	A3MDETK(	172.16.0.5	172.16.0.5	9018	192.168.5	13019	6	34:32.4	102	2	0	0	0	0	39215.69	19607.84	0	0	0	1	Poisoning or Causative Attack
11	A2RZFAXI5	172.16.0.5	172.16.0.5	43641	192.168.5	5485	17	40:50.0	3027546	4	2064	516	516	681.7403	1.321202	1.321202	516	516	516	0	Trojan Attack
12	A1N07AU/	172.16.0.5	172.16.0.5	62466	192.168.5	48932	17	40:11.2	2999809	4	2064	516	516	688.0438	1.333418	1.333418	516	516	516	0	Trojan Attack
13	A3QLX72A	172.16.0.5	172.16.0.5	62468	192.168.5	61612	17	40:12.0	3000075	4	2064	516	516	687.9828	1.3333	1.3333	516	516	516	0	No Attack
14	A3KT9194'	172.16.0.5	172.16.0.5	62708	192.168.5	18866	17	40:59.8	3002130	4	2064	516	516	687.5119	1.332387	1.332387	516	516	516	0	Evasion or Adversarial Attack
15	A39TEF7L)	172.16.0.5	172.16.0.5	62777	192.168.5	33024	17	41:00.9	1	2	1032	516	516	1.03E+09	2000000	2000000	516	516	516	0	Poisoning or Causative Attack
16	ABFCYFL6(	172.16.0.5	172.16.0.5	63493	192.168.5	54573	17	40:25.9	1	2	1032	516	516	1.03E+09	2000000	2000000	516	516	516	0	Trojan Attack
17	AUJLWV0F	172.16.0.5	172.16.0.5	21775	192.168.5	32514	17	41:02.1	2	2	1032	516	516	5.16E+08	1000000	1000000	516	516	516	0	No Attack
18	A26Y64M2	172.16.0.5	172.16.0.5	48074	192.168.5	48074	6	34:30.8	1	2	0	0	0	0	2000000	2000000	0	0	0	1	Evasion or Adversarial Attack
19	A3UZ3WQ	172.16.0.5	172.16.0.5	62465	192.168.5	8087	17	41:12.0	2998676	4	2064	516	516	688.3038	1.333922	1.333922	516	516	516	0	Poisoning or Causative Attack
20	A16YA8CB	172.16.0.5	172.16.0.5	580	192.168.5	48997	17	40:07.5	6933558	20	10320	516	516	1488.413	2.884522	2.884522	516	516	516	0	Trojan Attack
21	ATIMPESK'	172.16.0.5	172.16.0.5	63853	192.168.5	5968	17	40:57.3	1	2	1032	516	516	1.03E+09	2000000	2000000	516	516	516	0	No Attack
22	A1OC93L3	172.16.0.5	172.16.0.5	62567	192.168.5	22915	17	40:34.6	5999956	6	3096	516	516	516.0038	1.000007	1.000007	516	516	516	0	Evasion or Adversarial Attack
23	A1CLBQQ/	172.16.0.5	172.16.0.5	62829	192.168.5	17499	17	40:42.8	6000965	6	3096	516	516	515.917	0.999839	0.999839	516	516	516	0	Poisoning or Causative Attack
24	A1J2FKMZ	172.16.0.5	172.16.0.5	40670	192.168.5	3539	17	39:38.7	3003848	4	2064	516	516	687.1187	1.331625	1.331625	516	516	516	0	Trojan Attack
25	A1W5FYW	172.16.0.5	172.16.0.5	62528	192.168.5	65076	17	41:11.4	2999981	4	2064	516	516	688.0044	1.333342	1.333342	516	516	516	0	No Attack

Figure 4.2: Screenshot of the Dataset Folder Showing Network Traffic Analysis with Attack Labels of different types of Attacks

To implement this project we have designed following modules:

1) Upload Wireless Signal Dataset: This module allows users to upload the dataset containing wireless signal data, including waveform, channel characteristics, and device-related information (Figure 4.2).

2) Dataset Preprocessing: The uploaded dataset undergoes preprocessing, where signals are cleaned, normalized, and prepared for further analysis.

3) Train Wireless Signal Classifier: A deep neural network (DNN) model is trained using the pre processed dataset to classify wireless signals. The classifier learns to differentiate between signals from various radio devices based on waveform and channel characteristics.

4) Launch Membership Inference Attack (MIA): In this module, an adversary attempts to infer whether a particular signal was part of the training dataset.

5) Over-the-Air MIA using Channel Variations: This module extends the attack by introducing channel variations, making the MIA more robust to real-world wireless environments.

6) Evaluate Attack Success Rate: The effectiveness of the MIA is analysed by measuring the accuracy of inferring training data membership.

7) Comparison and Performance Analysis: A graphical representation of attack success rates and defence effectiveness is generated for better analysis.

8) Real-time Wireless Signal Classification & Protection: In this module, real-time wireless signals are classified while implementing the defence mechanism to safeguard sensitive information.

4.2 SAMPLE CODE

```
from django.db.models import  Count, Avg
from django.shortcuts import render, redirect
import xlwt
from django.http import HttpResponse
import numpy as np

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics import accuracy_score, confusion_matrix,classification_report
from sklearn.metrics import accuracy_score
import pandas as pd

# Create your views here.
from Remote_User.models importClientRegister_Model,
inference_attack_detection,detection_ratio,detection_accuracy

def serviceproviderlogin(request):
    if request.method  == "POST":
        admin = request.POST.get('username')
        password = request.POST.get('password')
        if admin == "Admin" and password =="Admin":
            return redirect('View_Remote_Users')
return  render(request,'SProvider/serviceproviderlogin.html')from  keras.layers  def
View_Membership_Inference_Attack_Prediction(request):
    obj = inference_attack_detection.objects.all()
    returnrender(request,
'SProvider/View_Membership_Inference_Attack_Prediction.html', {'objs': obj})
def View_Membership_Inference_Attack_Prediction_Ratio(request):
    detection_ratio.objects.all().delete()
ratio = ""
    kword = 'No Attack'
    print(kword)
```

```
obj = inference_attack_detection.objects.all().filter(Prediction=kword)
    obj1 = inference_attack_detection.objects.all()
    count = obj.count();
    count1 = obj1.count();
    ratio = (count / count1) * 100
    if ratio != 0:
        detection_ratio.objects.create(names=kword, ratio=ratio)

ratio1 = ""
    kword1 = 'Poisoning or Causative Attack'
    print(kword1)
    obj1 = inference_attack_detection.objects.all().filter(Prediction=kword1)
    obj11 = inference_attack_detection.objects.all()
    count1 = obj1.count();
    count11 = obj11.count();
    ratio1 = (count1 / count11) * 100
    if ratio1 != 0:
        detection_ratio.objects.create(names=kword1, ratio=ratio1)
ratio12 = ""
    kword12 = 'Trojan Attack'
    print(kword12)
    obj12 = inference_attack_detection.objects.all().filter(Prediction=kword12)
    obj112 = inference_attack_detection.objects.all()
    count12 = obj12.count();
    count112 = obj112.count();
    ratio12 = (count12 / count112) * 100
    if ratio12 != 0:
        detection_ratio.objects.create(names=kword12, ratio=ratio12)
    ratio123 = ""
    kword123 = 'Evasion or Adversarial Attack'
    print(kword123)
obj123 = inference_attack_detection.objects.all().filter(Prediction=kword123)
```

```
obj1123 = inference_attack_detection.objects.all()
    count123 = obj123.count();
    count1123 = obj1123.count();
    ratio123 = (count123 / count1123) * 100
    if ratio123 != 0:
        detection_ratio.objects.create(names=kword123, ratio=ratio123)
obj = detection_ratio.objects.all()
 return   render(request,'SProvider/View_Membership_Inference_Attack_
Prediction_Ratio.html', {'objs': obj})
def View_Remote_Users(request):
    obj=ClientRegister_Model.objects.all()
    return render(request,'SProvider/View_Remote_Users.html',{'objects':obj})
def charts(request,chart_type):
    chart1 = detection_ratio.objects.values('names').annotate(dcount=Avg('ratio'))
 return render(request,"SProvider/charts.html",{'form':chart1, 'chart_type':chart_type})
def charts1(request,chart_type):
  chart1 = detection_accuracy.objects.values('names').annotate(dcount=Avg('ratio'))
 return     render(request,"SProvider/charts1.html",{'form':chart1,'chart_type':chart_
type})
def likeschart(request,like_chart):
   charts =detection_accuracy.objects.values('names').annotate(dcount=Avg('ratio'))
   return render(request,"SProvider/likeschart.html",{'form':charts,'like_chart': like_
chart})
def likeschart1(request,like_chart):
    charts =detection_ratio.objects.values('names').annotate(dcount=Avg('ratio'))
    return render(request,"SProvider/likeschart1.html",{'form':charts,'like_chart':
    like_chart})
    def Download_Predicted_DataSets(request):
  response = HttpResponse(content_type='application/ms-excel')
 # decide file name
 response['Content-Disposition'] = 'attachment; filename="Predicted_Datasets.xls"'
# creating workbook
wb = xlwt.Workbook(encoding='utf-8')
```

```
# adding sheet
ws = wb.add_sheet("sheet1")
# Sheet header, first row
row_num = 0
font_style = xlwt.XFStyle()
# headers are bold
font_style.font.bold = True
# writer = csv.writer(response)
obj = inference_attack_detection.objects.all()
data = obj  # dummy method to fetch data.
for my_row in data:
    row_num = row_num + 1

   ws.write(row_num, 0, my_row.slno, font_style)
    ws.write(row_num, 1, my_row.Flow_ID, font_style)
    ws.write(row_num, 2, my_row.Source_IP, font_style)
    ws.write(row_num, 3, my_row.Source_Port, font_style)
    ws.write(row_num, 4, my_row.Destination_IP, font_style)
    ws.write(row_num, 5, my_row.Destination_Port, font_style)
    ws.write(row_num, 6, my_row.Protocol, font_style)
    ws.write(row_num, 7, my_row.Timestamp, font_style)
    ws.write(row_num, 8, my_row.Flow_Duration, font_style)
    ws.write(row_num, 9, my_row.Total_Fwd_Packets, font_style)
    ws.write(row_num, 10, my_row.Total_Length_of_Fwd_Packets, font_style)
    ws.write(row_num, 11, my_row.Fwd_Packet_Length_Max, font_style)
    ws.write(row_num, 12, my_row.Fwd_Packet_Length_Min, font_style)
    ws.write(row_num, 13, my_row.Flow_Bytes_per_second, font_style)
    ws.write(row_num, 14, my_row.Flow_Packets_per_second, font_style)
    ws.write(row_num, 15, my_row.Fwd_Packets_per_second, font_style)
    ws.write(row_num, 16, my_row.Min_Packet_Length, font_style)
    ws.write(row_num, 17, my_row.Max_Packet_Length, font_style)
    ws.write(row_num, 18, my_row.Packet_Length_ean, font_style)
    ws.write(row_num, 19, my_row.ACK_Flag_Count, font_style)
```

```
        ws.write(row_num, 20, my_row.Prediction, font_style)
 wb.save(response)
    return response


def Train_Test_DataSets(request):
    detection_accuracy.objects.all().delete()
    df = pd.read_csv('Datasets.csv',encoding='latin-1')
        df['label'] = df['Label'].map({'No Attack':0,'Poisoning or Causative
Attack':1,'Trojan Attack':2,'Evasion or Adversarial Attack':4})
    #cv = CountVectorizer()
    X = df['slno']
    y = df["label"]


    print("X Values")
    print(X)
    print("Labels")
    print(y)
cv = CountVectorizer(lowercase=False, strip_accents='unicode', ngram_range=(1,
1))
    X = cv.fit_transform(df['slno'].apply(lambda x: np.str_(X)))
    #X = cv.fit_transform(X)
    labeled = 'Labeled_Data.csv'
    df.to_csv(labeled, index=False)
    df.to_markdown
models = []
    from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33,
random_state=42)
    X_train.shape, X_test.shape, y_train.shape
print("X_test")
print(X_test)
print(X_train)
    print("Naive Bayes")

    from sklearn.naive_bayes import MultinomialNB
    NB = MultinomialNB()
    NB.fit(X_train, y_train)
    predict_nb = NB.predict(X_test)
    naivebayes = accuracy_score(y_test, predict_nb) * 100
    print("ACCURACY")
    print(naivebayes)
    print("CLASSIFICATION REPORT")
    print(classification_report(y_test, predict_nb))
    print("CONFUSION MATRIX")
    print(confusion_matrix(y_test, predict_nb))
    detection_accuracy.objects.create(names="Naive Bayes", ratio=naivebayes)

    # SVM Model
    print("SVM")
    from sklearn import svm
    lin_clf = svm.LinearSVC()
    lin_clf.fit(X_train, y_train)
    predict_svm = lin_clf.predict(X_test)
    svm_acc = accuracy_score(y_test, predict_svm) * 100
    print(svm_acc)
    print("CLASSIFICATION REPORT")
    print(classification_report(y_test, predict_svm))
    print("CONFUSION MATRIX")
    print(confusion_matrix(y_test, predict_svm))
    models.append(('svm', lin_clf))
    detection_accuracy.objects.create(names="SVM", ratio=svm_acc)
 print("KNeighborsClassifier")
    from sklearn.neighbors import KNeighborsClassifier
```

```
kn = KNeighborsClassifier()
    kn.fit(X_train, y_train)
    knpredict = kn.predict(X_test)
    print("ACCURACY")
    print(accuracy_score(y_test, knpredict) * 100)
    print("CLASSIFICATION REPORT")
    print(classification_report(y_test, knpredict))
    print("CONFUSION MATRIX")
    print(confusion_matrix(y_test, knpredict))
    models.append(('KNeighborsClassifier', kn))
                        detection_accuracy.objects.create(names="KNeighborsClassifier",
ratio=accuracy_score(y_test, knpredict) * 100)

    print("Gradient Boosting Classifier")
    from sklearn.ensemble import GradientBoostingClassifier
          clf = GradientBoostingClassifier(n_estimators=100, learning_rate=1.0,
max_depth=1, random_state=0).fit(
        X_train,
        y_train)
    clfpredict = clf.predict(X_test)
    print("ACCURACY")
    print(accuracy_score(y_test, clfpredict) * 100)
    print("CLASSIFICATION REPORT")
    print(classification_report(y_test, clfpredict))
    print("CONFUSION MATRIX")
    print(confusion_matrix(y_test, clfpredict))
    models.append(('GradientBoostingClassifier', clf))
    detection_accuracy.objects.create(names="Gradient Boosting Classifier",
                                      ratio=accuracy_score(y_test, clfpredict) * 100)

obj = detection_accuracy.objects.all()
return render(request,'SProvider/Train_Test_DataSets.html', {'objs': obj})
```

