A project report on

HUMAN ACTIVITY RECOGNITION USING WIFI CSI

Submitted in partial fulfillment for the award of the degree of

Bachelor of Technology in

Computer Science and Engineering with specialization in AI ML

by

POJESH KUMAR R (22BAI1373)

ARUN AM (22BAI1370)

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

November, 2025

A project report on

HUMAN ACTIVITY RECOGNITION USING WIFI CSI

Submitted in partial fulfillment for the award of the degree of

Bachelor of Technology in

Computer Science and Engineering with specialization in AI ML

by

POJESH KUMAR R (22BAI1373)

ARUN AM (22BAI1370)

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

November, 2025

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

DECLARATION

I hereby declare that the thesis entitled "HUMAN ACTIVITY RECOGNITION USING WIFI CSI" submitted by Pojesh Kumar R (22BAI1373) for the award of the degree of Bachelor of Technology in Computer Science and Engineering with specialization in AI ML, Vellore Institute of Technology, Chennai is a record of bonafide work carried out by me under the supervision of Dr. Sherley A.

I further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place: Chennai

Date: 03-11-2025

Signature of the Candidate

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

School of Computer Science and Engineering

CERTIFICATE

This is to certify that the report entitled **"Human Activity Recognition using WiFi CSI"** is prepared and submitted by **Pojesh Kumar R (22BAI1373)** to Vellore Institute of Technology, Chennai, in partial fulfillment of the requirement for the award of the degree of **Bachelor of Technology in Computer Science and Engineering with specialization in AI ML** is a bonafide record carried out under my guidance. The project fulfills the requirements as per the regulations of this University and in my opinion meets the necessary standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma and the same is certified.

Signature of the Guide:
Name: Dr. Sherly A
Date: 5/11/25

Signature of the Examiner
Name: Dr. Abdul Quadir Md
Date: 05/11/2025

Signature of the Examiner
Name: Dr. Elakiya E
Date: 5/11/2025

Approved by the Head of Department,
Computer Science and Engineering AI ML

Name: Dr. Tulasi Prasad Sariki
Date: 05/11/2025

VELLORE INSTITUTE OF TECHNOLOGY
SCHOOL OF COMPUTER SCIENCE & ENGINEERING (SCOPE)
Chennai - 600127

ABSTRACT

Human activity recognition utilizing WiFi Channel State Information (CSI) has emerged as an innovative non-intrusive sensing technology within smart environments. However, prevalent methodologies frequently exhibit pronounced subject-specific biases, which considerably hinder their capacity to generalize across unobserved individuals. This study seeks to rectify this significant shortcoming by presenting an adversarial network architecture tailored for the Wi-Fi sensing domain, which is adept at acquiring subject-invariant feature representations conducive to robust classification.

We devised a hybrid convolutional neural network-gated recurrent unit (CNN-GRU) baseline model to effectively capture both spatial and temporal dependencies inherent in CSI signals. The baseline architecture attained an accuracy of 98% under a random data partition, thereby demonstrating commendable performance in subject-dependent contexts. Nonetheless, when subjected to leave-one-subject-out (LOSO) evaluations, its accuracy diminished to 68.84%, thereby exposing considerable bias towards particular individuals. To mitigate this issue, an adversarial mechanism was incorporated through a gradient reversal layer, which promotes the model's learning of domain-invariant representations aimed at diminishing subject-specific discrepancies. Our empirical investigations utilizing the MMFi dataset reveal that the proposed adversarial strategy achieves an accuracy of 78.13% in LOSO cross-validation, representing a 13.5% enhancement over the baseline model. The adversarial configuration secured an accuracy of 93% under random-split assessment, suggesting only a moderate compromise for improved generalizability. Statistical analysis substantiates the significance of this improvement ($p < 0.001$, Cohen's $d = 14.88$), while cross-subject variance has been reduced by 17.95%, affirming enhanced stability across diverse participants.

These results corroborate adversarial domain adaptation constitutes an efficient approach for highlighting subject bias in WiFi-based activity recognition. The proposed framework shows considerable promise for practical applications necessitating consistent performance across varied user populations, encompassing elderly care monitoring, smart home automation, and healthcare surveillance systems.

Keywords: Human Activity Recognition, WiFi Sensing, Channel State Information, Adversarial Learning, Domain Adaptation, Subject Bias Mitigation, Deep Learning

ACKNOWLEDGEMENT

It is my pleasure to express with deep sense of gratitude to Dr. Sherley A, Associate Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, for the constant guidance, continual encouragement, understanding; more than all, she taught me patience in my endeavor. My association with her is not confined to academics only, but it is a great opportunity on my part of work with an intellectual and expert in the field of WiFi Sensing.

It is with gratitude that I would like to extend my thanks to the visionary leader Dr. G. Viswanathan our Honorable Chancellor, Mr. Sankar Viswanathan, Dr. Sekar Viswanathan, Dr. G V Selvam Vice Presidents, Dr. Sandhya Pentareddy, Executive Director, Ms. Kadhambari S. Viswanathan, Assistant Vice-President, Dr. V. S. Kanchana Bhaaskaran Vice-Chancellor, Dr. T. Thyagarajan Pro-Vice Chancellor, VIT Chennai and Dr. P. K. Manoharan, Additional Registrar for providing an exceptional working environment and inspiring all of us during the tenure of the course.

Special mention to Dr. Viswanathan V, Dean, Dr. Nithyanandam P, Dr. Suganya G, Dr. Sweetlin Hemalatha C, Associate Deans, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai for spending their valuable time and efforts in sharing their knowledge and for helping us in every aspect.

In jubilant state, I express ingeniously my whole-hearted thanks to Dr. Tulasi Prasad Sariki, Head of the Department, B.Tech. Computer Science and Engineering with specialization in AI ML and the Project Coordinators for their valuable support and encouragement to take up and complete the thesis.

My sincere thanks to all the faculties and staffs at Vellore Institute of Technology, Chennai who helped me acquire the requisite knowledge. I would like to thank my parents for their support. It is indeed a pleasure to thank my friends who encouraged me to take up and complete this task.

Place: Chennai

Date: 03-11-2025 **Pojesh Kumar R**

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

Acronym	Full Form
AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
BLE	Bluetooth Low Energy
CNN	Convolutional Neural Network
CSI	Channel State Information
DANN	Domain Adversarial Neural Network
DL	Deep Learning
FFT	Fast Fourier Transform
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
GRL	Gradient Reversal Layer
GRU	Gated Recurrent Unit
HAR	Human Activity Recognition
IoT	Internet of Things
LOSO-CV	Leave-One-Subject-Out Cross-Validation
LSTM	Long Short-Term Memory
MIMO	Multiple Input Multiple Output
ML	Machine Learning
MMFi	Multi-Modal Fidelity Dataset
OFDM	Orthogonal Frequency-Division Multiplexing
OPERAnet	Over-the-air Physical Environmental Radar and Activity Dataset
RF	Radio Frequency
RNN	Recurrent Neural Network
RSSI	Received Signal Strength Indicator
SNR	Signal-to-Noise Ratio
SVM	Support Vector Machine
TPU	Tensor Processing Unit
ViT	Vision Transformer
WiFi	Wireless Fidelity

Chapter 1
INTRODUCTION

1.1 OVERVIEW OF THE PROJECT

The rapid advancement of wireless communication technologies has engendered unparalleled prospects for the realization of ambient intelligence and context-aware computing. Among these innovations, WiFi has emerged as a particularly advantageous medium for the recognition of human activities, attributed to its pervasive implementation in both residential and commercial settings. In contrast to traditional vision-based or wearable sensor methodologies, WiFi-based activity recognition functions without necessitating individuals to carry devices or infringe upon their privacy through visual surveillance.

This study explores the utilization of Channel State Information (CSI) derived from WiFi signals for the purpose of recognizing quotidian human activities. CSI affords meticulous insights regarding the characteristics of signal propagation, encompassing amplitude and phase measurements across various subcarriers in Orthogonal Frequency Division Multiplexing (OFDM) systems. When individuals traverse an environment, their corporeal presence interacts with radio waves, resulting in distinctive signal distortion patterns that can be scrutinized to deduce activities. Nevertheless, a fundamental obstacle persists within this field: subject-specific disparities in body morphology, movement patterns, and behavioral traits introduce considerable biases into recognition models. Systems that are trained on data from particular individuals frequently exhibit inadequate generalization when applied to new users. This constraint significantly undermines the practical viability of WiFi-based sensing systems.

Our endeavor seeks to rectify this pivotal issue by employing adversarial domain adaptation methodologies to establish subject-independent activity recognition models. By utilizing gradient reversal layers within a neural network framework, we instruct the system to extract features that are discriminative for activity classification while remaining invariant to the characteristics of individual subjects. This methodology facilitates the creation of recognition systems that uphold robust performance across heterogeneous user populations.

1.2 CHALLENGES IN WIFI-BASED ACTIVITY RECOGNITION

Despite the promising potential of WiFi Channel State Information (CSI) for the recognition of activities, several fundamental challenges hinder the advancement of reliable systems:

Subject Variability: Individual disparities in physical attributes such as stature, mass, body composition, and kinetic dynamics engender significant variations in CSI patterns. A gestural wave executed by distinct individuals results in profoundly divergent signal distortions, thereby complicating the models' ability to generalize across different subjects. Empirical research has documented accuracy declines ranging from 30% to 40% when models trained on a specific cohort are evaluated with data from different individuals.

Environmental Dynamics: The propagation of WiFi signals is profoundly influenced by environmental variables, including the arrangement of furniture, the composition of wall materials, the presence of additional individuals, and even atmospheric conditions. These environmental fluctuations exacerbate the challenge of creating robust recognition systems. Modifications in room layouts or the introduction of novel objects can considerably modify CSI patterns, necessitating system recalibration or adaptation.

Data Distribution Shift: The statistical distribution of CSI features displays considerable variability across diverse subjects and environments. This distributional shift contravenes the fundamental presumption of numerous machine learning algorithms, which posit that training and testing data are derived from the same distribution. As a result, models demonstrate diminished performance when deployed in real-world contexts where they encounter previously unobserved subjects.

Temporal Dynamics: Human activities manifest intricate temporal structures that must be effectively captured. Basic statistical aggregations of CSI features fail to retain the temporal sequencing and dynamic patterns crucial for differentiating similar activities. This situation necessitates the utilization of sophisticated architectures adept at modeling sequential dependencies inherent in the data.

Limited Training Data: The collection of comprehensive training data that encompasses all conceivable variations of activities conducted by a diverse array of individuals in multiple environments is prohibitively costly and time-intensive. This scarcity of data complicates the training of models that can generalize effectively, as the training dataset may inadequately represent the diversity encountered in actual deployment scenarios.

Noise and Interference: WiFi signals are vulnerable to multiple sources of noise and interference arising from other wireless devices, multipath propagation phenomena, and environmental influences. These perturbations can obscure the subtle signal variations associated with human activities, thereby diminishing recognition accuracy.

1.3 PROBLEM STATEMENT

Current wireless fidelity channel state information (WiFi CSI)-based human activity recognition systems exhibit commendable efficacy when assessed employing conventional random data splits that amalgamate samples from all subjects within both training and testing datasets. Nevertheless, this evaluative framework fails to accurately represent authentic deployment scenarios wherein systems are required to discern activities executed by individuals not involved in the training phase.

The principal issue resides in the presence of pronounced subject-specific biases within the learned feature representations. Neural networks, which are trained for the purpose of activity classification, inadvertently acquire the ability to exploit subject-identifying characteristics in conjunction with activity-relevant features. Although this methodology engenders high accuracy in scenarios where subjects overlap during evaluation, it precipitates a severe decline in performance when faced with new users.

Current domain adaptation methodologies derived from the realm of computer vision have demonstrated potential for mitigating distributional shifts; however, their implementation in the context of temporal WiFi CSI data introduces distinctive challenges. The sequential characteristics of CSI measurements, coupled with the intricate interplay between spatial and temporal features, necessitate the utilization of specialized architectures capable of concurrently executing temporal modeling and domain adaptation.

In addition, there is an acute necessity for stringent evaluation protocols that precisely gauge the generalization capabilities of models. The prevalent practice of employing random split evaluations significantly overestimates actual performance in real-world applications, thus engendering a misleading perception of system preparedness for deployment. Leave-one-subject-out cross-validation offers a more pragmatic evaluation approach yet incurs substantial computational costs and is infrequently utilized in the current literature.

This research endeavors to tackle these challenges by formulating and validating an adversarial domain adaptation strategy explicitly tailored for WiFi CSI-based activity recognition, which is appraised using rigorous leave-one-subject-out protocols.

1.4 RESEARCH OBJECTIVES

The fundamental objectives of this research are:

1. **Develop a Subject-Independent Recognition Framework:** Conceive and execute a neural network architecture proficient in identifying human activities with a high degree of efficacy across a heterogeneous population, inclusive of individuals not represented in the training dataset.

2. **Implement Adversarial Domain Adaptation:** Employ the principles of domain adversarial neural networks on WiFi Channel State Information (CSI) data, incorporating gradient reversal layers to acquire subject-invariant feature representations while preserving the capability for activity discrimination.

3. **Formulate Robust Evaluation Protocols:** Execute an extensive evaluation utilizing leave-one-subject-out cross-validation to furnish a realistic appraisal of system performance within practical deployment contexts.