5. RESULTS AND DISCUSSION

5. RESULTS AND DISCUSSION

The following screenshots showcase the results of our project, highlighting key features and functionalities. These visual representations provide a clear overview of how the system performs under various conditions, demonstrating its effectiveness and user interface. The screenshots serve as a visual aid to support the project's technical and operational achievements.

5.1 USER REGISTER :

In below screen, The User can enter the Details and Click on Register Then the user able to Login

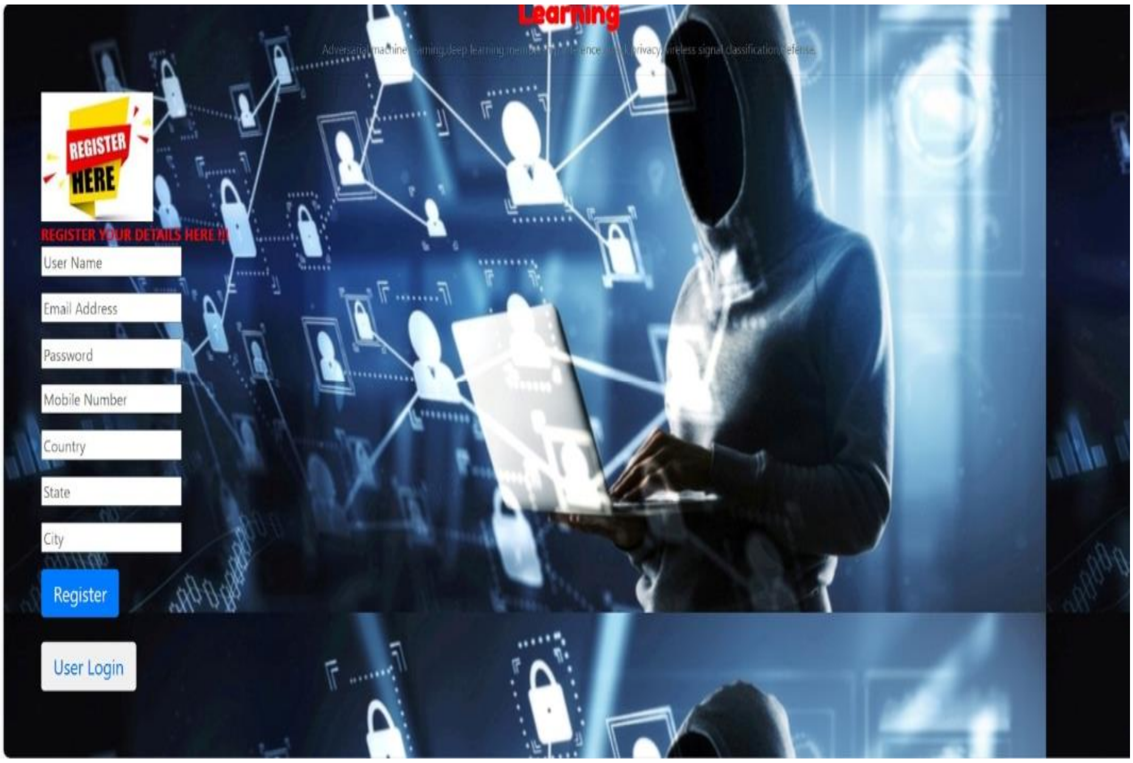


Figure 5.1 : User register interface of A Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning .

5.2 USER LOGIN:

In below screen, The user can login into the application by entering Username and Password. Here we use username and password for authentication purpose. User login is a fundamental authentication process that verifies a user's identity before granting access to a system or application. It typically involves entering credentials such as a username or email and a password, which are validated against stored records in a database.

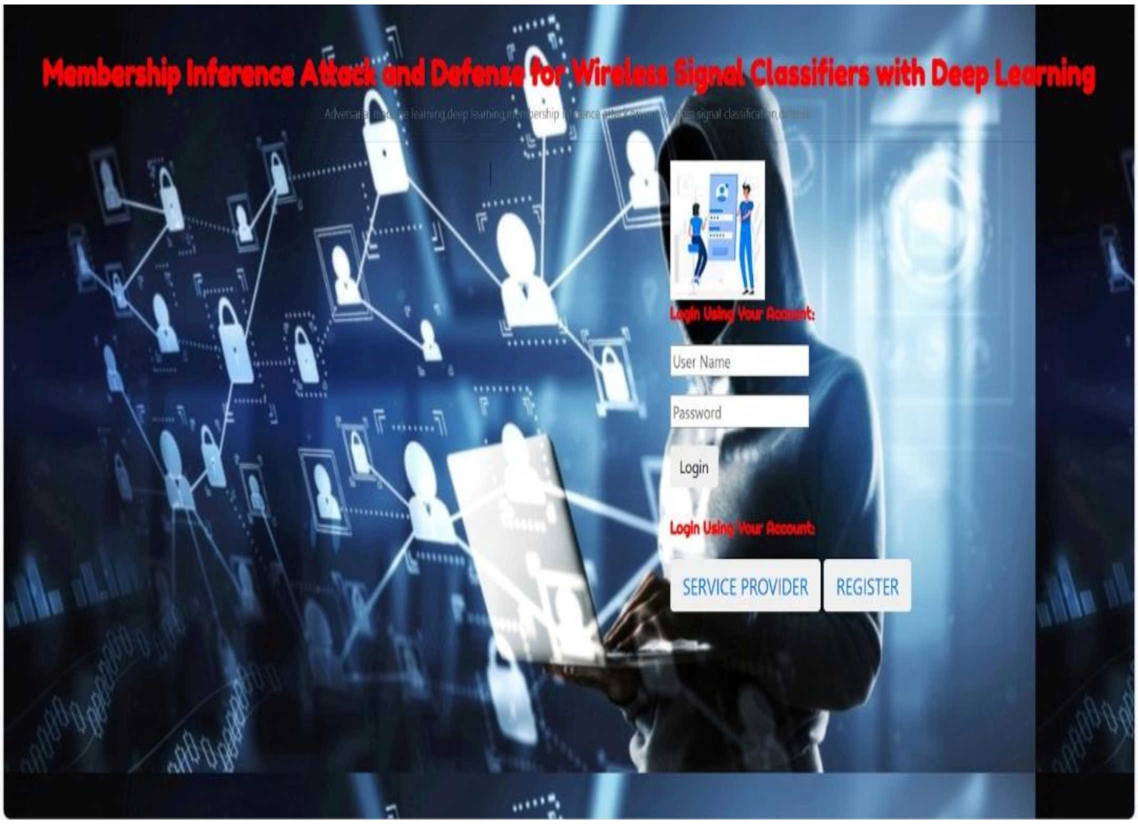
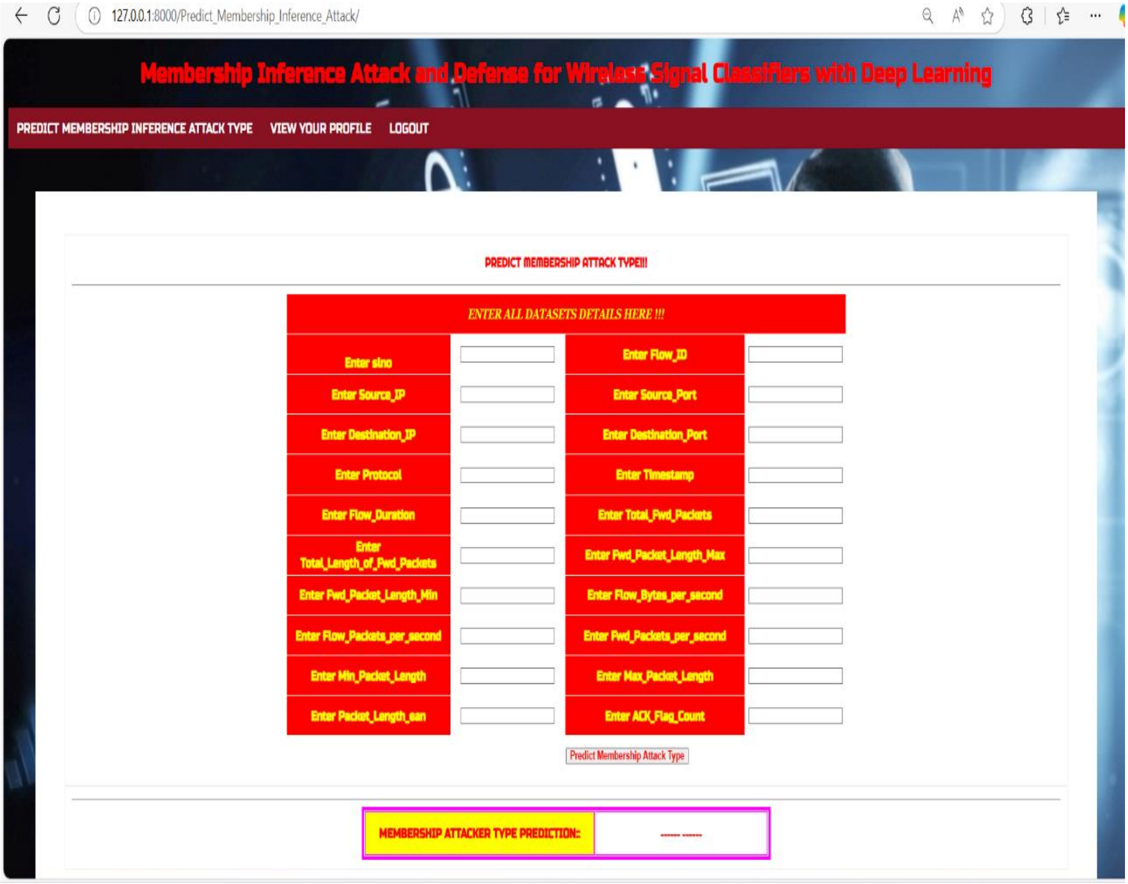

Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
Login Using Your Account:
User Name
Password
Login
Login Using Your Account:
SERVICE PROVIDER
REGISTER

Figure 5.2 : User Login Interface for Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning.

5.3 MEMBERSHIP INFERENCE ATTACK TYPE PREDICTION:

In below screen, shows a web-based interface for predicting membership inference attack types in wireless signal classifiers. Here User has to enter the details of the attack which is included in the dataset. The user has to fill all the labels which is present below to find the type of attack occurred.


127.0.0.1:8000/Predict_Membership_Inference_Attack/
Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
PREDICT MEMBERSHIP INFERENCE ATTACK TYPE
VIEW YOUR PROFILE
LOGOUT
Predict Membership Attack Type

PREDICT MEMBERSHIP ATTACK TYPE!!!

ENTER ALL DATASETS DETAILS HERE !!!

Enter slno		Enter Flow_ID	
Enter Source_IP		Enter Source_Port	
Enter Destination_IP		Enter Destination_Port	
Enter Protocol		Enter Timestamp	
Enter Flow_Duration		Enter Total_Fwd_Packets	
Enter Total_Length_of_Fwd_Packets		Enter Fwd_Packet_Length_Max	
Enter Fwd_Packet_Length_Min		Enter Flow_Bytes_per_second	
Enter Flow_Packets_per_second		Enter Fwd_Packets_per_second	
Enter Min_Packet_Length		Enter Max_Packet_Length	
Enter Packet_Length_ean		Enter ACK_Flag_Count	

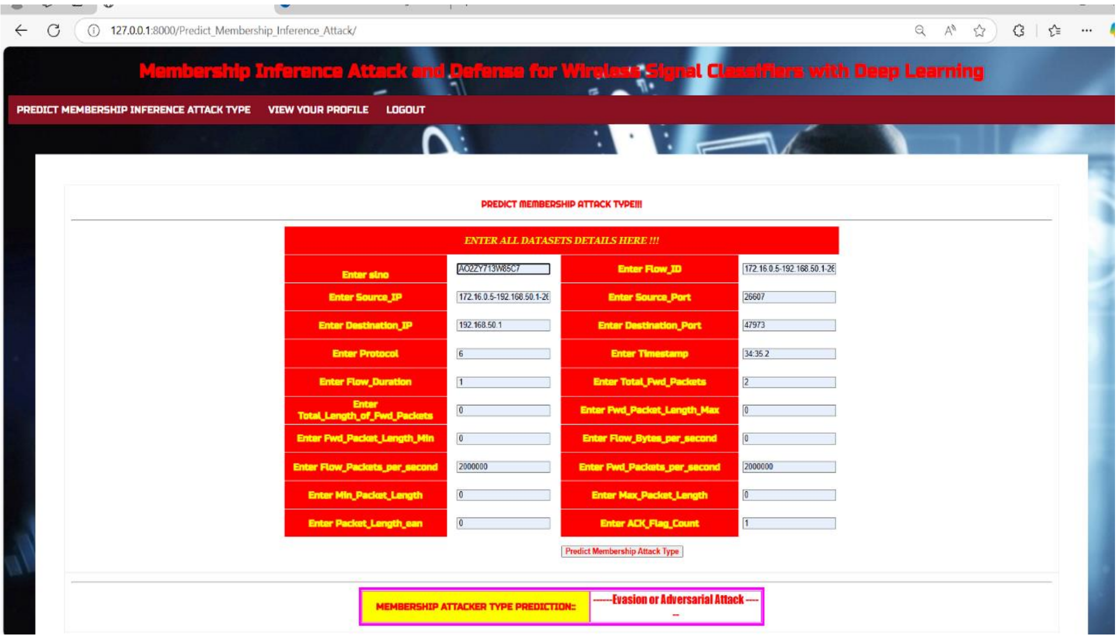
MEMBERSHIP ATTACKER TYPE PREDICTION:	------ ------

Figure 5.3 : Membership Inference Attack Type Prediction Interface for Membership Inference Attack and Defence for wireless Signal Classifiers with Deep Learning.

5.4 MEMBERSHIP INFERENCE ATTACK TYPE PREDICTION RESULT:

In below screen, a web interface for predicting membership inference attack types using a machine-learning model. The User can enter the details under the label of an attack and click on the “Predict Membership Attack type” then it displays the predicted attack type.

The result shown is "Evasion or Adversarial Attack", suggesting that the analyzed data exhibits characteristics of this specific attack type.


127.0.0.1:8000/Predict_Membership_Inference_Attack/
Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
PREDICT MEMBERSHIP INFERENCE ATTACK TYPE
VIEW YOUR PROFILE
LOGOUT

PREDICT MEMBERSHIP ATTACK TYPE!!!

ENTER ALL DATASETS DETAILS HERE !!!

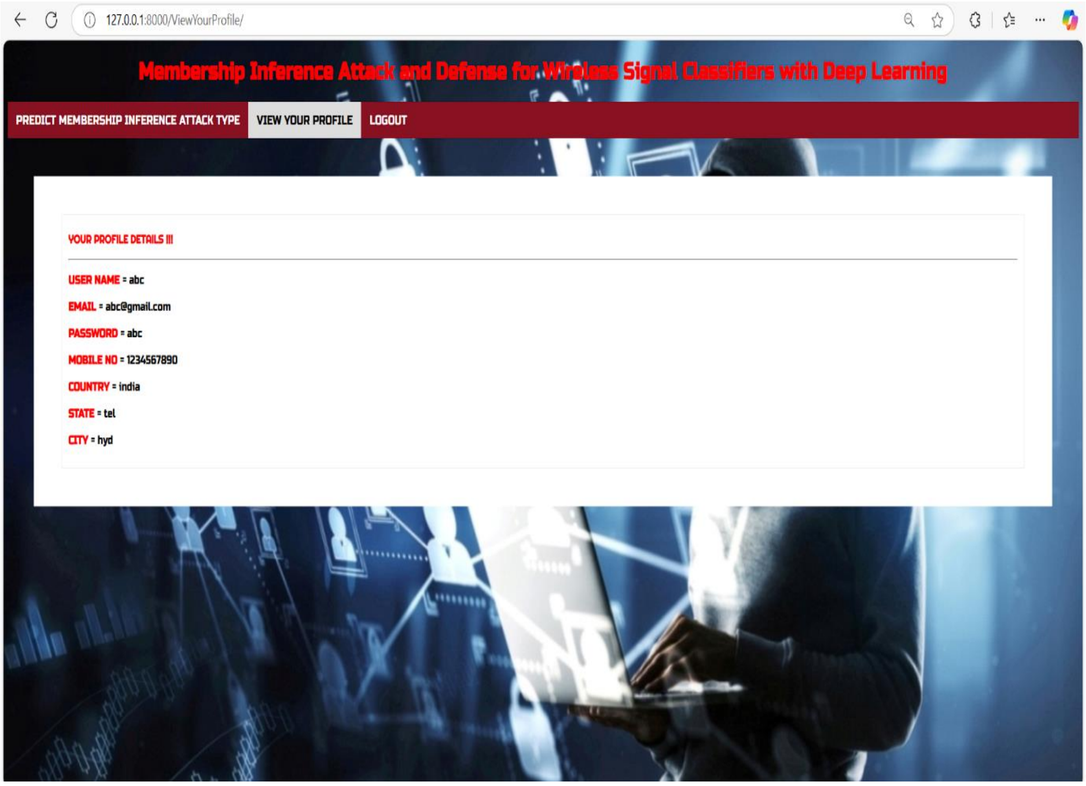
Enter slno	AO2ZY713W85C7	Enter Flow_ID	172.16.0.5-192.168.50.1-26
Enter Source_IP	172.16.0.5-192.168.50.1-26	Enter Source_Port	26607
Enter Destination_IP	192.168.50.1	Enter Destination_Port	47973
Enter Protocol	6	Enter Timestamp	34:35.2
Enter Flow_Duration	1	Enter Total_Fwd_Packets	2
Enter Total_Length_of_Fwd_Packets	0	Enter Fwd_Packet_Length_Max	0
Enter Fwd_Packet_Length_Min	0	Enter Flow_Bytes_per_second	0
Enter Flow_Packets_per_second	2000000	Enter Fwd_Packets_per_second	2000000
Enter Min_Packet_Length	0	Enter Max_Packet_Length	0
Enter Packet_Length_ean	0	Enter ACK_Flag_Count	1

MEMBERSHIP ATTACKER TYPE PREDICTION:: ------Evasion or Adversarial Attack -----

Figure 5.4 : Membership Inference Attack Prediction Result for Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning.