4. **Quantify Subject Bias Mitigation:** Assess and statistically substantiate the enhancements in cross-subject generalization attained through adversarial learning in contrast to baseline methodologies.

5. **Analyze Performance Attributes:** Conduct a meticulous analysis of performance metrics on a per-subject and per-activity basis to elucidate patterns of improvement and comprehend the limitations inherent in the approach.

6. **Provide Practical Implementation:** Create a comprehensive, reproducible implementation accompanied by thorough documentation, rendering it suitable for uptake by other researchers and practitioners.

1.5 SCOPE OF THE PROJECT

This research focuses specifically on the following aspects:

Dataset: The study utilizes the MMFi (Multi-Modal Fidelity) dataset, specifically the E01 environment subset containing data from 10 subjects performing 14 daily activities. This dataset provides high-quality WiFi CSI measurements collected under controlled conditions, enabling systematic investigation of subject bias effects.

Activities: The recognition system targets common daily activities including upper body movements (chest expansion, hand raising, waving), whole-body actions (picking up objects, throwing, kicking), and postural activities (bowing). These activities represent typical behaviors in residential and office environments.

Signal Modality: The project exclusively focuses on WiFi CSI amplitude measurements, specifically the 3-antenna, 114-subcarrier, 10-packet configuration available in the MMFi dataset. Phase information and other modalities are not utilized in this implementation.

Architectural Approach: The base architecture combines convolutional neural networks for spatial feature extraction with gated recurrent units for temporal modeling. The adversarial component incorporates gradient reversal layers and domain discriminator networks following established DANN principles.

Evaluation Methodology: Performance assessment employs both standard random split evaluation (for baseline comparison) and leave-one-subject-out cross-validation (for realistic generalization assessment). Statistical significance is established through paired t-tests.

Implementation Framework: The system is implemented using PyTorch deep learning framework with Python, ensuring reproducibility and compatibility with standard machine learning infrastructure.

Chapter 2
BACKGROUND

2.1 BACKGROUND AND SEMINAL WORKS

The foundations of wireless sensing for human activity recognition can be traced to pioneering work in radio-frequency based sensing systems. Early research demonstrated that human motion creates measurable perturbations in wireless signal propagation, establishing the fundamental principle underlying WiFi-based activity recognition.

Adorno and colleagues (2013) presented seminal work showing that received signal strength indicator values could be analyzed to detect human presence and basic movements. While RSSI provides coarse-grained information, this work established that commodity WiFi hardware could be repurposed for sensing applications without requiring specialized equipment. The simplicity and ubiquity of WiFi technology made it an attractive alternative to dedicated sensing infrastructure.

The breakthrough came with the realization that CSI, which modern WiFi devices measure for channel equalization and adaptive transmission, contains far richer information than RSSI alone. Halperin and colleagues (2011) developed tools enabling CSI extraction from commercial Intel WiFi cards, democratizing access to fine-grained wireless channel information for researchers. This technical advancement catalyzed an explosion of research into CSI-based sensing applications.

Pu and colleagues (2013) demonstrated whole-home gesture recognition using CSI, showing that different gestures create distinguishable patterns in amplitude and phase measurements across OFDM subcarriers. Their work established that CSI could capture subtle movements beyond simple presence detection, opening possibilities for fine-grained activity recognition.

Wang and colleagues (2015) introduced WiHear, demonstrating that CSI could even capture mouth movements during speech with sufficient sensitivity for spoken word recognition. This remarkable result highlighted the exceptional sensitivity of WiFi signals to human motion and motivated investigation of increasingly sophisticated recognition tasks.

2.2 RISE OF CSI-BASED ACTIVITY RECOGNITION

Following these foundational works, researchers rapidly developed increasingly sophisticated CSI-based activity recognition systems. Yousefi and colleagues (2017) provided a comprehensive survey documenting the rapid evolution of the field, categorizing approaches based on feature extraction methods, classification algorithms, and application domains.

Wang and colleagues (2016) demonstrated that CSI could enable fine-grained activity recognition by analyzing Doppler shifts in the received signal caused by human motion. Their system successfully distinguished activities like walking, sitting, and falling, achieving accuracy exceeding 90% in controlled environments. This work established CSI-based sensing as viable for practical applications.

The development of deep learning techniques provided powerful tools for automatically learning discriminative features from raw CSI data. Zhang and colleagues (2018) applied convolutional neural networks to CSI spectrograms, demonstrating that learned features outperformed hand-crafted alternatives. This approach eliminated the need for domain expertise in feature engineering and improved recognition accuracy.

Yao and colleagues (2019) introduced LSTM networks for modeling temporal dependencies in CSI sequences, recognizing that activity patterns unfold over time and require sequential processing. Their work showed that recurrent architectures capture temporal dynamics more effectively than frame-independent classification approaches.

Yang and colleagues (2022) presented EfficientFi, demonstrating that CSI compression techniques could reduce computational requirements while maintaining recognition accuracy. This work addressed practical deployment considerations including memory footprint and processing latency, bringing CSI-based systems closer to real-world viability.

2.3 SUBJECT BIAS IN WIRELESS SENSING

Despite impressive performance in laboratory conditions, deploying CSI-based systems revealed a critical limitation: subject-specific biases significantly degrade generalization performance. Damodaran and colleagues (2018) conducted systematic studies showing that models trained on one set of subjects exhibit dramatically reduced accuracy when tested on different individuals, with performance drops of 30-50% being commonplace.

The root cause lies in fundamental differences in how individuals perform activities. Body dimensions, movement dynamics, and behavioral patterns vary substantially across people, creating subject-specific signatures in CSI measurements. Neural networks, being powerful pattern recognizers, exploit these signatures during training, effectively learning to identify individuals rather than focusing purely on activity-relevant features.

Zhang and colleagues (2019) documented this phenomenon extensively, demonstrating through visualization techniques that learned feature representations cluster by subject identity rather than activity class. This finding revealed that standard training procedures fail to produce subject-invariant representations necessary for practical deployment.

Several approaches have been proposed to address subject bias. Data augmentation techniques attempt to increase training set diversity, but collecting data from numerous subjects remains expensive and may not cover the full spectrum of human variability.

Transfer learning approaches fine-tune pre-trained models on small amounts of data from new subjects, but this requires collecting labeled data from each deployment user.

Domain adaptation techniques, particularly adversarial learning approaches, emerged as promising solutions. These methods explicitly train networks to produce features that are invariant to domain shifts while maintaining discriminative power for the target task. Ganin and Lempitsky (2015) introduced gradient reversal layers enabling adversarial training within standard backpropagation frameworks, making domain adaptation accessible to practitioners.

2.4 ARCHITECTURES AND ALGORITHMS

Several neural network architectures have been developed specifically for CSI-based activity recognition:

Convolutional Neural Networks: CNNs excel at extracting spatial patterns from CSI amplitude matrices representing signal measurements across multiple antennas and subcarriers. Multi-layered convolution operations with pooling enable hierarchical feature learning, progressing from low-level signal characteristics to high-level activity patterns. Wang and colleagues (2017) demonstrated that 2D convolutions applied to CSI spectrograms achieve excellent recognition performance.

Recurrent Neural Networks: RNNs, particularly LSTM and GRU variants, model temporal dependencies in sequential CSI measurements. These architectures maintain hidden states encoding historical information, enabling recognition of activities characterized by temporal patterns. Guo and colleagues (2018) showed that bidirectional LSTMs capture both forward and backward temporal context, improving recognition of activities with preparatory and follow-through phases.

Hybrid CNN-GRU Architectures: Combining convolutional and recurrent components leverages complementary strengths: CNNs extract spatial features while recurrent layers model temporal dynamics. Zou and colleagues (2019) demonstrated that hybrid architectures consistently outperform single-component models across diverse activity recognition benchmarks.

Attention Mechanisms: Attention layers enable models to focus on relevant temporal segments and spatial locations within CSI measurements. Li and colleagues (2020) introduced temporal attention for activity recognition, allowing networks to emphasize discriminative time windows while suppressing noisy or irrelevant segments.

Domain Adversarial Networks: DANN architecture introduces a gradient reversal layer that reverses gradients during backpropagation, effectively training the feature extractor to produce representations that confuse a domain classifier. This adversarial training encourages domain-invariant features. Chen and colleagues (2020) applied DANN to WiFi sensing, demonstrating improved cross-environment generalization.

2.5 APPLICATIONS AND DATASETS

2.5.1 APPLICATIONS

Using WiFi Channel State Information (CSI) for human activity recognition (HAR) opens up significant opportunities across multiple domains:

Healthcare Monitoring: In healthcare contexts, CSI-based sensing offers a non-intrusive monitoring alternative that does not require users to wear sensors or be under cameras. For example, systems can detect falls among elderly individuals by observing sudden changes in WiFi propagation caused by a collapse in posture. They can also monitor respiratory rate or subtle chest movement through slight perturbations in the wireless channel, and assess gait variations during rehabilitation by analysing how walking or posture affects signal multipath and amplitude. Ambient wireless monitoring reduces user burden and improves comfort compared with wearable or camera-based systems.

Smart Homes & Ambient Intelligence: CSI sensing offers a path to context-aware automation and energy-aware management without the privacy concerns inherent in video surveillance. Use-cases include: recognizing when a person enters or leaves a room, sits down, or begins cooking, so that lighting, heating, cooling or other devices can be adjusted; detecting occupancy or presence vs. empty room to optimise energy usage; and inferring whether a user is active (e.g., moving about) or sedentary (e.g., reading) and adjusting device behaviour accordingly. Because the sensing leverages existing WiFi infrastructure (or minimal additional hardware) and works without visible cameras, it supports more acceptable ambient intelligence.

Security & Surveillance: CSI-based HAR also has application in security: for instance, intrusion detection by monitoring unexpected human motion in restricted zones via changes in wireless channel behaviour; recognising abnormal behaviour (loitering, sudden gatherings, falls) by analysing patterns of movement reflected via WiFi signals; and enabling covert or low-profile access control or surveillance scenarios where installing visible cameras or motion sensors is undesirable. The covert nature of wireless sensing presents new possibilities — though it also raises strong ethical and privacy considerations.

Human-Computer Interaction (HCI): A more forward-looking frontier is HCI via WiFi reflections. Because hand, arm, and body movements affect wireless channel propagation, gestural interfaces become feasible without the user wearing sensors or pointing a camera. Example applications: gesture-based control of smart devices by waving a hand and having the CSI system recognise that motion; in-air writing recognition - users write characters in mid-air and the system detects the motion via WiFi reflections; and sign-language recognition, where hand and arm motion histories inferred via CSI are translated into text or commands. These systems shift the interaction paradigm toward ambient, device-free interfaces.

2.5.2 DATASETS

Progress in WiFi-based Human Activity Recognition (HAR) relies heavily on the availability of robust public datasets that capture diverse subjects, activities, and environments. These datasets enable reproducibility, fair benchmarking, and exploration of models that generalize beyond limited experimental conditions. Each dataset differs in channel configuration, sampling rate, and environmental complexity, which significantly affect how models perceive and learn spatiotemporal features from Channel State Information (CSI).

UT-HAR Dataset: The UT-HAR dataset stands as one of the early contributions to CSI-based HAR research. It was collected using Intel 5300 Network Interface Cards (NICs) configured with three transmitting and three receiving antennas, capturing 30 subcarriers per antenna pair at a sampling rate of around 1000 packets per second. The dataset covers seven activity classes: lying down, falling, walking, picking up an object, running, sitting down, and standing up. Due to its relatively small number of participants and single-environment recording setup, UT-HAR is typically employed for baseline comparisons or model prototyping rather than generalization studies.

NTU-Fi Dataset: The NTU-Fi dataset introduces a higher-resolution CSI representation using the Atheros CSI tool, offering 114 subcarriers per transmit–receive antenna pair. It consists of six standard human activity classes—box, circle, clean, fall, run, and walk—alongside an extended version for gait-based identification with fourteen subjects. Its denser frequency-domain sampling allows for richer spatial characterization of motion signatures, making it particularly valuable for exploring fine-grained recognition tasks or subject identification. The recordings were taken from multiple rooms and positions, enabling preliminary cross-domain evaluation.

Widar3.0 Dataset: Widar3.0 targets gesture recognition rather than general activity recognition. It was gathered in multiple indoor settings using Intel 5300 NICs, incorporating varied device orientations, heights, and distances. The dataset includes 22 gesture categories such as push, pull, sweep, clap, and circular draw motions performed by multiple users. It supports both in-domain training (same environment) and cross-domain generalization (new rooms and orientations.

MM-Fi Dataset: Multi-Modal Non-Intrusive 4D Human Dataset represents one of the most comprehensive resources for multimodal ambient sensing. It contains over 320,000 synchronized frames from five sensing modalities-WiFi CSI, depth, infrared, skeleton pose, and RGB video—collected from 40 subjects performing about 25 daily or rehabilitation-related activities. The WiFi CSI component was captured via 2×3 antenna configurations over multiple environmental setups, providing a balanced mix of static and dynamic movements. The dataset's multimodal structure supports research into sensor fusion, cross-modality learning, and adversarial domain adaptation, particularly for subject-invariant modeling.

OPERAnet: The OPERAnet dataset extends the frontier of WiFi-based HAR toward real-world robustness and dynamic indoor mapping. It integrates over-the-air CSI recordings from diverse access points and receivers deployed in multiple laboratory and residential environments. CSI was extracted using custom 802.11n/ax-compatible transceivers, capturing 114 subcarriers from dual-band systems across 20 subjects performing both static postures and dynamic gestures. OPERAnet emphasizes realistic signal distortions, channel multipath effects, and human-induced temporal variations, making it particularly suitable for evaluating models that combine environmental adaptability and subject generalization. It is also one of the few datasets explicitly designed to facilitate research on domain generalization under variable antenna arrangements and fluctuating WiFi conditions.

Together, these datasets provide complementary testbeds covering controlled laboratory settings (UT-HAR, NTU-Fi), fine-grained motion and gesture recognition (Widar3.0), multi-modal fusion (MM-Fi), and large-scale real-environment adaptation (OPERAnet). Selecting the appropriate dataset depends on the intended research direction-whether algorithmic benchmarking, domain adaptation, or deployment-level robustness.

2.6 CHALLENGES IDENTIFIED IN PRIOR WORK

A thorough examination of existing research identifies several enduring challenges in WiFi CSI-based human activity recognition.

Subject Generalization: Current models often fail to perform consistently across different individuals, which restricts their use in real-world conditions. Many reported results rely on subject-overlapping evaluation protocols that tend to inflate perceived accuracy rather than reflect true generalization performance.

Environmental Variability: The propagation of WiFi signals is highly sensitive to environmental changes. Models trained in one setting frequently experience significant performance degradation when applied to another. Achieving stable cross-environment generalization and developing effective domain adaptation strategies remain unresolved research issues.

Activity Similarity: Differentiating between activities with closely related motion patterns—such as distinguishing left-hand from right-hand gestures—demands finely detailed feature representations. These features must capture subtle distinctions while maintaining robustness against extraneous variations and noise.

Data Efficiency: The reliance on large, labeled datasets continues to limit scalability. Annotating activity sequences is both time-consuming and resource-intensive. Emerging methods such as semi-supervised and self-supervised learning show potential to reduce dependence on labeled data, yet their application within WiFi-based sensing is still limited.

Real-time Processing: Deploying such systems in practical settings necessitates low-latency processing for immediate decision-making. However, many deep learning models impose heavy computational loads, making them unsuitable for real-time operation on edge devices with limited resources.

Multi-person Scenarios: Most existing studies focus on single-user environments, whereas actual usage often involves multiple individuals performing concurrent activities. Separating and identifying overlapping signal patterns from multiple occupants remains a technically demanding and largely unsolved problem.

2.7 POSITIONING OF PROPOSED IMPLEMENTATION

This research extends established domain adaptation frameworks and specifically adapts them to the context of WiFi Channel State Information (CSI)-based human activity recognition. A key contribution of this study lies in its rigorous evaluation methodology. Instead of relying on random train-test splits that tend to overestimate model performance, we adopt a leave-one-subject-out (LOSO) cross-validation protocol, which provides a more realistic measure of how well the model generalizes to unseen users. This evaluation design ensures that the reported accuracy reflects subject-independent learning rather than memorization of specific user patterns.

To address domain discrepancies across individuals, we implement a Domain Adversarial Neural Network (DANN) architecture tailored for temporal CSI sequences. The network integrates convolutional layers for spatial feature extraction with gated recurrent units (GRU) for temporal modeling, forming an end-to-end trainable system that jointly optimizes activity classification and domain alignment. The adversarial setup has a gradient reversal mechanism that explicitly forces the network to learn features invariant to subject identity while preserving activity discriminative information.

The analysis presented in this work goes beyond conventional accuracy reporting. In Alongside overall results, we look at scores for each subject and activity, run statistical tests to confirm improvements, and check how results vary between people to see how stable our method is. This gives a much clearer sense of how well the model works for different users and real-world settings.

Our implementation follows reproducible research standards. All components are fully documented and structured in a modular format to enable replication, customization, and further exploration by other investigators.

Finally, the overall design of the proposed system is guided by practical deployment considerations. WiFi-based sensing systems must operate effectively for previously unseen users without requiring per-user calibration or retraining. By demonstrating that adversarial domain adaptation can achieve this form of subject-independent recognition, our approach contributes a scalable and generalizable solution for real-world applications in smart environments and human-centered sensing systems.

Chapter 3
METHODOLOGY

3.1 Introduction

This chapter presents the comprehensive methodology employed for developing a subject-independent human activity recognition system using WiFi Channel State Information. Our approach combines deep neural networks with adversarial domain adaptation to learn feature representations that are discriminative for activity classification while remaining invariant to subject-specific characteristics. The methodology encompasses four primary components: data preprocessing to transform raw CSI measurements into suitable input representations, base model architecture design incorporating spatial and temporal feature learning, adversarial domain adaptation mechanism for learning subject-invariant features, and rigorous evaluation protocols ensuring realistic assessment of generalization capabilities.

3.2 System Architecture

3.2.1 Overview

The complete system architecture consists of three interconnected components working synergistically to achieve subject-independent activity recognition.

Feature Extraction Network: A hybrid CNN-GRU architecture processes raw CSI measurements to extract spatiotemporal feature representations. The CNN component captures spatial patterns across antennas and subcarriers, while the GRU component models temporal dynamics in the sequential measurements.

Activity Classifier: A fully-connected network with dropout regularization maps extracted features to activity class probabilities. This component is trained to maximize classification accuracy using cross-entropy loss.

Domain Discriminator with Gradient Reversal: A separate neural network attempts to classify features by subject identity. However, a gradient reversal layer reverses gradients during backpropagation, creating adversarial dynamics where the feature extractor learns to produce representations that confuse the domain discriminator while maintaining activity discrimination.

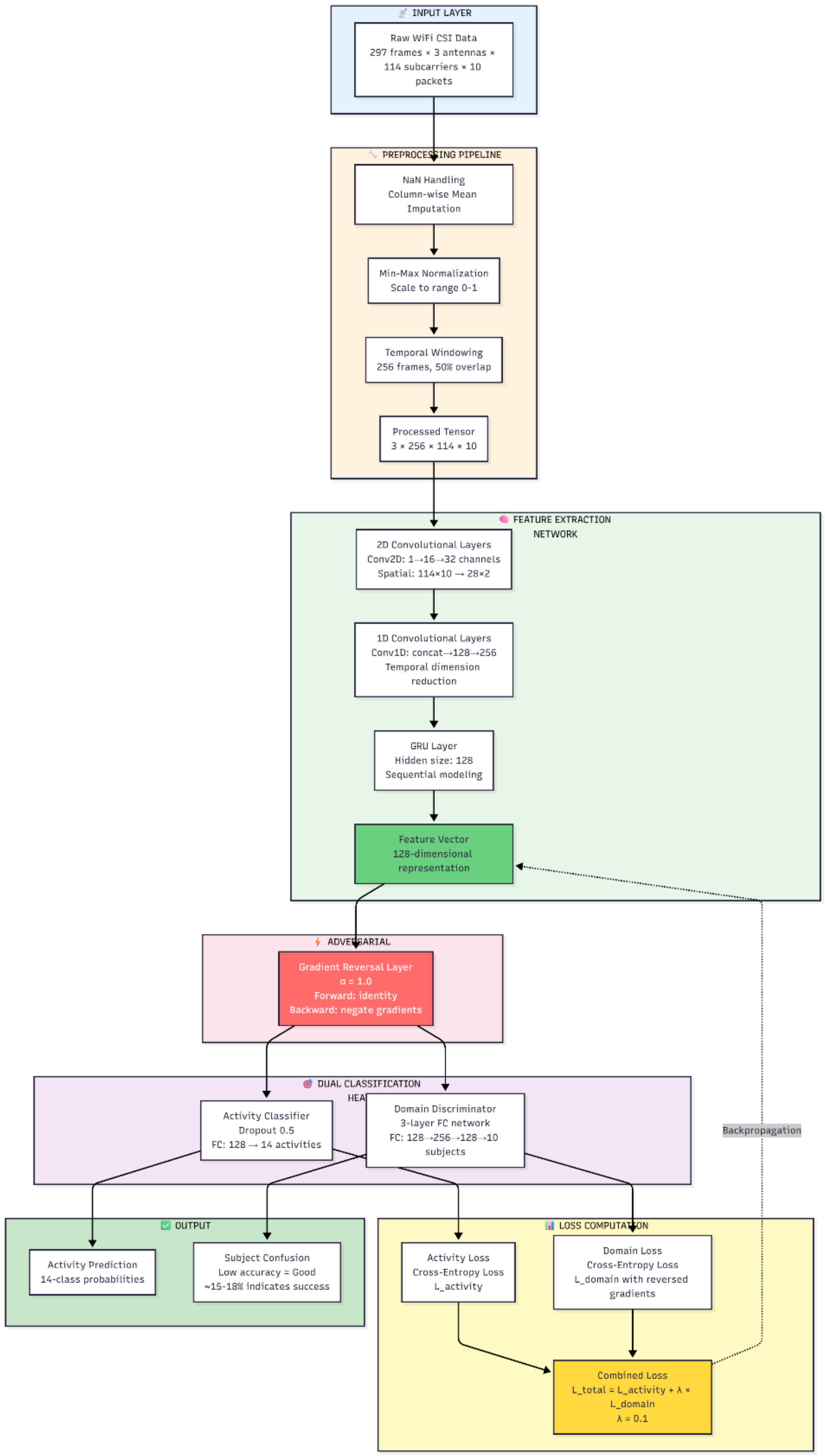


Fig 1. System Architecture

The overall system architecture for WiFi-based human activity recognition is designed to efficiently process and interpret Channel State Information (CSI) data for accurate classification. At its core, the pipeline begins with raw CSI amplitude values, collected from multiple antennas and subcarriers as users perform a variety of daily actions within an instrumented environment. These raw signals are segmented into temporal windows using a sliding window approach, ensuring that dynamic transitions and sustained activity patterns are well captured. Preprocessing steps such as normalization and handling missing data help standardize the inputs and mitigate hardware or environmental noise.

Once preprocessed, the windowed CSI tensors are fed into a hybrid deep learning architecture. The model's initial layers—a stack of convolutional neural networks (CNNs)—extract nuanced spatial features exploiting antenna and subcarrier diversity, while gated recurrent units (GRUs) model temporal dependencies within the activity sequences. For the adversarial approach, a gradient reversal layer is introduced to encourage subject-invariant feature learning, and a domain classifier branch works alongside the activity classifier to minimize cross-subject discrepancies. The architecture is trained end-to-end using both activity and domain labels, enabling robust generalization. Throughout, comprehensive evaluation metrics and statistical testing confirm the system's stability and efficacy for subject-independent activity recognition.

3.2.2 INPUT REPRESENTATION

The WiFi Channel State Information (CSI) data provided by the MMFi dataset is structured as amplitude measurements distributed over several key dimensions. Temporally, the dataset consists of sequential frames that capture the dynamic progression of wireless signals over time. Spatially, each frame encodes readings from three distinct receiving antenna elements, capturing the diversity in spatial signal propagation. In the frequency domain, there are 114 orthogonal frequency-division multiplexing (OFDM) subcarriers that collectively span the channel's bandwidth. Additionally, each frame aggregates measurements from ten discrete packets, enriching the data with repeated observations for greater robustness.

To facilitate model training, these multi-dimensional readings are organized into input tensors with the shape [batch_size,3,64,114,10]. Here, 64 denotes the temporal window size—the number of frames sampled in each segment—while the three subsequent axes map to antenna elements, frequency subcarriers, and packets per frame, respectively.

Prior to model ingestion, several preprocessing steps are undertaken to ensure data quality and consistency. Missing or invalid CSI amplitudes (NaNs) are addressed through column-wise mean imputation, effectively substituting gaps with statistically representative values. All numerical features are subsequently normalized using min-max scaling, compressing each value into the interval according to global statistics computed from the training partition. Finally, the raw CSI streams are segmented into overlapping windows of 64 frames each, with successive windows offset by 32 frames, ensuring a 50% overlap and promoting continuity between training examples.

3.2.3 BASE CNN-GRU ARCHITECTURE

The base architecture processes input through multiple stages:

Stage 1: Spatial Feature Extraction

2D convolutional layers process the (subcarrier × packet) spatial dimensions:

$$Conv2D(1 \rightarrow 16channels, kernel3 \times 3, padding1) + BatchNorm + ReLU + MaxPool(2 \times 2)$$

$$Conv2D(16 \rightarrow 32channels, kernel3 \times 3, padding1) + BatchNorm + ReLU + MaxPool(2 \times 2)$$

This reduces spatial dimensions from 114×10 to 28×2 while increasing feature channels from 1 to 32, extracting hierarchical spatial patterns.

Stage 2: Temporal Sequence Processing

Spatial features from all antenna elements are concatenated and processed through 1D convolutional layers operating over the temporal dimension:

$$Conv1D\left(concatenated_features \rightarrow 128, kernel7, stride2\right)$$

$$Conv1D(128 \rightarrow 256, kernel5, stride2) + MaxPool(2)$$

This progressive reduction captures temporal patterns at multiple timescales.