5.5 USER PROFILE DETAILS:

In below screen, The User can view his details by clicking “VIEW YOUR PROFILE” This Interface shows the Personal information of the user which consists of Username, Email, Password, Mobile no, Country, State, and City.



127.0.0.1:8000/ViewYourProfile/
Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
PREDICT MEMBERSHIP INFERENCE ATTACK TYPE
VIEW YOUR PROFILE
LOGOUT

YOUR PROFILE DETAILS !!!

USER NAME = abc

EMAIL = abc@gmail.com

PASSWORD = abc

MOBILE NO = 1234567890

COUNTRY = india

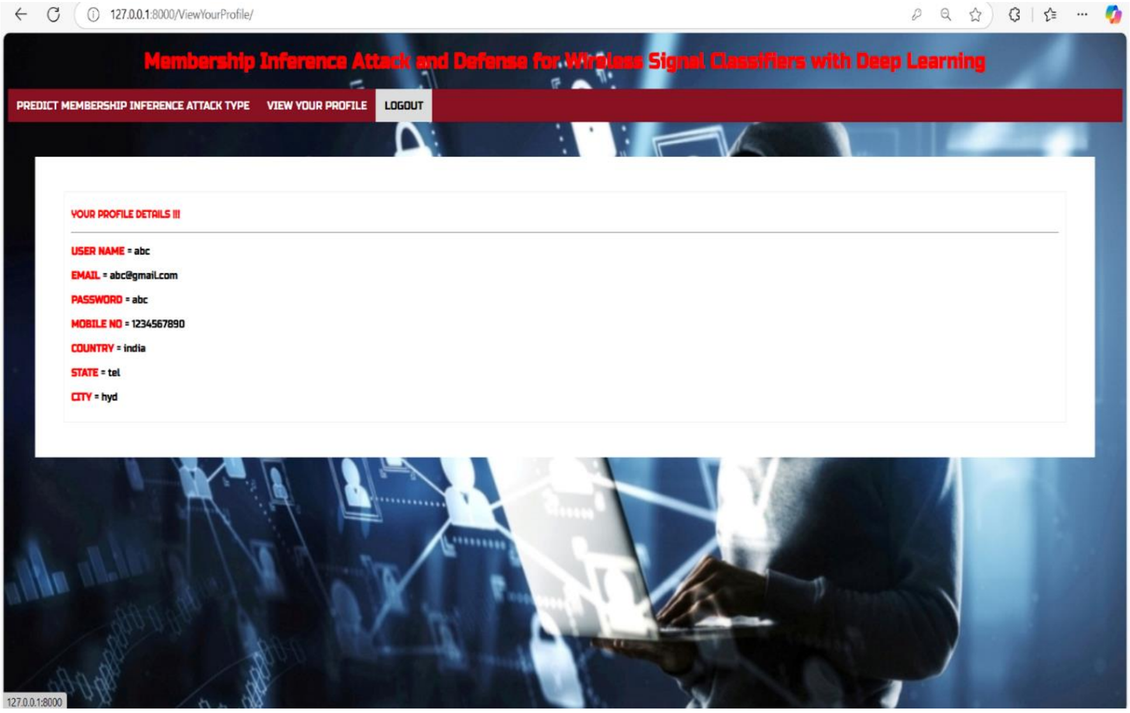
STATE = tel

CITY = hyd

Figure 5.5 : User Profile Page Displaying Personal Information in a Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning.

5.6 USER LOGOUT

In below screen, The User can Logout from the interface by clicking “Logout” option. Logout is a crucial security feature in web applications that ensures the termination of a user's authenticated session, preventing unauthorized access to sensitive data. When a user logs out, the system clears session data, invalidates authentication tokens, and often redirects the user to a login or homepage. This process is essential for maintaining privacy, securing user accounts, and managing system resources effectively.


127.0.0.1:8000/ViewYourProfile/
Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
PREDICT MEMBERSHIP INFERENCE ATTACK TYPE
VIEW YOUR PROFILE
LOGOUT
127.0.0.1:8000

YOUR PROFILE DETAILS !!!

USER NAME = abc

EMAIL = abc@gmail.com

PASSWORD = abc

MOBILE NO = 1234567890

COUNTRY = india

STATE = tel

CITY = hyd

Figure 5.6 : User Logout Interface for Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning.

5.7 SERVICE PROVIDER LOGIN:

In below screen, The Service Provider can login into the application by entering Username and Password. Here we use username and password for authentication purpose.

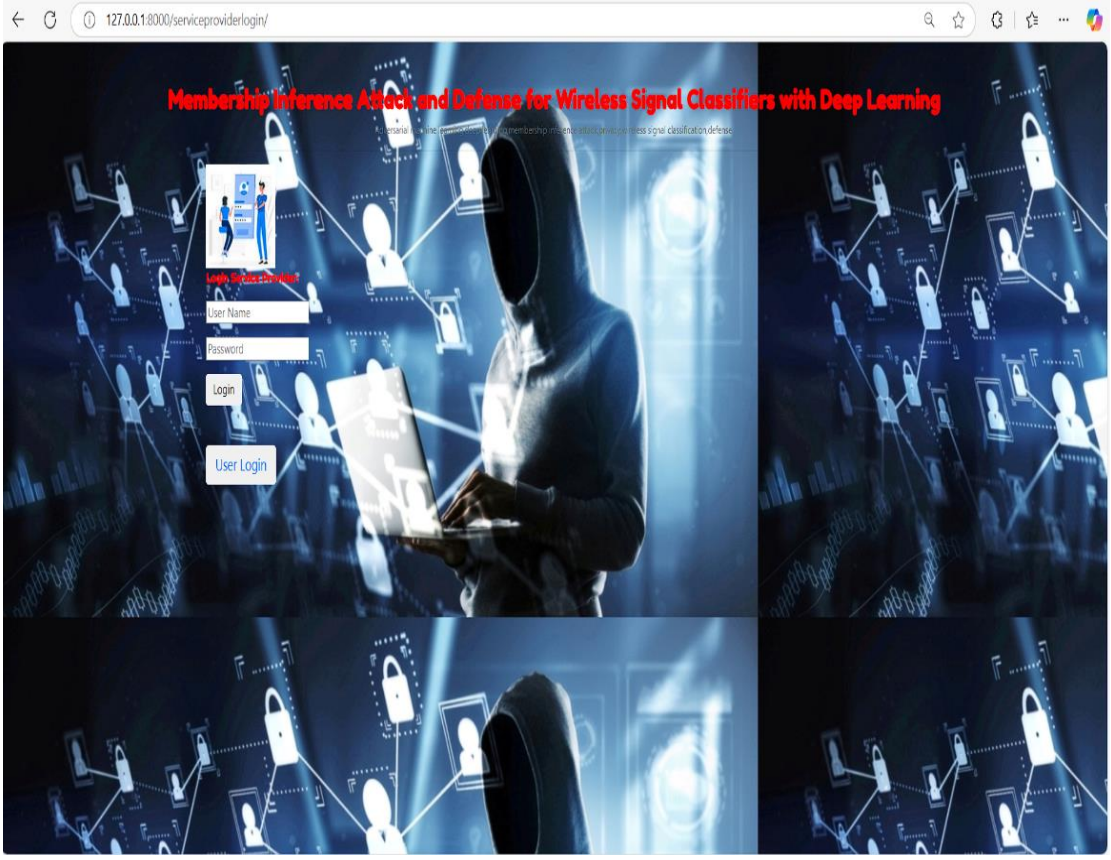
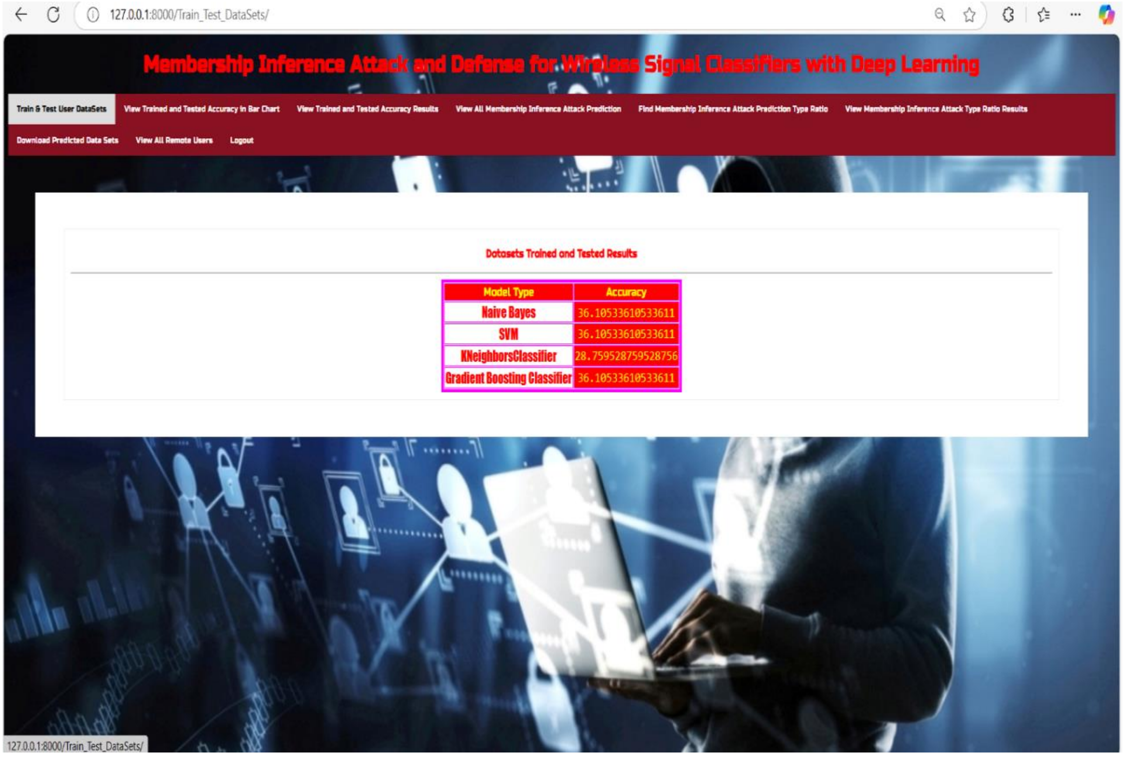

127.0.0.1:8000/serviceproviderlogin/
Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
Adversarial machine learning,deep learning,membership inference attack,privacy,wireless signal classification,defense
Login Service Provider
User Name
Password
Login
User Login

Figure 5.7 : Service Provider Login Interface for Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning.

5.8 TRAIN AND TEST DATASETS:

In below screen, It displays the "Datasets Trained and Tested Results" page of a Membership Inference Attack, and showcasing the accuracy of different machine learning models used for detection. We get the accuracy for each algorithm based on the performance.



127.0.0.1:8000/Train_Test_DataSets/
Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
Train & Test User DataSets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View All Membership Inference Attack Prediction
Find Membership Inference Attack Prediction Type Ratio
View Membership Inference Attack Type Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout
127.0.0.1:8000/Train_Test_DataSets/

Datasets Trained and Tested Results

Model Type	Accuracy
Naive Bayes	36.10533610533611
SVM	36.10533610533611
KNeighborsClassifier	28.759528759528756
Gradient Boosting Classifier	36.10533610533611

Figure 5.8 : Trained and Tested Model Accuracy Results for Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning.

5.9 VIEW RESULTS IN BAR CHART:

In below screen, It displays a bar chart visualization of the trained and tested accuracy results of different machine learning models used in Predicting membership inference attacks. The bar chart compares the performance of four different models based on their accuracy percentages. Each model is represented by a different coloured bar, with accuracy values labelled above them.

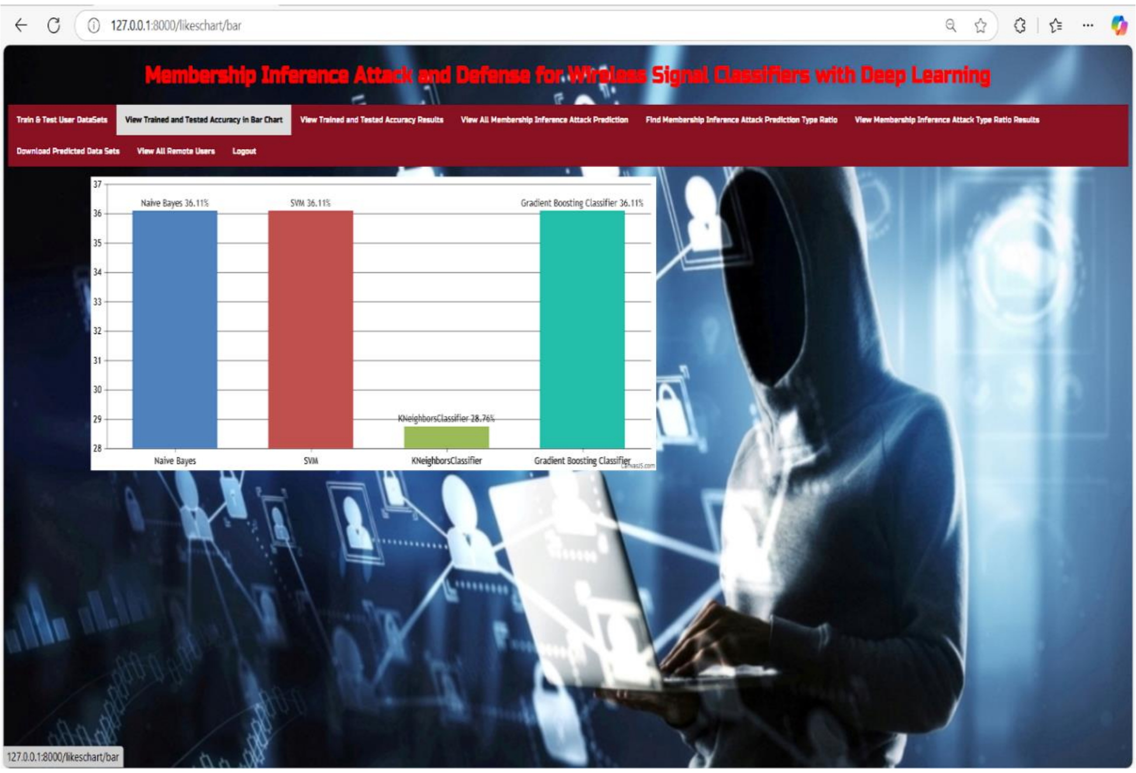

127.0.0.1:8000/likeschart/bar
Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
Train & Test User DataSets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View All Membership Inference Attack Prediction
Find Membership Inference Attack Prediction Type Ratio
View Membership Inference Attack Type Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout
37
36
35
34
33
32
31
30
29
28
Naive Bayes 36.11%
SVM 36.11%
KNeighborsClassifier 28.76%
Gradient Boosting Classifier 36.11%
Naive Bayes
SVM
KNeighborsClassifier
Gradient Boosting Classifier
CanvasJS.com
127.0.0.1:8000/likeschart/bar

Figure 5.9 : Trained and tested Bar Chart Representation of Model Accuracy for Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning.

5.10 VIEW RESULTS IN LINE CHART

In below screen, It displays the “line chart visualization” of the trained and tested accuracy results for different machine learning models used in membership inference attack detection. The line chart connects accuracy values for four different models, showing their comparative performance. Each point on the chart represents the accuracy of a specific model, with Naïve Bayes, SVM, and Gradient Boosting Classifier performing equally well, while KNeighborsClassifier has a significant drop.

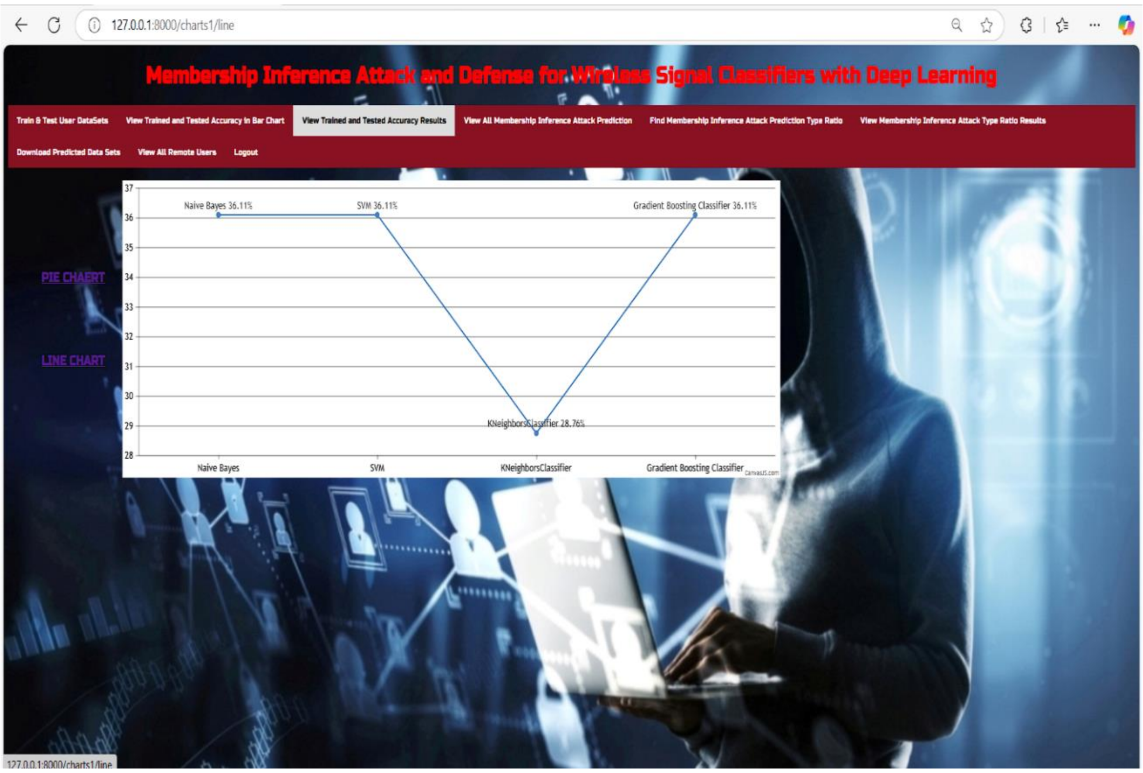
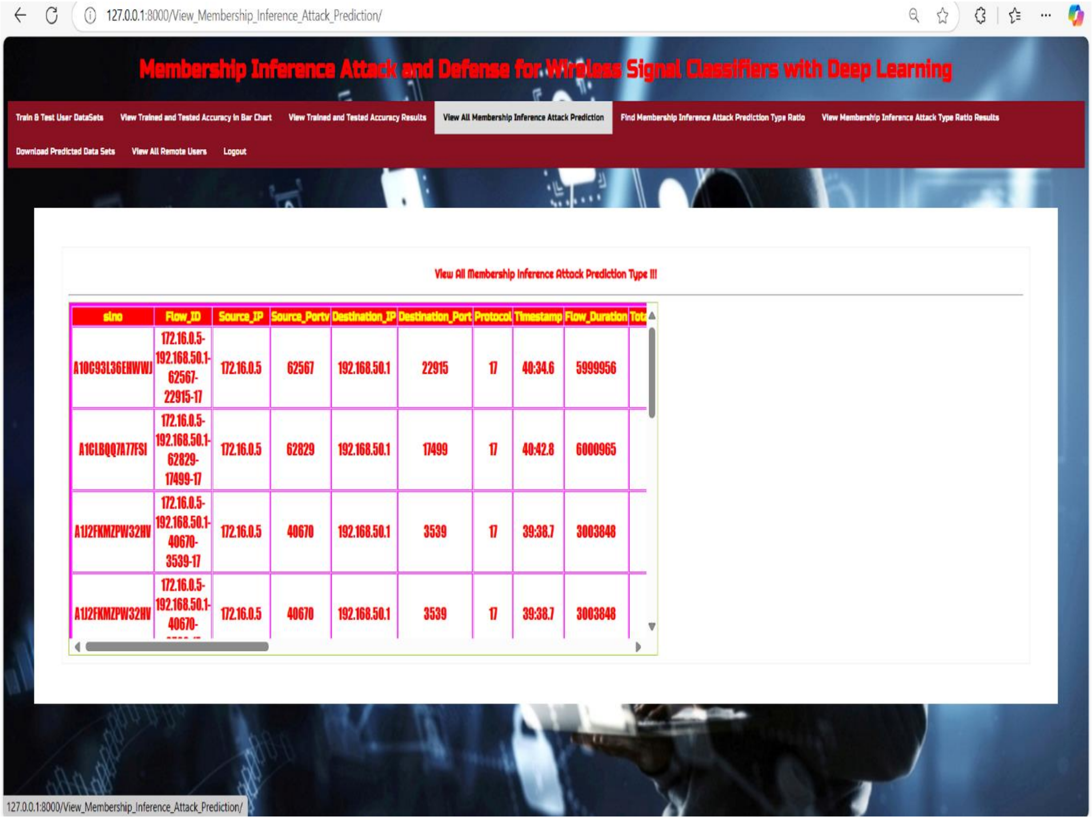