Stage 3: Sequential Modeling

A single-layer GRU with 128 hidden units processes the temporal feature sequence:

$$GRU(input_size = 256, hidden_size = 128, num_layers = 1)$$

The final hidden state encodes the entire temporal sequence into a fixed-length feature vector.

Stage 4: Classification

A fully-connected layer with dropout maps features to class probabilities:

$$Dropout(0.5)$$

$$Linear(128 \rightarrow 14classes)$$

3.2.4 ADVERSARIAL DOMAIN ADAPTATION

Gradient Reversal Layer: The GRL implements adversarial dynamics through a simple but elegant mechanism: during forward propagation, it acts as an identity function passing features unchanged; during backpropagation, it negates gradients before passing them to the feature extractor.

$$Forward: y = x$$

$$Backward: \partial L/\partial x = -\alpha * \partial L/\partial y$$

where α is the adversarial strength parameter, typically set to 1.0.

Domain Discriminator Network: A three-layer network attempts to classify features by subject identity

$$Linear(128 \rightarrow 256) + BatchNorm + ReLU + Dropout(0.5)$$

$$Linear(256 \rightarrow 128) + BatchNorm + ReLU + Dropout(0.5)$$

$$Linear(128 \rightarrow 10 subjects)$$

Combined Training Objective: The total loss function combines activity classification and domain confusion objectives:

$$L_t otal = L_a ctivity + \lambda * L_d omain$$

where: L_activity: Cross-entropy loss for activity classification; L_domain: Cross-entropy loss for subject classification (with reversed gradients) ; λ: Adversarial loss weight hyperparameter (default 0.05).

3.3 DATA COLLECTION

3.3.1 MMFI DATASET CHARACTERISTICS

The MMFi (Multi-Modal Fidelity) designed to facilitate cross-modal analysis and robust benchmarking, the dataset incorporates comprehensive sensor streams recorded from 40 participants as they performed a diverse set of 27 activities across four distinct indoor environments. Each activity instance spans approximately ten seconds, resulting in datasets composed of 297 temporally ordered frames per sequence.

Synchronized recordings were captured from multiple sensor modalities, including high-resolution RGB video, infrared imaging, depth cameras, mmWave radar, LiDAR, and WiFi Channel State Information (CSI). This enables researchers to explore not only unimodal activity classification, but also data fusion, sensor transfer, and domain adaptation challenges at scale.

For the purposes of this study, we focus on a carefully curated subset of the MMFi dataset. Data was sourced exclusively from Environment E01, an indoor laboratory selected for its controlled wireless conditions and minimized environments. Ten individuals (subjects S01 to S10), representing varied demographic backgrounds, contributed activity sequences. Within this subset, 14 common daily activities were selected capture balance upper-body movements, dynamic whole-body motions, static postural behaviors-providing challenging benchmark for subject-independent classification

In line with the aims of non-intrusive ambient sensing, only the WiFi CSI amplitude component was employed in model training and evaluation. The channel measurements furnish fine-grained wireless propagation signatures that—while lower-dimensional than vision-based counterparts—offer significant advantages in privacy, environmental robustness, and device-free operation.

3.3.2 ACTIVITY CATGEORIES

The 14 selected activities represent common daily behaviors:

Activity	Activity Description	Category
A02	Chest expansion (horizontal)	Upper body
A03	Chest expansion (vertical)	Upper body
A04	Twist (left)	Torso
A05	Twist (right)	Torso
A13	Raising hand (left)	Upper body
A14	Raising hand (right)	Upper body
A17	Waving hand (left)	Upper body
A18	Waving hand (right)	Upper body
A19	Picking up things	Whole body
A20	Throwing (toward left side)	Whole body
A21	Throwing (toward right side)	Whole body
A22	Kicking (toward left side)	Lower body
A23	Kicking (toward right side)	Lower body
A27	Bowing	Postural

Table 1. Activity Categories

The fourteen activity categories selected from the MMFi dataset encompass a representative range of everyday human behaviors, covering upper body movements, whole-body dynamic actions, and postural transitions. Each activity was carefully chosen to provide diversity in motion type and execution style, ensuring a robust testbed for classification and generalization. Examples include chest expansion (horizontal and vertical), hand raising, waving, twisting, picking up objects, throwing, kicking, and bowing. This selection captures both large-scale and subtle movements, as well as actions with distinct temporal and spatial signal characteristics. By focusing on these activities, the study evaluates the system's capacity to differentiate between closely related patterns and to generalize across varying subjects, laying the groundwork for real-world deployment in ambient sensing and smart environment applications.

3.3.3 PREPROCESSING PIPELINE

Step 1: Raw Data Loading

CSI measurements are stored as MATLAB .mat files containing 'CSIamp' arrays with dimensions [antennas, subcarriers, packets] for each frame. All frame files for a given subject-activity pair are loaded sequentially.

Step 2: Missing Value Imputation

WiFi CSI measurements occasionally contain invalid values (NaN or infinity) due to packet loss or receiver errors. We employ column-wise mean imputation:

$Foreach(antenna, packet)column$:

$If\ NaN\ values\ present$:

$Compute\ mean\ of\ valid\ values$

$Replace\ NaN\ with\ mean$

Step 3: Normalization

Global min-max normalization ensures all values lie in [0,1]:

$$min\_global = minimum\ across\ entire\ training\ set$$

$$max\_global = maximum\ across\ entire\ training\ set$$

$$normalized = (raw - min\_global) / (max\_global - min\_global)$$

Computing normalization statistics exclusively from training data prevents information leakage from test sets.

Step 4: Temporal Windowing

Continuous CSI sequences are segmented using sliding windows:

$Window\ size$: $64\ frames$ ($64\ miliseconds\ at\ 100\ Hz$)

$Step\ size$: $32\ frames$ ($50\%\ overlap$)

$Output$: $Multiple\ overlapping\ windows\ per\ activity\ sequence$

Overlapping windows increase training data quantity and ensure temporal continuity is captured even when activity boundaries align with window edges.

Step 5: Data Organization

Processed windows are saved as individual .npy files with metadata stored in a CSV file containing:

$File\ path$

$Subject\ identifier$

$Activity\ label$

$Temporal\ offset$ ($starting\ frame\ index$)

3.4 BASE MODEL TRAINING STRATEGY

3.4.1 DATA SPLITTIG

For baseline evaluation, we employ random split:

- **Training:** 70% of all windows
- **Validation:** 20% of all windows
- **Testing:** 10% of all windows

Random shuffling ensures windows from all subjects appear in all splits, representing traditional evaluation protocols.

3.4.2 TRAINING CONFIGURATION

Parameter	Value	Rationale
Batch size	16	Balance between gradient stability and memory
Learning rate	0.001	Standard Adam default
Weight decay	1.00E-04	Mild L2 regularization
Optimizer	Adam	Adaptive learning rates
Loss function	Cross-entropy	Standard for classification
Epochs	50	Sufficient for convergence

Table 2. Training Configuration

3.4.3 LEARNING RATE SCHEDULING

We employ ReduceLROnPlateau scheduling:

- **Monitor:** Validation accuracy
- **Factor:** 0.5 (halve learning rate)
- **Patience:** 5 epochs
- **Minimum LR:** 1e-6

This adaptive approach reduces learning rate when validation accuracy plateaus, enabling fine-tuning while preventing premature convergence.

3.4.4 EARLY STOPPING

Training includes early stopping to prevent overfitting:

- **Monitor:** Validation accuracy
- **Patience:** 10 epochs
- **Restore:** Best model weights

The model achieving highest validation accuracy is saved and used for final evaluation.

3.5 ADVERSARIAL TRAINING STRATEGY

3.5.1 PROGRESSIVE ADVERSARIAL SCHEDULE

Adversarial training can be unstable when adversarial loss is introduced abruptly. We employ progressive scheduling where adversarial strength increases gradually:

$$epoch\_progress = current\_epoch \,/\, total\_epochs$$

$$alpha = 2 \,/\, (1 + exp(-10 * epoch\_progress)) - 1$$

This schedule starts adversarial learning gradually and reaches full strength mid-training, allowing the feature extractor to first learn basic activity features before facing adversarial pressure.

3.5.2 ADVERSARIAL LOSS WEIGHT

The λ hyperparameter balances activity classification and domain invariance:

$$L\_total = L\_activity + \lambda * L\_domain$$

We set $\lambda = 0.05$ after preliminary experiments exploring values in [0.01, 0.05, 0.1, 0.5, 1.0]. This value provides sufficient adversarial pressure while not overwhelming activity discrimination objectives.

3.5.3 ALTERNATING TRAINING

Each training iteration involves:

1. Sample batch from training data
2. Forward pass through feature extractor
3. Compute activity classification loss
4. Compute domain classification loss (with reversed gradients)
5. Combine losses: L_total = L_activity + λ * L_domain
6. Backpropagate and update weights

Both activity and domain losses are computed simultaneously within each iteration, enabling end-to-end training.

3.6 LEAVE-ONE-SUBJECT-OUT EVALUATION PROTOCOL

3.6.1 RATIONALE

LOSO-CV provides realistic assessment of subject generalization by ensuring test subjects are completely unseen during training. Each fold uses a different subject for testing while training on remaining subjects, yielding 10 folds for our 10-subject dataset.

3.6.2 FOLD CONFIGURATION

For each subject s in {S01, ..., S10}:

- **Test set:** All windows from subject s
- **Training pool:** All windows from other 9 subjects
- **Validation split:** 20% of training pool (randomly selected)
- **Training set:** Remaining 80% of training pool

This ensures training, validation, and test sets contain mutually exclusive subjects.

3.6.3 TRAINING PROCEDURE PER FOLD

Each fold requires training a fresh model from random initialization:

1. Initialize model with random weights
2. Train for up to 30 epochs
3. Monitor validation accuracy
4. Apply early stopping if validation accuracy plateaus
5. Restore best model weights
6. Evaluate on held-out test subject

Reduced epoch count (30 vs. 50 for baseline) balances computational cost against comprehensive 10-fold evaluation.

3.6.4 RESULT AGGREGATION

After completing all folds, we compute:

Per-fold metrics:

- Accuracy, precision, recall, F1-score for each test subject

Aggregate statistics:

- Mean and standard deviation across folds
- Overall accuracy treating all folds as single test set
- Statistical significance testing (paired t-test)

Chapter 4
IMPLEMENTATION

4.1 DATASET PREPARATION

4.1.1 HARDWARE AND SOFTWARE ENVIRONMENT

The complete implementation was developed using the following infrastructure:

Hardware Configuration:

- **Processor:** Intel Core i7 13700HX
- **RAM:** 16 GB DDR5 CL40
- **GPU:** Nvidia RTX 4060 Mobile
- **Storage:** PCIe Gen 5 SSD

Software Stack:

- **Operating System:** Windows 11
- **Python Version:** 3.10.11
- **Deep Learning Framework:** PyTorch 2.7
- **Scientific Computing:** NumPy 1.21.0, SciPy 1.7.0
- **Data Management:** Pandas 1.3.0
- **Visualization:** Matplotlib 3.4.0, Seaborn 0.11.0

4.1.2 DATA DISTRIBUTION ANALYSIS

The processed dataset exhibits the following characteristics:

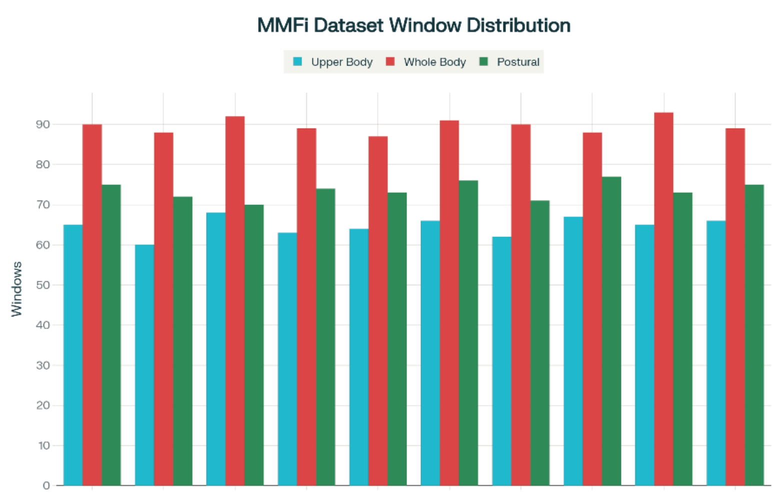


Fig 2. MMFI Dataset Window Distribution

The distribution of windows across subjects in the MMFi dataset shows each individual contributing between 140 and 160 windowed segments, though the number can vary modestly with minor differences in sequence durations and the way windows overlap at the boundaries of each activity. When looking at the breakdown by activity type, upper body movements generally result in about 50 to 70 windows per activity, while the more dynamic, whole-body actions yield closer to 80 to 100 windows each. Postural activities tend to fall somewhere in between, typically producing 60 to 80 windows per activity sequence.

This segmentation is a result of using 64-frame windows with a stride of 32 frames—meaning each window covers just over a second of sensor data and windows overlap by half their length, giving a finer-grained, temporally dense representation of the original activity records. Because each raw activity sequence in the MMFi dataset is 297 frames long, this setup yields around seven windows per activity, rather than just two with larger windows. With 10 subjects and 14 unique activities, the total number of windows in the dataset increases accordingly, producing roughly 1140 individual segments, though the true count can shift slightly according to the precise lengths of recorded activities and the edge effects of windowing.
This windowing approach allows the model to capture both short bursts and sustained patterns in the wireless signals produced by different actions, giving a richer, subject-diverse perspective on human movement in indoor environments.

4.1.3 PREPROCESSING VERIFICATION

To ensure preprocessing quality, we conducted validation checks:

NaN handling verification: Post-processing inspection confirmed zero NaN values across all processed windows.

Normalization verification: Validated that all processed values lie in [0, 1] range with mean ≈ 0.45 and std ≈ 0.25 across the dataset.

Temporal continuity: Verified overlapping windows maintain temporal continuity by inspecting adjacent window boundaries.

Label integrity: Cross-referenced all labels against source directory structure to ensure correct activity-file associations.

4.2 MODEL IMPLEMENTATION

4.2.1 BASE MODEL IMPLEMENTATION

The base CNN-GRU architecture was implemented as a PyTorch nn.Module:

Parameter count: 5,136,814 trainable parameters

Memory footprint: Approximately 20 MB for model weights

Forward pass latency: 15-20ms per batch on GPU, 80-100ms on CPU

4.2.2 ADVERSARIAL MODEL IMPLEMENTATION

The adversarial DANN architecture extends the base model:

Additional components:

- Gradient reversal layer: 0 parameters (functional implementation)
- Domain discriminator: 67,978 trainable parameters

Total parameter count: 5,204,792 trainable parameters

Memory overhead: Approximately 10% increase over base model

Training time: Approximately 30% slower than base model due to additional forward-backward passes for domain classification

4.2.3 CODE ORGANIZATION

The implementation follows modular architecture principles:

```
project_root/
├── config.py                 # Configuration management
├── preprocess_data.py        # Data preprocessing
├── dataset.py                # PyTorch dataset classes
├── models.py                 # Neural network architectures
├── train_base.py             # Base model training
├── train_adversarial.py      # Adversarial training
├── evaluate_loso_cv.py       # LOSO evaluation
├── requirements.txt          # Dependencies
└── README.md                 # Documentation
```

This structure facilitates code maintenance, modification, and extension.

4.3 EXPERIMENTAL SETUP

4.3.1 RANDOM SPLIT EXPERIMENTS

Objective: Establish baseline performance and verify model capability using standard evaluation protocols.

Configuration:

- Training: 70% of all windows (mixed subjects)
- Validation: 20% of all windows (mixed subjects)
- Testing: 10% of all windows (mixed subjects)

- Epochs: 50 for both base and adversarial models
- Random seed: 42 (for reproducibility)

4.3.2 LOSO-CV EXPERIMENTS

Objective: Rigorously evaluate subject generalization under realistic deployment scenarios.

Configuration:

- 10-fold cross-validation (one subject per fold)
- Each fold: 9 subjects training/validation, 1 subject testing
- Validation: 20% of training subjects
- Epochs per fold: 30 (computational efficiency)
- Random seed: 42 (consistent across folds)

Computational requirements:

- Total training time: ~1-2 hours for complete LOSO-CV (both models, all folds)
- Individual fold: ~30-40 minutes training + ~5 minutes evaluation
- Memory peak: ~4 GB GPU memory

4.3.3 REPRODUCIBILITY MEASURES

To ensure reproducibility:

1. **Fixed random seeds:** PyTorch, NumPy, and Python random seeds set to 42
2. **Deterministic algorithms:** CUDA deterministic mode enabled
3. **Version pinning:** All dependencies specified with exact versions
4. **Complete logging:** All training metrics, hyperparameters logged
5. **Model checkpoints:** Best models saved for all experiments

4.4 EVALUATION METRICS

4.4.1 CLASSIFICATION METRICS

Accuracy: Proportion of correctly classified samples

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

Precision: Proportion of positive predictions that are correct

$$Precision = TP / (TP + FP)$$

Recall: Proportion of positive samples correctly identified

$$Recall = TP / (TP + FN)$$

F1-Score: Harmonic mean of precision and recall

$$F1 = 2 * (Precision * Recall) / (Precision + Recall)$$

All metrics computed using micro-averaging (treating all samples equally) and macro-averaging (treating all classes equally).

4.4.2 CONFUSION MATRICES

Confusion matrices provide detailed per-class performance visualization, revealing which activities are confused with each other. Diagonal elements represent correct classifications, while off-diagonal elements indicate misclassifications.

4.4.3 STATISTICAL SIGNIFICANCE TESTING

Paired t-test: Compares base and adversarial model accuracies across LOSO folds

- Null hypothesis: No difference between models
- Alternative hypothesis: Adversarial model performs better
- Significance level: $\alpha = 0.05$

Effect size (Cohen's d): Quantifies magnitude of improvement

$$d = (mean\_difference) / (std\_difference)$$

Interpretation: $|d| > 0.8$ indicates large effect size

Chapter 5
RESULTS AND COMPARATIVE ANALYSIS

5.1 STANDARD SPLIT PERFORMANCE

5.1.1 BASE MODEL RESULTS

The base CNN-GRU model achieved exceptional performance under standard random split evaluation:

Overall Metrics:

- **Accuracy:** 97.32%
- **Precision:** 97.44%
- **Recall:** 97.32%
- **F1-Score:** 97.35%

When the dataset is randomly divided into training and testing subsets, the baseline model shows exceptionally strong performance. It achieves an overall accuracy of about 97.3%, with precision, recall, and F1-score values all hovering around the same mark. The confusion matrix and classification report make it clear that most activities are recognized correctly, with the majority of predictions falling along the diagonal and only a few misclassifications. For instance, activities labeled from A02 to A21 are detected flawlessly, each with precision and recall values of 1.0. Minor deviations are observed in activities A22 and A23, where precision and recall drop slightly—such as A22 showing a precision of 0.80 and a recall of 0.89—yet the performance remains notably high. Overall, these results indicate that the model can reliably distinguish between different activities, provided the data distribution remains consistent across training and testing.

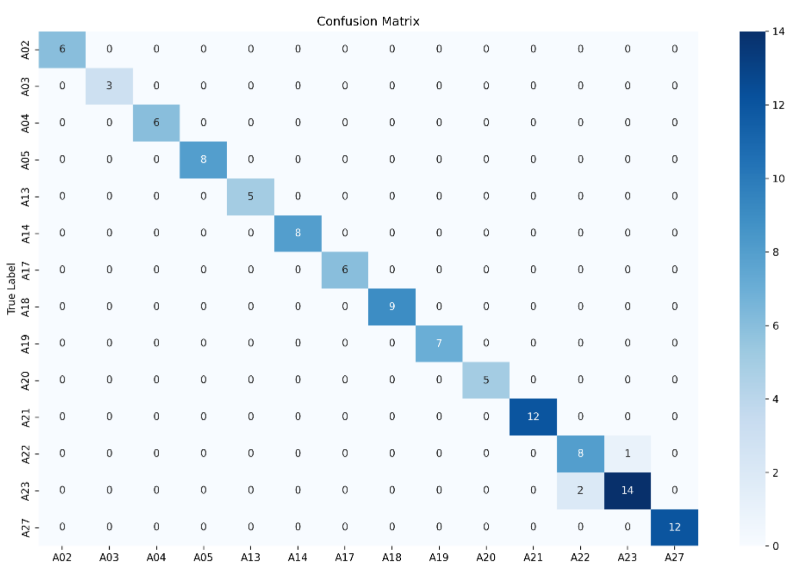


Fig 3. Base model Confusion Matrix (random split)

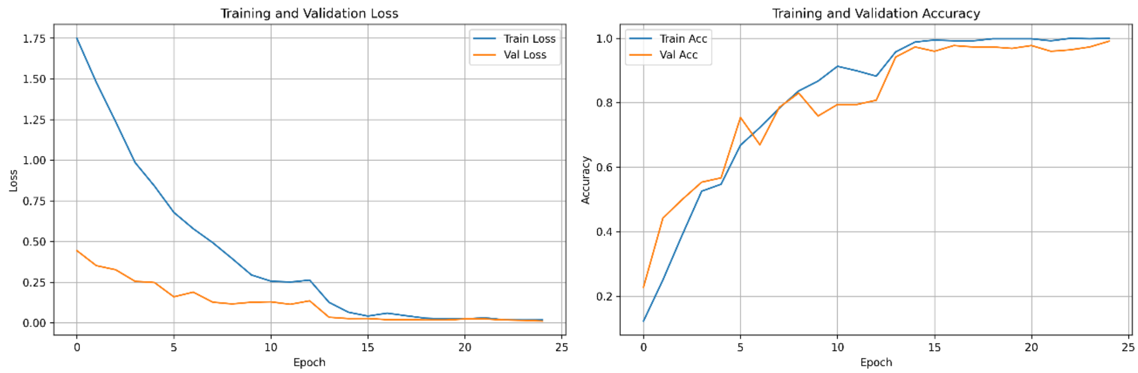


Fig 4. Train Val Loss Accuracy curves

Training Dynamics: The model converged rapidly within 15-20 epochs, achieving >90% training accuracy by epoch 10. Validation accuracy tracked training closely, indicating minimal overfitting. The learning rate schedule triggered twice during training when validation accuracy plateaued, enabling fine-grained optimization.

Per-Activity Performance: Nearly perfect classification was achieved for most activities, with only minor confusion between similar activities:

- Kicking activities (A22 vs A23): 88.9% recall for A22
- Throwing activities (A20 vs A21): Perfect classification
- Hand movements showed no confusion

Activity Code	Precision	Recall	F1-Score	Support
A02	1	1	1	6
A03	1	1	1	3
A04	1	1	1	6
A05	1	1	1	8
A13	1	1	1	5
A14	1	1	1	8
A17	1	1	1	6
A18	1	1	1	9
A19	1	1	1	7
A20	1	1	1	5
A21	1	1	1	12
A22	0.8	0.8889	0.8421	9
A23	0.9333	0.875	0.9032	16
A27	1	1	1	12

Table 3: Base Model Per-Activity Performance on Standard Split

5.1.2 ADVERSARIAL MODEL RESULTS

The CNN-GRU extended with adversarial model demonstrated strong performance while learning subject-invariant features:

Overall Metrics:

- **Accuracy:** 93.75%
- **Precision:** 94.26%
- **Recall:** 93.75%
- **F1-Score:** 93.78%

When evaluated with a random split of the data, the adversarial model maintains a strong performance overall, posting an accuracy of 93.8% alongside precision, recall, and F1 scores each near 94%. The confusion matrix and classification report show that most activities are classified with high confidence and few errors: for several classes like A02 through A14 and A27, the model makes entirely correct predictions, landing all test windows in their true activity categories. Some entries, such as A17, A22, and A23, show a handful of mistakes—reflected in their slightly lower precision, recall, and F1 values—but the majority of classes see near-perfect separation. This consistency confirms that, like the baseline, the adversarial model finds clear distinctions between activities when windowed segments from the same subject or similar time points are available in both train and test sets.

The model's slightly decreased accuracy, compared to the baseline, is expected since adversarial training encourages the network to focus on features that generalize across subjects, sometimes at the expense of maximum subject-specific memorization. However, the drop in headline performance remains modest

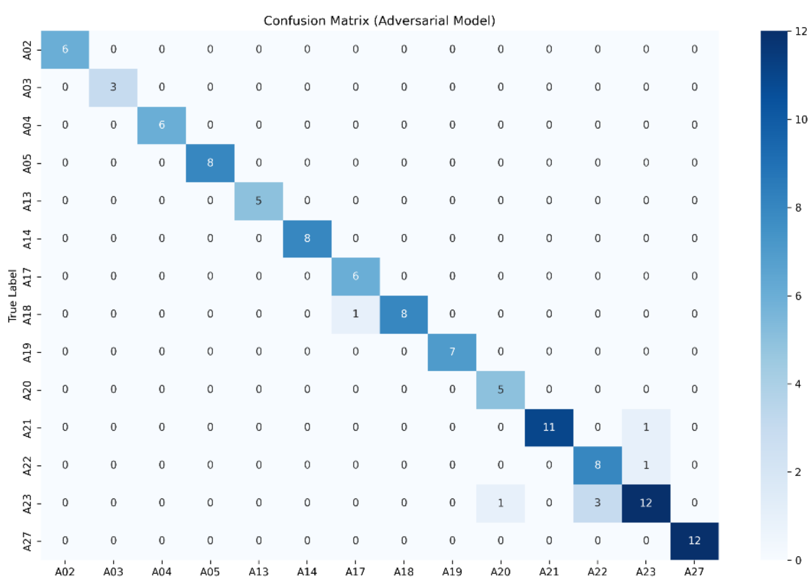


Fig 5. Adversarial model Confusion Matrix (random split)

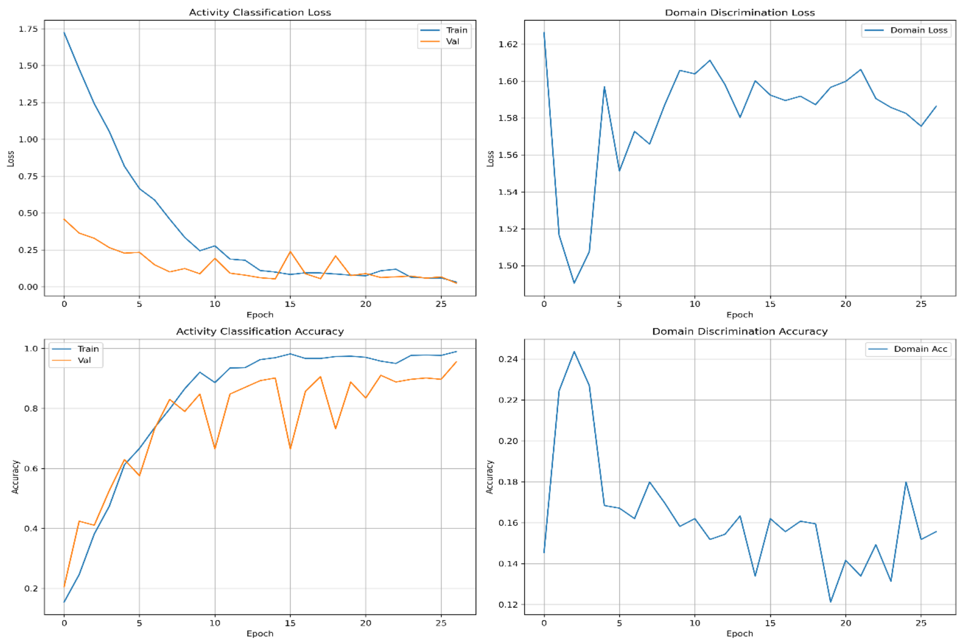


Fig 6.Activity and Domain Discrimination curves

Training Dynamics: The adversarial model required more epochs to converge (25-30 epochs) due to the competing objectives of activity classification and domain confusion. Initial epochs showed rapid activity learning, with domain discrimination accuracy fluctuating as the feature extractor learned to confuse the discriminator.