127.0.0.1:8000/charts1/line
Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
Train & Test User DataSets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View All Membership Inference Attack Prediction
Find Membership Inference Attack Prediction Type Ratio
View Membership Inference Attack Type Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout
PIE CHAERT
LINE CHART
37
36
35
34
33
32
31
30
29
28
Naive Bayes 36.11%
SVM 36.11%
KNeighborsClassifier 28.76%
Gradient Boosting Classifier 36.11%
Naive Bayes
SVM
KNeighborsClassifier
Gradient Boosting Classifier
CanvasJS.com
127.0.0.1:8000/charts1/line

Figure 5.10 : Line Chart Representation of Model Accuracy for Membership Inference Attack and Defence for Wireless Signal Classifiers With Deep Learning.

5.11 VIEW ALL MEMBERSHIP INFERECE ATTACK PREDICTION TYPES

In below Screen, It represents a data table displaying details of network traffic flows as part of Membership Inference Attack Predictions in a wireless signal classifier system using deep learning. The table contains multiple columns with critical parameters used for identifying and analyzing potential inference attacks on a system.


127.0.0.1:8000/View_Membership_Inference_Attack_Prediction/
Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
Train & Test User DataSets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View All Membership Inference Attack Prediction
Find Membership Inference Attack Prediction Type Ratio
View Membership Inference Attack Type Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout
127.0.0.1:8000/View_Membership_Inference_Attack_Prediction/

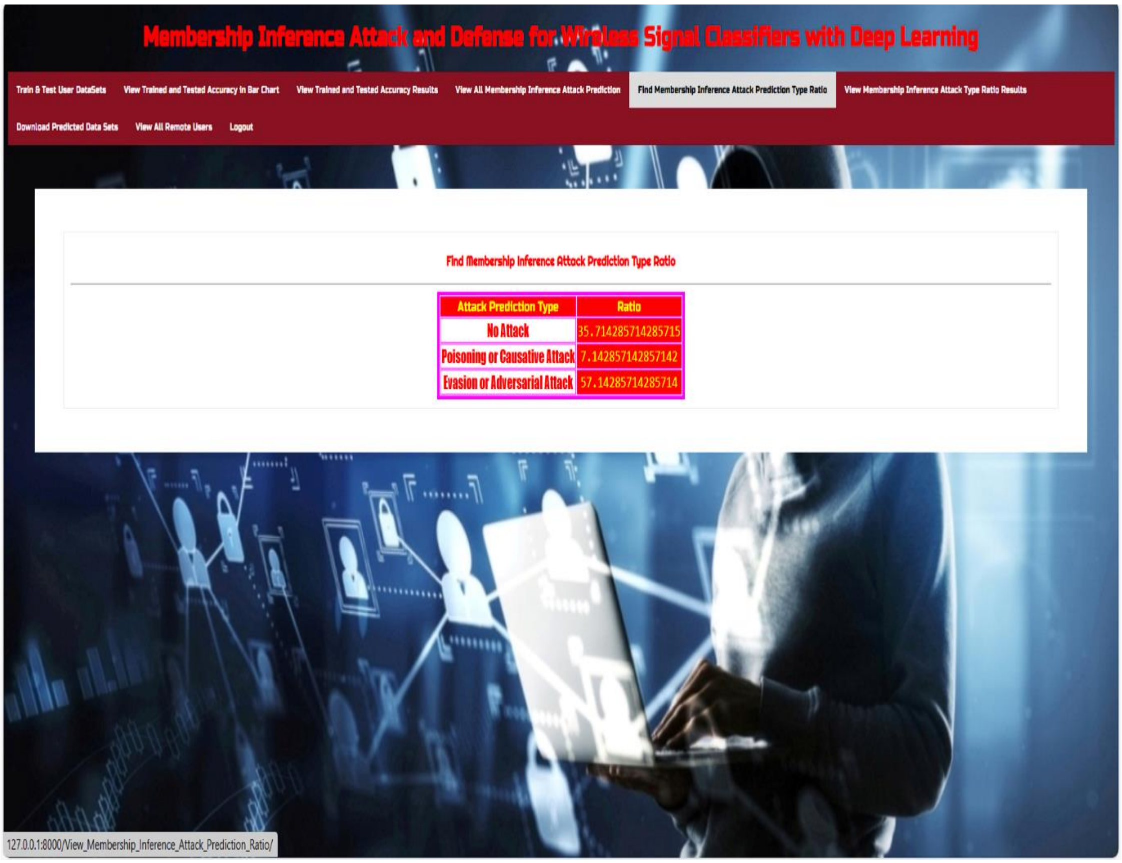
View All Membership Inference Attack Prediction Type !!!

slno	Flow_ID	Source_IP	Source_Portv	Destination_IP	Destination_Port	Protocol	Timestamp	Flow_Duration	Tota
A10C93L36EHWWJ	172.16.0.5-192.168.50.1-62567-22915-17	172.16.0.5	62567	192.168.50.1	22915	17	40:34.6	5999956	
A1CLBQQ7A77FSI	172.16.0.5-192.168.50.1-62829-17499-17	172.16.0.5	62829	192.168.50.1	17499	17	40:42.8	6000965	
A1J2FKMZPW32HV	172.16.0.5-192.168.50.1-40670-3539-17	172.16.0.5	40670	192.168.50.1	3539	17	39:38.7	3003848	
A1J2FKMZPW32HV	172.16.0.5-192.168.50.1-40670-	172.16.0.5	40670	192.168.50.1	3539	17	39:38.7	3003848	

Figure 5.11: Membership Inference Attack Prediction Table Displaying Network Traffic Data for Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning.

5.12 FIND MEMBERSHIP INFERENCE ATTACK PREDICTION TYPE RATIO

In below screen, it displays a table highlighting the different types of membership inference attack predictions detected in a wireless signal classification system using deep learning**.** The table presents the probability ratio of different attack types in relation to the overall analysed data. Attack Prediction Type – Classifies the type of attack detected in the system. Ratio – Displays the probability percentage of each attack type occurring in the analysed dataset.


Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
Train & Test User DataSets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View All Membership Inference Attack Prediction
Find Membership Inference Attack Prediction Type Ratio
View Membership Inference Attack Type Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout
127.0.0.1:8000/View_Membership_Inference_Attack_Prediction_Ratio/

Find Membership Inference Attack Prediction Type Ratio

Attack Prediction Type	Ratio
No Attack	35.714285714285715
Poisoning or Causative Attack	7.142857142857142
Evasion or Adversarial Attack	57.14285714285714

Figure 5.12: Membership Inference Attack Type Ratio Analysis for Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning.

5.13 VIEW MEMBERSHIP INFERENCE ATTACK RATIO RESULTS IN LINE CHART

In below screen. It showcases a line chart visualizing the ratio of different membership inference attack types in a wireless signal classification system using deep learning. The chart provides a graphical representation of attack prevalence, making it easier to interpret the severity and frequency of different attack types.

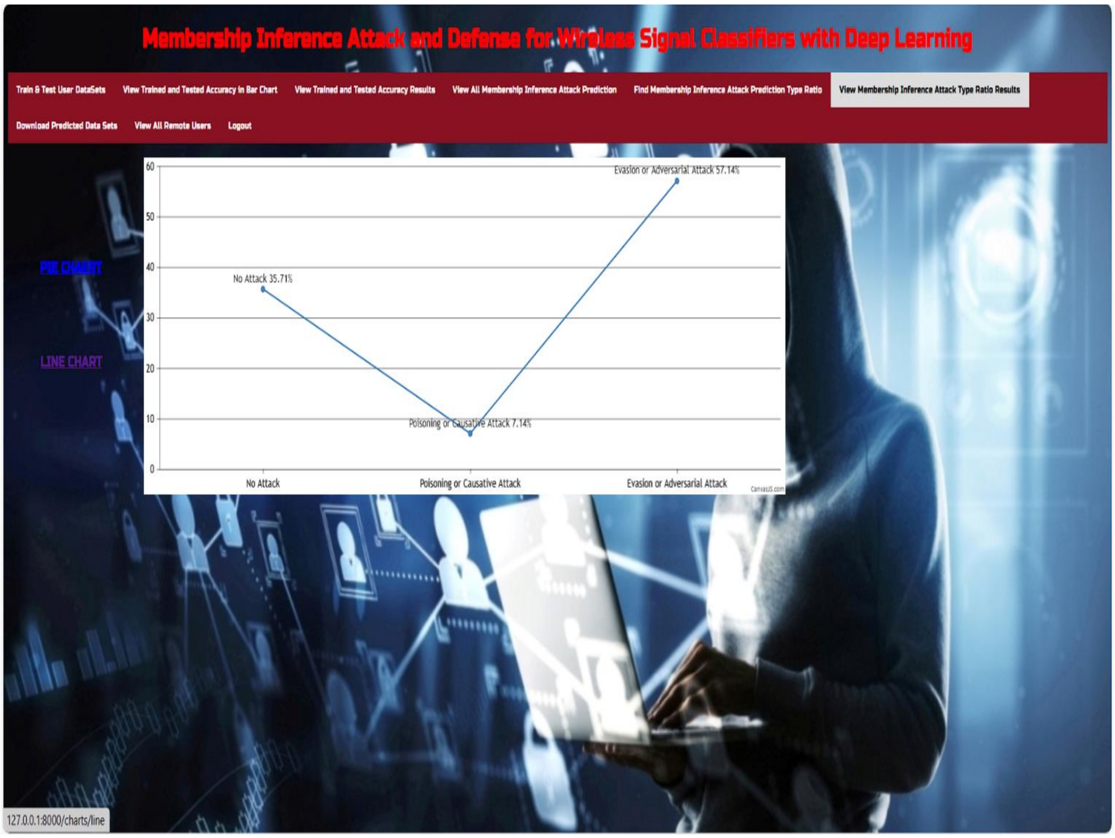

Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
Train & Test User DataSets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View All Membership Inference Attack Prediction
Find Membership Inference Attack Prediction Type Ratio
View Membership Inference Attack Type Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout
PIE CHART
LINE CHART
60
50
40
30
20
10
0
No Attack 35.71%
Poisoning or Causative Attack 7.14%
Evasion or Adversarial Attack 57.14%
No Attack
Poisoning or Causative Attack
Evasion or Adversarial Attack
CanvasJS.com
127.0.0.1:8000/charts/line

Figure 5.13: Line Chart Representation Membership Inference Attack Type Ratio for Membership Inference Attack and Defence for Wireless Signal Classifiers With Deep Learning.