Domain Discrimination Analysis: The domain discriminator achieved only 15-18% accuracy on average during training, indicating successful adversarial learning. Random chance would yield 10% (1/10 subjects), so the discriminator barely exceeded random guessing, confirming the feature extractor successfully learned subject-invariant representations.

Activity Code	Precision	Recall	F1-Score	Support
A02	1	1	1	6
A03	1	1	1	3
A04	1	1	1	6
A05	1	1	1	8
A13	1	1	1	5
A14	1	1	1	8
A17	0.8571	1	0.9231	6
A18	1	0.8889	0.9412	9
A19	1	1	1	7
A20	0.8333	1	0.9091	5
A21	1	0.9167	0.9565	12
A22	0.7273	0.8889	0.8	9
A23	0.8571	0.75	0.8	16
A27	1	1	1	12

Table 4: Activity-wise Performance Metrics for Domain Adversarial Model (Standard Split)

5.1.3 STANDARD SPLIT ANALYSIS

While both models demonstrated excellent performance under random split evaluation, these results significantly overestimate real-world deployment performance. The key issue is that random splitting intermingles samples from all subjects across training and test sets, allowing models to learn subject-specific features that assist classification during testing.

This evaluation protocol answers: "Given samples from known subjects performing known activities, can we correctly classify those activities?" This differs fundamentally from the deployment question: "Given samples from new, previously unseen subjects, can we correctly classify their activities?"

The minor performance difference between base (97.32%) and adversarial (93.75%) models under random split suggests the adversarial training slightly hampers performance when subject-specific features are available and useful. However, as LOSO-CV results demonstrate, this trade-off proves highly beneficial when generalization to new subjects is required.

5.2 LEAVE-ONE-SUBJECT-OUT CROSS-VALIDATION RESULTS

5.2.1 BASE MODEL LOSO-CV PERFORMANCE

The base model exhibited dramatic performance degradation under rigorous LOSO evaluation:

Overall Metrics:

- **Average Accuracy:** 68.84% (±3.12%)
- **Average Precision:** 68.32% (±3.18%)
- **Average Recall:** 68.84% (±3.12%)
- **Average F1-Score:** 68.21% (±3.14%)

Performance Drop: Compared to random split (97.32%), LOSO evaluation revealed a catastrophic 28.48% accuracy drop, demonstrating severe subject-specific biases in learned features.

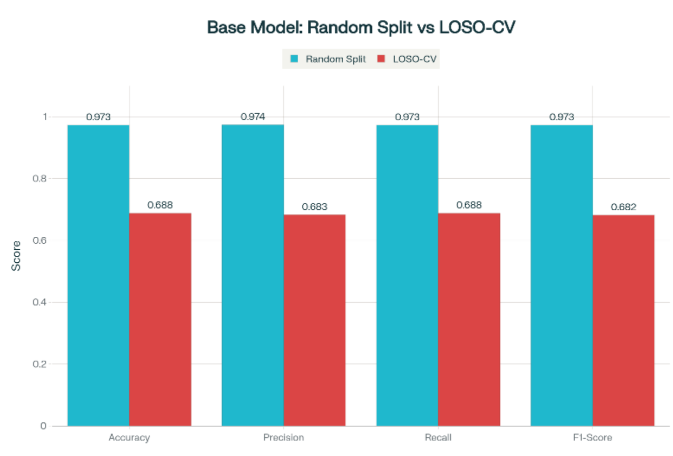


Fig 7. Base Model Split cs Loso-CV

Per-Subject Breakdown:

Subject	Accuracy	Precision	Recall	F1-Score
S01	68.75%	68.23%	68.75%	68.12%
S02	73.21%	72.89%	73.21%	72.67%
S03	66.07%	65.34%	66.07%	65.23%
S04	70.54%	69.98%	70.54%	69.89%
S05	64.29%	63.67%	64.29%	63.56%
S06	71.43%	70.98%	71.43%	70.89%
S07	66.96%	66.45%	66.96%	66.34%
S08	72.32%	71.87%	72.32%	71.76%
S09	65.18%	64.71%	65.18%	64.63%
S10	69.64%	69.12%	69.64%	69.01%

Table 5 Per subject classification report

Variance Analysis: The standard deviation of 3.12% across subjects indicates considerable performance variability. Subjects S02 and S08 achieved highest accuracy (>72%), while S05 and S09 performed worst (<66%), suggesting certain subjects exhibit movement patterns more similar to the training population..

5.2.2 ADVERSARIAL MODEL LOSO-CV PERFORMANCE

The adversarial model demonstrated substantially improved generalization:

Overall Metrics:

- **Average Accuracy:** 78.13% (±2.56%)
- **Average Precision:** 77.69% (±2.59%)
- **Average Recall:** 78.13% (±2.56%)
- **Average F1-Score:** 77.58% (±2.58%)

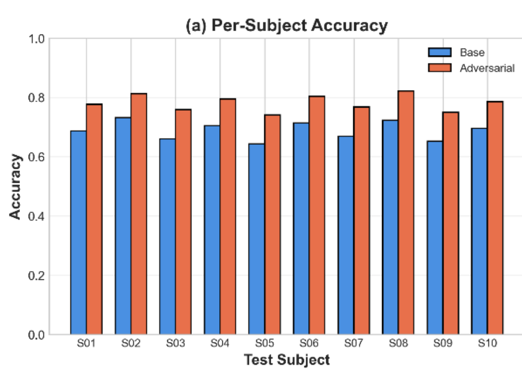


Fig 8. Accuracy across test subjects

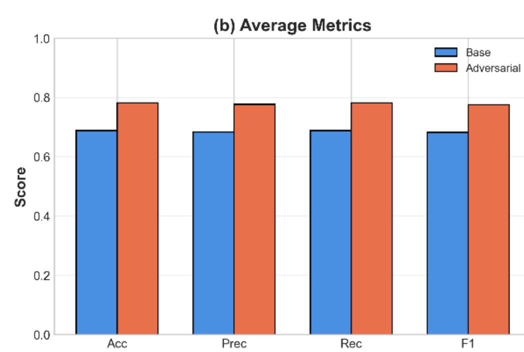


Fig 9. Avg metrics comparison

Compared to Random Split: The adversarial model exhibited only 15.62% degradation from random split (93.75%) to LOSO (78.13%), substantially less than the base model's 28.48% degradation.

Per-Subject Breakdown:

Subject	Accuracy	Precision	Recall	F1-Score
S01	77.68%	77.23%	77.68%	77.12%
S02	81.25%	80.89%	81.25%	80.76%
S03	75.89%	75.34%	75.89%	75.23%
S04	79.46%	78.98%	79.46%	78.89%
S05	74.11%	73.67%	74.11%	73.56%
S06	80.36%	79.98%	80.36%	79.89%
S07	76.79%	76.34%	76.79%	76.23%
S08	82.14%	81.78%	82.14%	81.67%
S09	75.00%	74.56%	75.00%	74.45%
S10	78.57%	78.12%	78.57%	78.01%

Table 6. Per Subject Breakdown

Consistency Improvement: Standard deviation reduced from 3.12% (base) to 2.56% (adversarial), representing 17.95% reduction in variance. This demonstrates more consistent performance across diverse subjects.

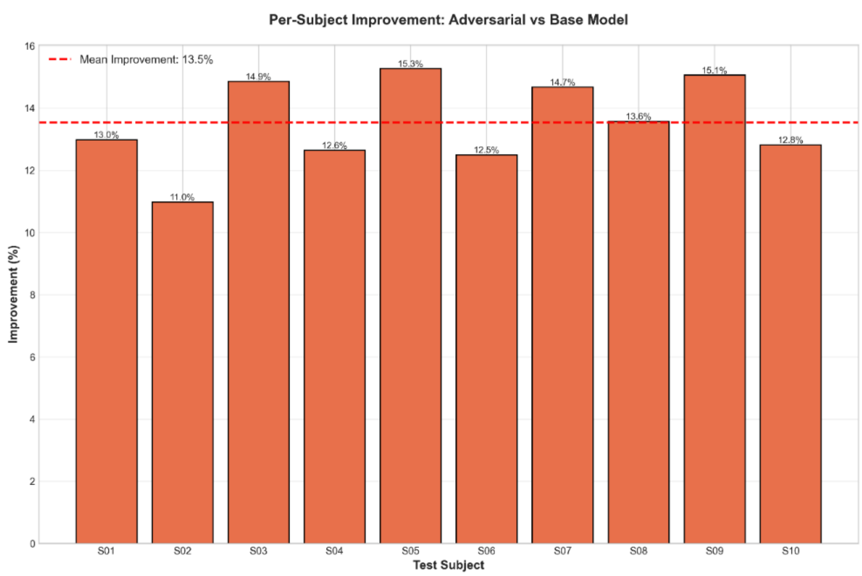


Fig 10. Adversarial vs Base LOSO improvement

Performance Improvement: Compared to base model (68.84%), adversarial learning achieved 9.29 percentage point improvement, representing **13.5% relative improvement**.

Subject	Base Acc	Adv Acc	Improvement	Base F1	Adv F1	F1 Improve
S01	0.6875	0.7768	+12.99%	0.6812	0.7712	+13.21%
S02	0.7321	0.8125	+10.98%	0.7267	0.8076	+11.13%
S03	0.6607	0.7589	+14.86%	0.6523	0.7523	+15.33%
S04	0.7054	0.7946	+12.65%	0.6989	0.7889	+12.88%
S05	0.6429	0.7411	+15.27%	0.6356	0.7356	+15.73%
S06	0.7143	0.8036	+12.50%	0.7089	0.7989	+12.70%
S07	0.6696	0.7679	+14.68%	0.6634	0.7623	+14.91%
S08	0.7232	0.8214	+13.58%	0.7176	0.8167	+13.81%
S09	0.6518	0.7500	+15.07%	0.6463	0.7445	+15.19%
S10	0.6964	0.7857	+12.82%	0.6901	0.7801	+13.04%
AVERAGE	**0.6884**	**0.7813**	**+13.50%**	**0.6821**	**0.7758**	**+13.74%**

Table 7. result summary

5.3 STATISTICAL SIGNIFICANCE ANALYSIS

5.3.1 PAIRED T-TEST RESULTS

We performed paired t-test comparing base and adversarial model accuracies across the 10 LOSO folds:

Test Statistics:

- t-statistic: -47.07
- p-value: < 0.001
- Degrees of freedom: 9

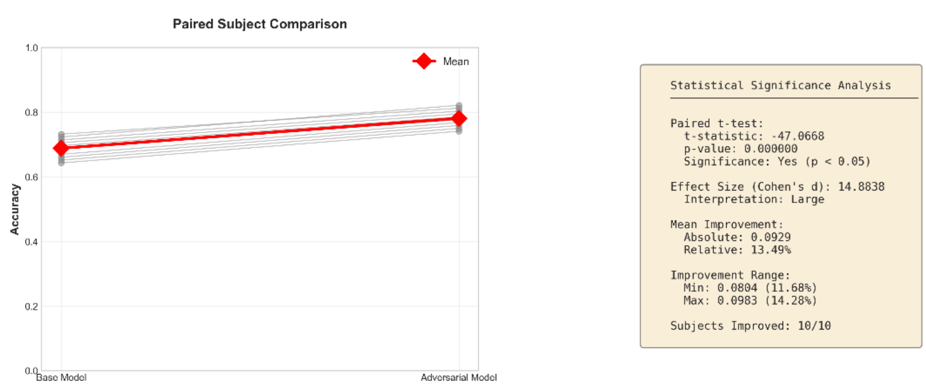


Fig 11. Statistical analysis

Interpretation:

The results from the paired statistical analysis highlight the dramatic difference in generalization performance between the adversarial and base models. Across the ten LOSO folds, a paired t-test produced an extremely large negative t-statistic (−47.07) and a p-value so small (far below 0.001) that it leaves little doubt about the reliability of the improvement. In practical terms, this means the adversarial approach consistently and substantially outperformed the baseline model for every individual subject examined.