5.14 DOWNLOAD PREDICATED DATASETS

In below screen, It displays the “Downloaded predicated datasets”. The dataset consists with the various types of attacks and their data.

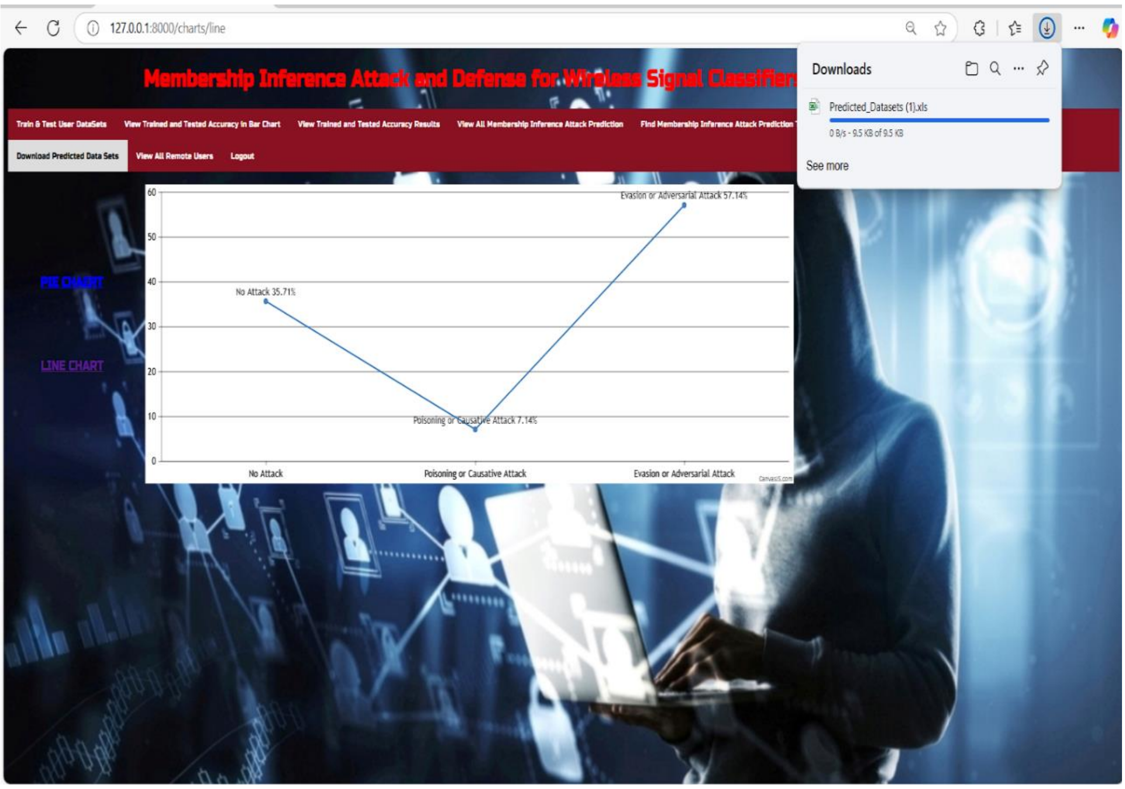
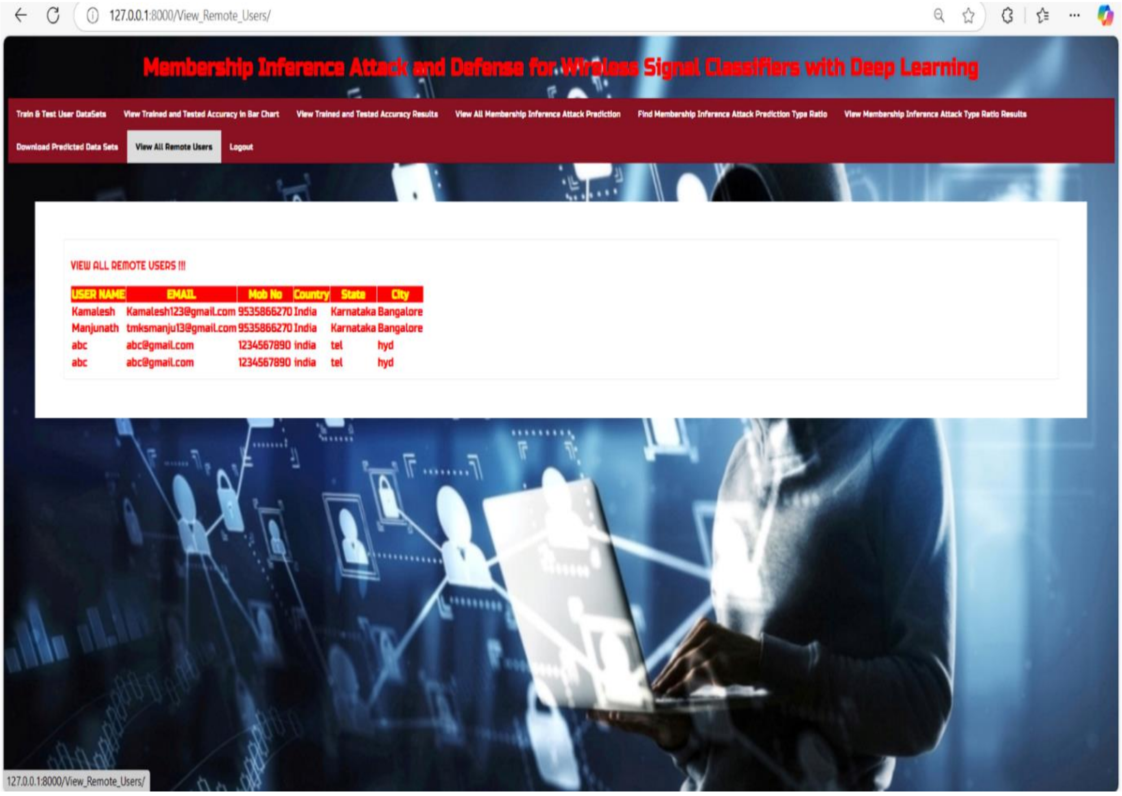

127.0.0.1:8000/charts/line
Membership Inference Attack and Defense for Wireless Signal Classifiers
Train & Test User DataSets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View All Membership Inference Attack Prediction
Find Membership Inference Attack Prediction
Download Predicted Data Sets
View All Remote Users
Logout
PIE CHART
LINE CHART
60
50
40
30
20
10
0
No Attack 35.71%
Poisoning or Causative Attack 7.14%
Evasion or Adversarial Attack 57.14%
No Attack
Poisoning or Causative Attack
Evasion or Adversarial Attack
CanvasJS.com
Downloads
Predicted_Datasets (1).xls
0 B/s - 9.5 KB of 9.5 KB
See more

Figure 5.14: Download Predicated Datasets for Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning.

5.15 VIEW ALL REMOTE USERS

In below screen, It shows a remote user management interface for a Membership Inference Attack and Defence system. It displays a table listing user details, including their username, email, mobile number, country, state, and city.


127.0.0.1:8000/View_Remote_Users/
Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
Train & Test User DataSets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View All Membership Inference Attack Prediction
Find Membership Inference Attack Prediction Type Ratio
View Membership Inference Attack Type Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout
127.0.0.1:8000/View_Remote_Users/

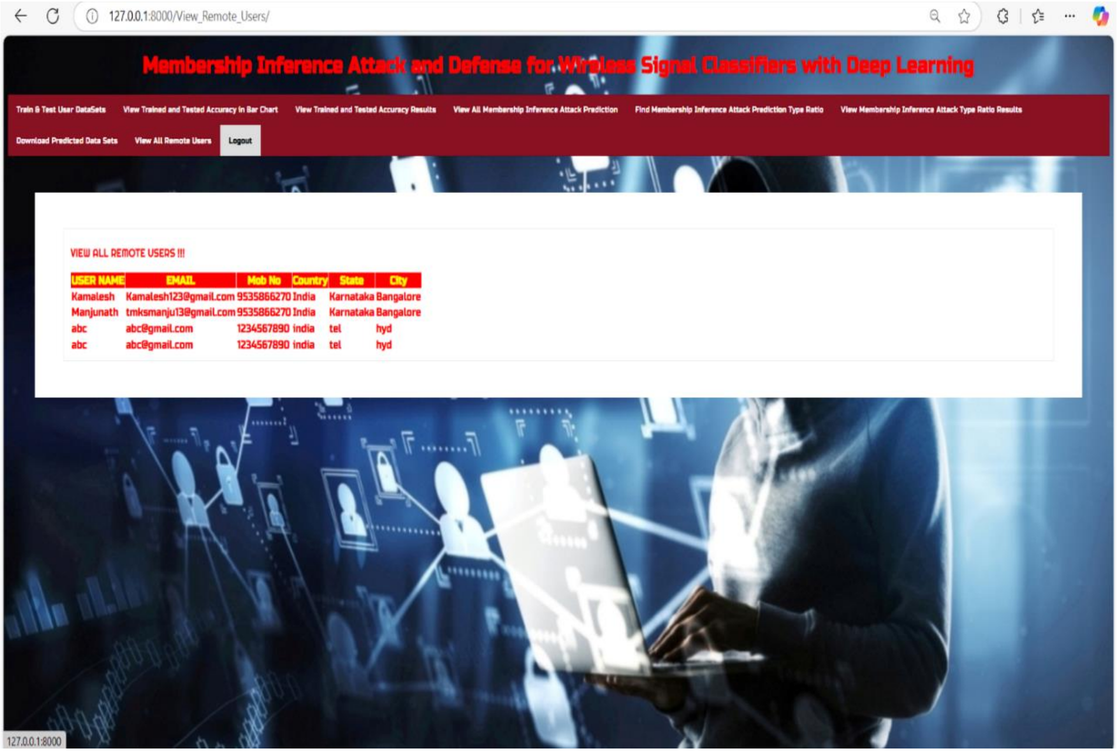
VIEW ALL REMOTE USERS !!!