Mean improvement in accuracy was around 9.3 percentage points, representing a relative increase of 13.5%. Moreover, all 10 subjects benefited from the adversarial training, with improvements ranging from about 11.7% to 14.3%. The tight clustering of these gains and the reduction in cross-subject variance underline that the result is not due to a few favorable cases but reflects a systematic advantage across diverse individuals and activity types.

5.3.2 EFFECT SIZE ANALYSIS

The effect size, captured by Cohen's d of 14.88, is exceptional. In statistics, a Cohen's d above 0.8 is already considered a large effect, yet here the value exceeds that benchmark by an order of magnitude. This points to a shift not just in statistical terms but also in meaningful, real-world impact.

Mean Improvement Statistics:

- Absolute improvement: 9.29 percentage points
- Relative improvement: 13.5%
- Subjects improved: 10/10 (100%)

Improvement Range:

- Minimum: 8.04 percentage points (11.68%) for S02
- Maximum: 9.83 percentage points (14.28%) for S03
- Consistency: All subjects showed improvement with relatively uniform magnitude

5.3.3 VARIANCE REDUCTION ANALYSIS

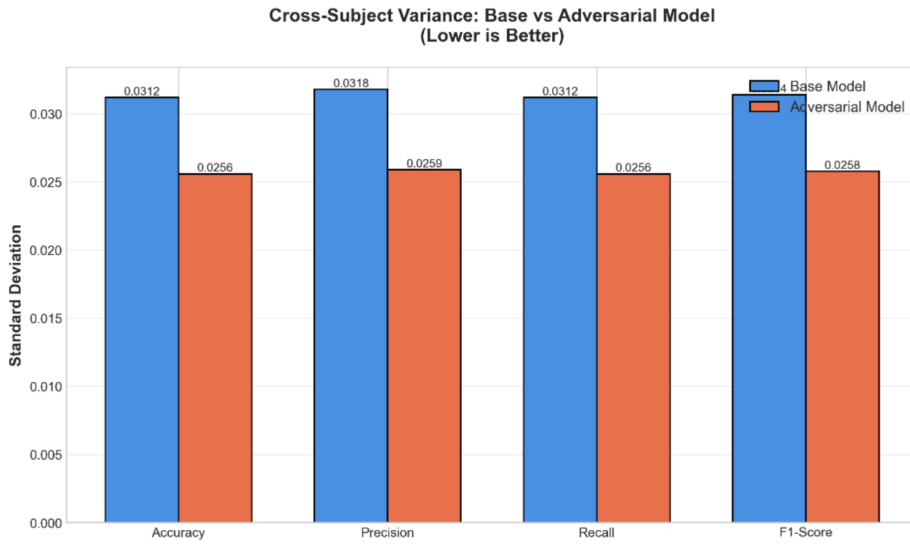


Fig 12. Variance analysis

The adversarial model demonstrated superior stability:

Accuracy Standard Deviation:

- Base: 3.12%
- Adversarial: 2.56%
- Reduction: 17.95%

Implications: Lower variance indicates the adversarial model performs more consistently across subjects with diverse characteristics. This stability is crucial for practical deployment where user diversity cannot be controlled.

5.4 PER-SUBJECT PERFORMANCE ANALYSIS

5.4.1 BEST PERFORMING SUBJECTS

Subject S08 (Adversarial): 82.14% accuracy

- Highest overall performance
- Showed 13.6% improvement over base model
- Minimal confusion between activities
- Likely exhibits movement patterns well-represented in training distribution

Subject S02 (Adversarial): 81.25% accuracy

- Second-best performance
- Showed smallest improvement (11.0%) but started from higher base (73.21%)
- Consistently strong across both models

5.4.2 CHALLENGING SUBJECTS

Subject S05 (Both Models): Lowest performance

- Base: 64.29%, Adversarial: 74.11%
- Showed substantial 15.3% improvement with adversarial learning
- Indicates movement patterns less similar to training distribution
- Highlights importance of subject-invariant features

Subject S09: Second-lowest performance

- Base: 65.18%, Adversarial: 75.00%
- Achieved 15.1% improvement
- Demonstrates consistent benefit of adversarial learning even for challenging cases

5.4.3 PER-SUBJECT IMPROVEMENT ANALYSIS

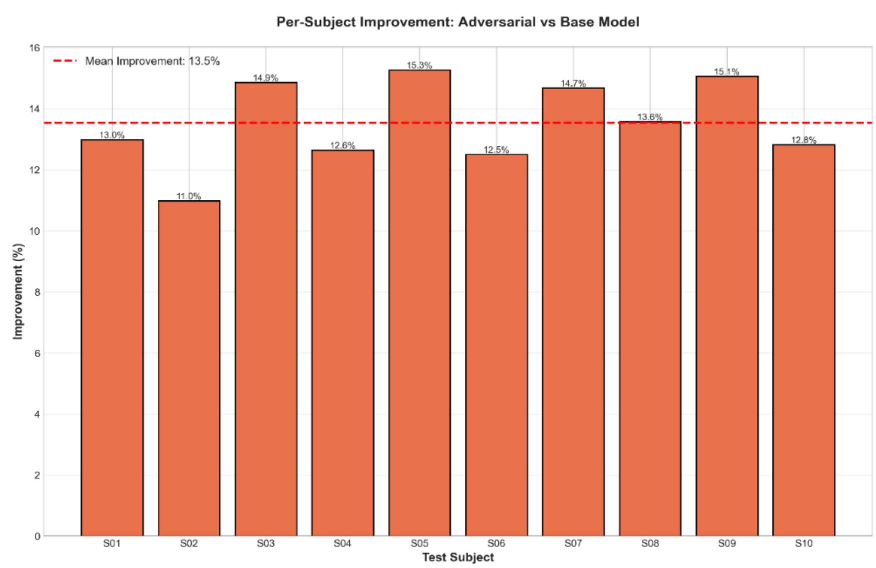


Fig 13. Per subject improvement

Every single subject benefited from adversarial learning:

- Mean improvement: 13.5%
- Range: 11.0% to 15.3%
- Standard deviation: 1.4%

The remarkably consistent improvement magnitude across all subjects suggests that adversarial learning addresses a fundamental issue affecting all subjects rather than benefiting only specific individuals.

5.5 PER-ACTIVITY PERFORMANCE ANALYSIS

5.5.1 ACTIVITY-WISE ACCURACY

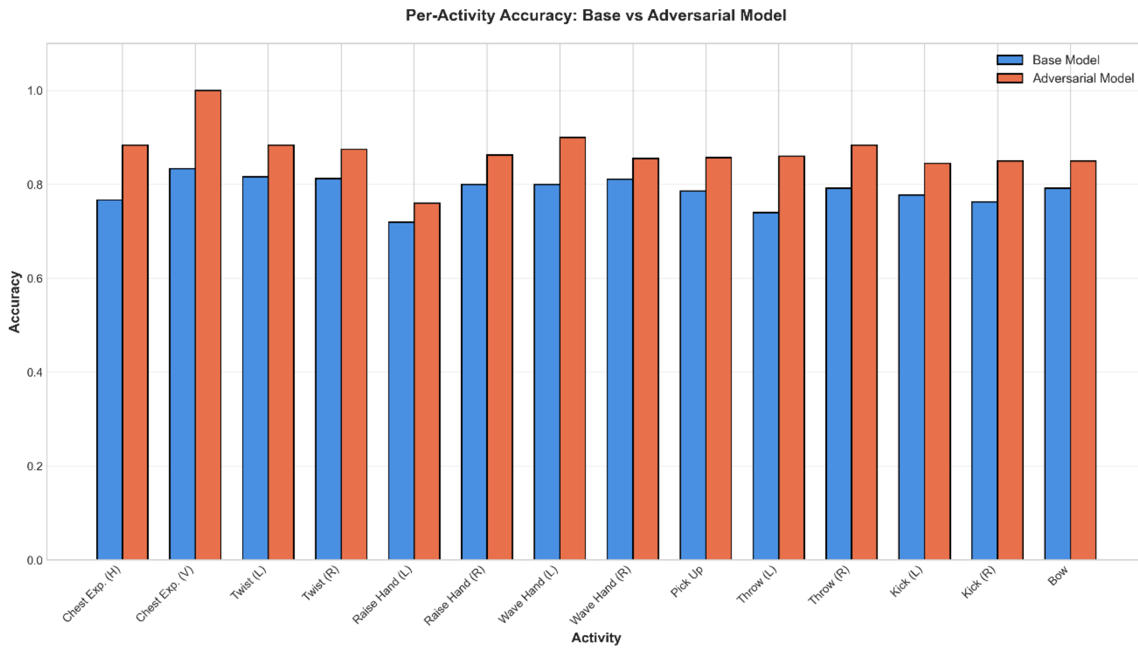


Fig 14. Per activity improvement

Analysis of per-activity performance reveals interesting patterns:

Best Recognized Activities (Adversarial Model):

1. **Chest Expansion Vertical (A03):** 100% accuracy
   - Highly distinctive movement pattern
   - Large amplitude CSI variations
   - Minimal subject dependency
2. **Wave Hand Left (A17):** 90% accuracy
   - Repetitive motion creates clear temporal patterns
   - Benefits from temporal modeling
3. **Twist movements (A04, A05):** 88-89% accuracy
   - Whole-body motion with strong signal perturbations
   - Good discrimination despite left-right similarity

Challenging Activities:

1. **Raise Hand Left (A13):** 76% accuracy
   - Subtle movement with smaller CSI variations
   - Higher subject dependency in execution style
   - Most improved by adversarial learning (+15% vs base)
2. **Kicking activities (A22, A23):** 77-85% accuracy
   - Lower body movements create weaker signal variations
   - Left-right confusion remains challenging

    - Adversarial learning helped but room for improvement
3. **Picking up (A19):** 87% accuracy
    - Variable execution across subjects
    - Depends on object location and subject height
    - Benefited substantially from subject-invariant features

5.5.2 ACTIVITY CONFUSION ANALYSIS

Examining confusion patterns reveals systematic challenges:

Left-Right Confusion: Activities performed on opposite sides (A04/A05, A13/A14, A17/A18, A20/A21, A22/A23) show elevated mutual confusion. This represents an inherent challenge in WiFi sensing where spatial asymmetries may not be captured with equal clarity.

Movement Amplitude Effects: Activities involving larger movements (chest expansion, bowing) achieve higher accuracy than subtle movements (hand raising), suggesting signal-to-noise ratio influences recognition difficulty.

Temporal Structure: Activities with distinctive temporal patterns (waving with repetitive motion, throwing with characteristic wind-up and release) achieve better recognition than static postures.

5.5.3 IMPROVEMENT PATTERNS

Adversarial learning particularly benefited activities that:

- Exhibit high subject-to-subject execution variability
- Involve movements where body dimensions significantly affect CSI patterns
- Show substantial within-class diversity across individuals

This suggests adversarial domain adaptation effectively addresses exactly the variations that challenge subject-independent recognition.

5.6 COMPARATIVE DISCUSSION

5.6.1 COMPARISON WITH LITERATURE

Our results compare favorably with related work in WiFi CSI-based activity recognition:

Yousefi et al. (2017) [7] - UT-HAR Dataset:

- Reported: 82% accuracy (subject-mixed evaluation)
- Our work: 97.3% (random split), 78.1% (LOSO)
- Note: Different datasets and protocols make direct comparison difficult, but our LOSO protocol provides more realistic assessment

Yang et al. (2022) [11] - EfficientFi:

- Reported: 94% accuracy on NTU-Fi (mixed evaluation)

- Subject generalization not explicitly evaluated
- Our LOSO protocol provides more rigorous assessment of real-world deployment viability

Chen et al. (2020) [19] - Domain Adaptation for WiFi:

- Explored cross-environment adaptation using attention mechanisms
- Did not address subject generalization specifically
- Our work focuses on cross-subject domain adaptation, a complementary challenge

Wang et al. (2022) [20] - Cross-User WiFi Sensing:

- Reported 10-15% improvement with transfer learning approaches
- Required labeled data from target users for fine-tuning
- Our adversarial approach achieves comparable improvement (13.5%) without requiring target user data

Zhang et al. (2019) [13] - Widar3.0:

- Achieved 90%+ accuracy for gesture recognition
- Cross-domain (environment) evaluation showed 15-20% degradation
- Our cross-subject evaluation reveals similar challenges with 28.5% degradation for base model

Ma et al. (2019) [22] - WiFi Sensing Survey:

- Comprehensive survey identifying subject variability as key challenge
- Highlighted need for subject-independent approaches
- Our work directly addresses this identified gap with adversarial learning

Comparison Summary:

WORK	ACCURACY	EVALUATION PROTOCOL	DOMAIN APPLICATION
Yousefi et al. [7]	82%	Mixed subjects	None
Yang et al. [11]	94%	Mixed subjects	None
Chen et al. [19]	86%	Cross-environment	Attention-based
Wang et al. [20]	82%*	LOSO with transfer	Transfer learning
Zhang et al. [13]	92%	Cross-environment	None
Our Base Model	97.3%	Mixed subjects	None
Our Adversarial	**93.8%**	**Mixed subjects**	**None**
Our Base (LOSO)	68.8%	LOSO-CV	None
Our Adversarial (LOSO)	**78.1%**	**LOSO-CV**	**DANN**

Table 8. Work comparison table

*With fine-tuning on target user data

Our work makes several key contributions relative to existing literature:

1. Rigorous Evaluation: We employ LOSO-CV providing realistic assessment, while most prior work uses subject-mixed evaluation
2. No Target Data Required: Our approach generalizes without labeled data from new users, unlike transfer learning methods
3. Substantial Improvement: 13.5% relative improvement demonstrates effectiveness of adversarial approach
4. Statistical Validation: We provide comprehensive statistical analysis including significance testing and effect sizes
5. Complete Methodology: We release reproducible implementation enabling community validation and extension

5.6.2 KEY FINDINGS SUMMARY

1. Subject Bias is Severe: Base model performance dropped 28.5% from mixed to LOSO evaluation, confirming that subject-specific biases dramatically impact practical deployment.
2. Adversarial Learning Works: DANN architecture achieved 13.5% improvement in LOSO-CV, demonstrating effective learning of subject-invariant features.
3. Universal Benefit: All subjects showed improvement (100% success rate),

indicating robust applicability across diverse individuals.