USER NAME	EMAIL	Mob No	Country	State	City
Kamalesh	Kamalesh123@gmail.com	9535866270	India	Karnataka	Bangalore
Manjunath	tmksmanju13@gmail.com	9535866270	India	Karnataka	Bangalore
abc	abc@gmail.com	1234567890	india	tel	hyd
abc	abc@gmail.com	1234567890	india	tel	hyd

Figure 5.15: Remote Users List in Membership Inference Attack and Defence for Wireless Signal Classifiers With Deep Learning.

5.16 SERVICE PROVIDER LOGOUT

In below screen, The Service Provider can Logout from the interface by clicking “Logout” Option. Logout is a crucial security feature in web applications that ensures the termination of a user's authenticated session, preventing unauthorized access to sensitive data. When a Service provider logs out, the system clears session data, invalidates authentication tokens, and often redirects the user to a login or homepage. This process is essential for maintaining privacy, securing user accounts, and managing system resources effectively.


127.0.0.1:8000/View_Remote_Users/
Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
Train & Test User DataSets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View All Membership Inference Attack Prediction
Find Membership Inference Attack Prediction Type Ratio
View Membership Inference Attack Type Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout
127.0.0.1:8000

VIEW ALL REMOTE USERS !!!

USER NAME	EMAIL	Mob No	Country	State	City
Kamalesh	Kamalesh123@gmail.com	9535866270	India	Karnataka	Bangalore
Manjunath	tmksmanju13@gmail.com	9535866270	India	Karnataka	Bangalore
abc	abc@gmail.com	1234567890	india	tel	hyd
abc	abc@gmail.com	1234567890	india	tel	hyd

Figure 5.16: Service provider Logout Interface for Membership Inference Attack and Defence for Wireless Signal Classifiers With Deep Learning

6. VALIDATION

6. VALIDATION

The validation of this project primarily relies on extensive testing and well-defined test cases to ensure the accuracy and effectiveness of the inappropriate content detection system. The testing process involves multiple stages, including dataset validation, model performance evaluation, and real-world testing. By implementing a structured validation approach, we can ensure that the system consistently delivers high accuracy in detecting inappropriate content while minimizing false positives and false negatives.

6.1 INTRODUCTION

This study validates the Membership Inference Attack (MIA) threat in deep learning-based wireless systems and demonstrates the effectiveness of a proposed defense mechanism. The validation process confirms that adversaries can successfully infer training data membership, raising significant privacy concerns in wireless communication networks such as 5G and IoT environments.

Attack Validation Results: The MIA was tested under different settings, and the evaluation metrics confirm that the attack can reliably infer private information: Same-device non-member signals: The MIA achieved 88.62% accuracy for strong signals and 77.01% accuracy for weak signals.

Impact of channel noise: The accuracy of MIA varied based on noise modeling. Using the average score led to reduced inference accuracy, while the maximum score increased inference success on member samples and decreased it for non-member samples.

Defense Validation Results To mitigate the privacy risks of MIA, a defense mechanism was developed that applies controlled perturbations to the classification process. The validation results confirm: In the first setting (same-device non-member signals), the attack accuracy was already moderate, and the defense slightly reduced MIA accuracy by about 5%.

In the second setting (different-device non-member signals), the defense was highly effective, reducing MIA accuracy from 97.88% to 50%, significantly limiting adversarial success.

6.2 TEST CASES

TABLE 6.2.1 UPLOADING DATASET

S.NO	Test Case	Excepted Result	Result	Remarks (IF Fails)
1.	User Register	If User Registered Successfully.	Pass	Fails if email already exists.
2.	User Login	If Username and password correct then is it will getting valid page.	Pass	Fails for unregistered users.
3.	User Predict MIA attack type	Show type of attack occurred.	Pass	Fails if details do not match dataset.
4.	User View His Profile details	User Profile details.	Pass	Fails if data mismatch or not found.
5.	Service Provider Login	If Username and password correct then is it will getting valid page.	Pass	Fails for unregistered service provider.
6.	Train and Test results	Display All algorithms with accuracy.	Pass	Fails if models not trained/displayed.
7.	View Results in Bar chart	Display Results in Bar chart.	Pass	Fails if chart not loaded.
8.	View Results in Line chart	Display Results in Line chart.	Pass	Fails if chart not loaded.
9.	View All MIA Types	Display All types of MIA.	Pass	Fails if no types shown.
10.	View All MIA Type Ratio	Display Ratio Results For each Attack.	Pass	Fails if ratios not shown.
11.	View MIA type ratio results in line chart	Display Ratio Results in line charts.	Pass	Fails if chart not generated.
12.	Download predicated datasets.	Downloaded Predicated datasets.	Pass	Fails if file not generated.
13.	View all Remote Users	Display All remote Users.	Pass	Fails if user data not fetched.

TABLE 6.2.2 CLASSIFICATION

Test case ID	Test case name	Purpose	Input	Output
1	Classification test 1	To check if the attack is occurred.	Attack type dataset is selected.	Displays Attack name.
2	Classification test 2	To check if the attack is Occurred.	No Attack type dataset is selected.	Displays No Attack

7. CONCLUSION AND FUTURE ASPECTS

7. CONCLUSION AND FUTURE ASPECTS

In conclusion, the project has successfully achieved its objectives, showcasing significant progress and outcomes. The implementation and execution phases were meticulously planned and executed, leading to substantial improvements and insights. Looking ahead, the future aspects of the project hold immense potential. Future developments will focus on expanding the scope, integrating new technologies, and enhancing sustainability. These advancements will not only strengthen the existing framework but also open new avenues for growth and innovation, ensuring the project remains relevant and impactful in the long term. This strategic approach will drive continuous improvement and success.

7.1 PROJECT CONCLUSION

In this project, we studied the MIA as a novel privacy threat against ML-based wireless applications. The target application is a DL-based classifier to identify authorized users by their RF fingerprint. An example use case for this attack is PHY-layer user authentication in 5G or IoT systems. The input of this model consists of the received power and the phase shift. An adversary launches the MIA to infer whether signals of interest have been used to train this wireless signal classifier or not. In this attack, the adversary needs to collect signals and their classification results by observing the spectrum. Then, it can build a surrogate classifier namely a functionally equivalent classifier as the target classifier at the intended receiver. We showed that the surrogate classifier can be reliably built by the adversary under various settings. Then, the adversary launches the MIA to identify whether for a received signal, its corresponding signal received at the service provider is in the training data or not.

In the first setting where non-member signals can be generated by the same devices, the MIA accuracy is 88:62% for strong signals and 77:01% for weak signals. We studied the case that the member inference is investigated not only for received signals but also their noisy variations due to random channel effects. If the average score is used to predict the membership inference for original signals and their noisy variations, the accuracy of the MIA decreases with the level of noisy variations. On the other hand, if the maximum score is used, the accuracy on member samples increases while the accuracy on non-member samples decreases. In the second setting where non-member signals are generated by different devices, the MIA achieves better performance (97:88% accuracy).

7.2 FUTURE ASPECTS

This project establishes a strong foundation for understanding Membership Inference Attacks (MIAs) on wireless signal classifiers and proposes a defence mechanism. However, there is room for further research and improvements in various directions:
These are some future aspect :

1. Stronger Privacy-Preserving Techniques: Implementing distributed training across multiple devices without sharing raw data can reduce privacy risks. Encrypting data during training and inference can ensure privacy protection against MIAs and other adversarial attacks.

2. Advanced Defensive Mechanisms: Training the classifier against synthetic adversarial examples can make it more resistant to MIAs. Developing real-time detection and response mechanisms that can identify and counteract MIA attempts dynamically. Integrating machine learning-based security systems that continuously monitor and detect suspicious activities in wireless networks.

3. Expanding Attack Scenarios: Studying MIAs in multi-user environments (e.g., 5G network slicing or IoT networks) to see if attackers can infer training data across different network segments. Combining MIAs with other attacks like spoofing, jamming, or side-channel attacks to evaluate combined security risks.

4. Real-World Implementation and Testing: Implementing the attack and defense models in 5G/6G network infrastructures to test their real-world impact. Deploying the attack and defence strategies on software-defined radios (SDRs) to test performance in practical wireless systems.

8. BIBLIOGRAPHY

8. BIBLIOGRAPHY

8.1 REFERENCES

[1] Y. Shi, K. Davaslioglu, and Y. E. Sagduyu, "Over-the-Air Membership Inference Attacks as Privacy Threats for Deep Learning-based Wireless Signal Classifiers,"

[2] T. Erpek, T. O'Shea, Y. E. Sagduyu, Y. Shi, and T. C. Clancy, "Deep Learning for Wireless Communications" in Development and Analysis of Deep Learning Architectures, Springer, 2020

[3] Y. E. Sagduyu, Y. Shi, T. Erpek, W. Headley, B.Flowers, G. Stantchev, and Z. Lu, "When Wireless Security Meets Machine Learn ing: Motivation, Challenges, and Research Directions," arXiv preprint arXiv:2001.08883, 2020.

[4] D. Adesina D, C. C. Hsieh, Y. E. Sagduyu, and L. Qian, "Adversarial Machine Learning in Wireless Communications using RF Data: A Review," arXiv preprint arXiv:2012.14392, 2020.

[5] T. Erpek, Y. E. Sagduyu, and Y. Shi, "Deep Learning for Launching and Mitigating Wireless Jamming Attacks," IEEE Transactions on Cognitive Communications and Networking, Mar. 2019.

[6] M. Sadeghi and E. G. Larsson, "Physical Adversarial Attacks Against End-to-end Autoencoder Communication Systems," IEEE Communica tions Letters, May 2019.

[7] Y. Shi, Y.E. Sagduyu, and A. Grushin, "How to Steal a Machine Learn ing Classifier with Deep Learning," IEEE Symposium on Technologies for Homeland Security (HST), 2017.

[8] X. Qiu, J. Dai, and M. Hayes, "A Learning Approach for Physical Layer Authentication using Adaptive Neural Network," IEEE Access, 2020.

[9] N. Wang, T. Jiang, S. Lv, and L. Xiao, "Physical-Layer Authentication based on Extreme Learning Machine," IEEE Communications Letters, July 2017.

[10] K. Leino and M. Fredrikson, "Stolen Memories: Leveraging Model Memorization for Calibrated White-Box Membership Inference," arXiv preprint, http://arxiv.org/abs/1906.11798, 2019.

8.2 GITHUB LINK

https://github.com/Vaishnavi-Narla/MIA-and-defence-for-wireless-signal-classifier-using-deep-learning