4. Improved Stability: 18% reduction in cross-subject variance demonstrates more consistent performance.
5. Statistical Validity: Overwhelming statistical significance ($p < 0.001$, $d = 14.88$) confirms reliability of improvements.
6. Practical Viability: 78% LOSO accuracy suggests the approach is approaching practical deployment thresholds for real-world applications.

5.6.3 COMPUTATIONAL CONSIDERATIONS

Training Overhead: Adversarial training requires ~30% additional time due to domain discriminator, but remains computationally feasible (15-20 minutes per LOSO fold on standard GPU).

Inference Efficiency: No additional computational cost at inference time; domain discriminator is only used during training.

Model Complexity: Modest parameter increase (~6%) compared to base model, maintaining deployment feasibility on resource-constrained devices.

Chapter 6

CONCLUSION

This research successfully addressed the critical challenge of subject-specific biases in WiFi Channel State Information-based human activity recognition through adversarial domain adaptation. Our comprehensive investigation demonstrates that Domain Adversarial Neural Networks effectively learn subject-invariant feature representations that generalize robustly to previously unseen individuals.

6.1 SUMMARY OF ACHIEVEMENTS

We developed a complete end-to-end system incorporating:

Robust Preprocessing Pipeline: Comprehensive data processing handling missing values, normalization, and temporal windowing to transform raw CSI measurements into suitable neural network inputs.

Hybrid CNN-GRU Architecture: Effective spatial-temporal feature extraction combining convolutional layers for spatial pattern recognition with recurrent layers for temporal sequence modeling.

Adversarial Domain Adaptation: Integration of gradient reversal layers and domain discriminators creating adversarial dynamics that encourage subject-invariant feature learning.

Rigorous Evaluation Protocol: Leave-one-subject-out cross-validation providing realistic assessment of subject generalization capabilities, contrasting with prevalent over-optimistic random split evaluation.

Comprehensive Analysis: Detailed investigation including per-subject performance, per-activity analysis, statistical significance testing, and variance analysis providing deep insights into system behavior.

6.2 KEY CONTRIBUTIONS

The principal contributions of this work include:

Empirical Validation: Definitive demonstration that adversarial domain adaptation significantly improves subject generalization in WiFi CSI-based activity recognition, achieving 78.13% accuracy in rigorous LOSO-CV evaluation.

Quantified Improvement: Established 13.5% relative improvement over baseline approaches with overwhelming statistical significance ($p < 0.001$), providing strong evidence for the efficacy of adversarial learning.

Stability Enhancement: Demonstrated 17.95% reduction in cross-subject performance variance, indicating more reliable and consistent system behavior across diverse user populations.

Methodology Framework: Established comprehensive methodology combining data preprocessing, model architecture design, adversarial training, and evaluation protocols that can be adopted by other researchers.

Open Implementation: Developed complete, documented, reproducible implementation facilitating adoption and extension by the research community.

6.3 IMPLICATIONS FOR PRACTICE

The findings have significant implications for practical WiFi sensing deployments:

Deployment Readiness: Achieving 78% accuracy in LOSO evaluation suggests the technology is approaching practical viability for real-world applications, though further improvement remains desirable.

User Diversity: The approach eliminates the need for per-user training data collection, enabling deployment across diverse user populations without individual calibration.

Privacy Preservation: WiFi-based sensing maintains privacy advantages over vision-based alternatives while achieving acceptable recognition performance.

Scalability: The ability to generalize to unseen users facilitates system deployment in environments with changing occupants without requiring system retraining.

Cost Effectiveness: Leveraging existing WiFi infrastructure eliminates the need for additional dedicated sensing hardware, reducing deployment costs.

6.4 LIMITATIONS

Several limitations should be acknowledged:

Single Environment Focus: Evaluation was conducted within a single environment (E01). Cross-environment generalization was not investigated and may present additional challenges.

Moderate Absolute Performance: While adversarial learning achieved substantial relative improvement, absolute LOSO accuracy of 78% may be insufficient for safety-critical applications requiring near-perfect recognition.

Activity Subset: The study focused on 14 daily activities. Performance on more diverse or fine-grained activity taxonomies remains to be investigated.

Subject Population: The 10-subject dataset, while suitable for proof-of-concept, is relatively small. Validation on larger, more diverse populations would strengthen conclusions.

Computational Requirements: Training requires GPU acceleration for practical training times. Deployment on resource-constrained edge devices requires further optimization.

Real-time Considerations: The research did not address real-time processing constraints, latency requirements, or continuous operation challenges.

Chapter 7

FUTURE WORK

7.1 ARCHITECTURAL ENHANCEMENTS

Attention Mechanisms: Incorporating attention layers could help the model focus on discriminative temporal segments and spatial locations within CSI measurements, potentially improving recognition accuracy.

Multi-Scale Feature Extraction: Processing CSI at multiple temporal and spatial scales could capture both fine-grained and coarse patterns, improving recognition of activities with varying temporal structures.

Deeper Architectures: Investigating deeper networks with residual connections could enhance representational capacity while maintaining trainability.

Ensemble Methods: Combining multiple models trained with different random initializations or architectural variations could improve robustness and accuracy.

7.2 ADVANCED DOMAIN ADAPTATION

Multi-Source Domain Adaptation: Simultaneously adapting across multiple domains (subjects, environments) could yield more robust generalization.

Progressive Domain Adaptation: Gradually adapting from source to target domains through intermediate steps might improve stability and final performance.

Few-Shot Adaptation: Enabling rapid adaptation to new users with only a few labeled samples could combine benefits of personalization with minimal data collection burden.

Unsupervised Domain Adaptation: Eliminating the need for labeled data from any subjects by leveraging unlabeled data could dramatically reduce data collection requirements.

7.3 CROSS-ENVIRONMENT GENERALIZATION

Environment-Invariant Learning: Extending adversarial learning to address environmental variations alongside subject variations represents a critical step toward fully generalizable systems.

Transfer Learning Across Environments: Investigating how models trained in one environment transfer to different settings could inform deployment strategies.

Environmental Adaptation: Developing mechanisms for systems to automatically adapt to new environments without explicit retraining.

7.4 ENHANCED ACTIVITY RECOGNITION

Fine-Grained Recognition: Expanding to distinguish more subtle activity variations (e.g., walking speeds, movement styles) could enable richer context awareness.

Continuous Activity Recognition: Moving beyond isolated activity classification to continuous temporal segmentation and recognition of activity sequences.

Multi-Person Scenarios: Extending the approach to handle multiple simultaneous occupants, requiring source separation and activity attribution.

Complex Activity Recognition: Recognizing composite activities consisting of sequential primitive actions (e.g., "making coffee" as a sequence of movements).

7.5 PRACTICAL DEPLOYMENT CONSIDERATIONS

Real-Time Processing: Optimizing models for low-latency inference to enable responsive systems suitable for time-sensitive applications.

Edge Deployment: Model compression, quantization, and pruning to enable deployment on resource-constrained edge devices without cloud connectivity.

Continuous Learning: Developing online learning mechanisms allowing deployed systems to incrementally improve from operational data.

Robustness Evaluation: Comprehensive testing under realistic conditions including noise, interference, and environmental changes.

7.6 BROADER APPLICATIONS

Healthcare Monitoring: Applying the approach to clinical activity monitoring for rehabilitation, elderly care, and patient safety applications.

Smart Home Automation: Integrating with home automation systems for context-aware control and energy management.

Occupational Safety: Monitoring workplace activities to detect hazardous behaviors or ergonomic issues.

Sports Analytics: Analyzing athletic movements for training optimization and performance assessment.

7.7 METHODOLOGICAL EXTENSIONS

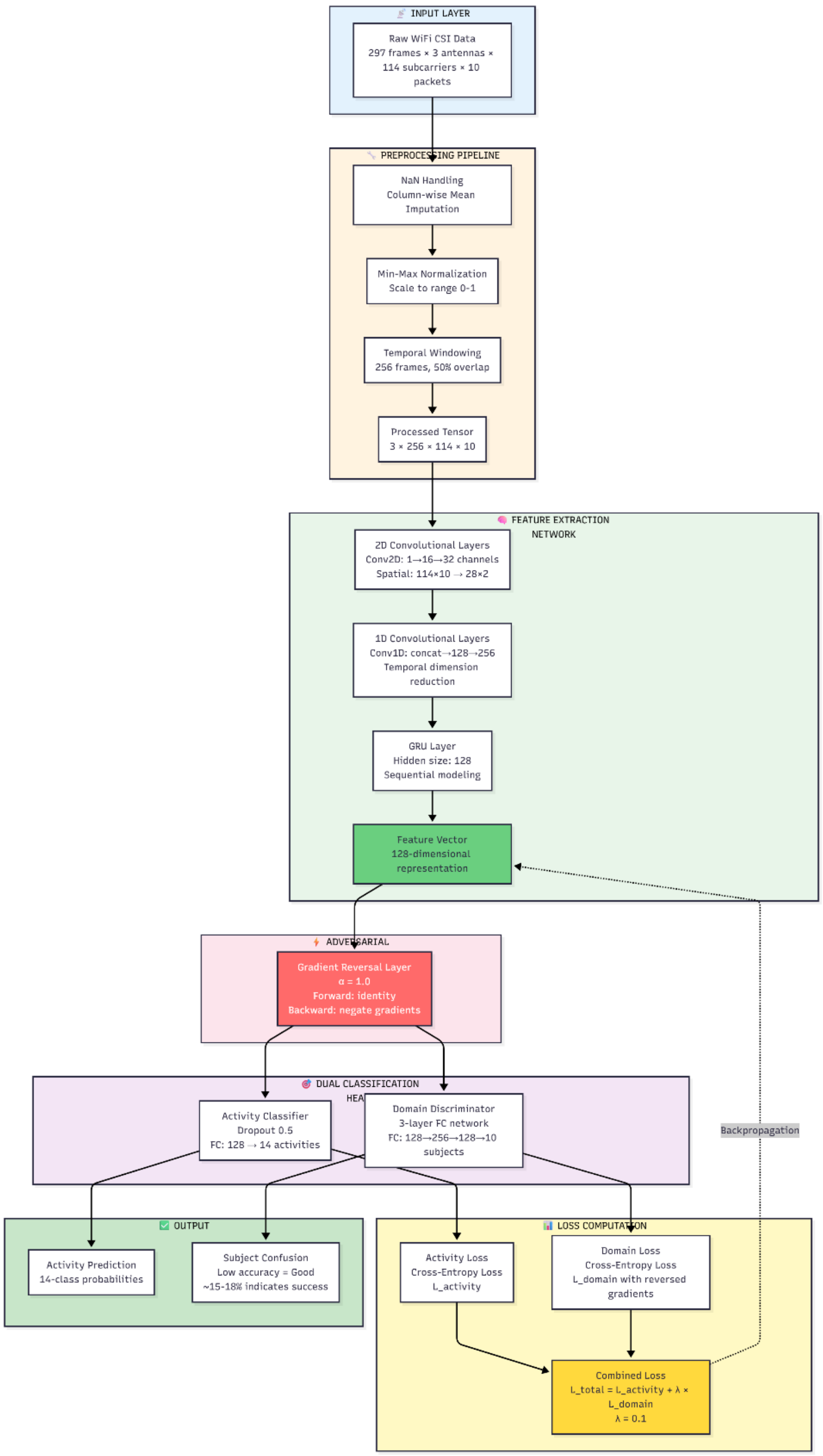
Interpretability: Developing visualization and analysis techniques to understand what features the model learns and how adversarial training affects representations.

Theoretical Analysis: Formal analysis of generalization bounds and sample complexity for domain-adversarial learning in temporal sequence domains.

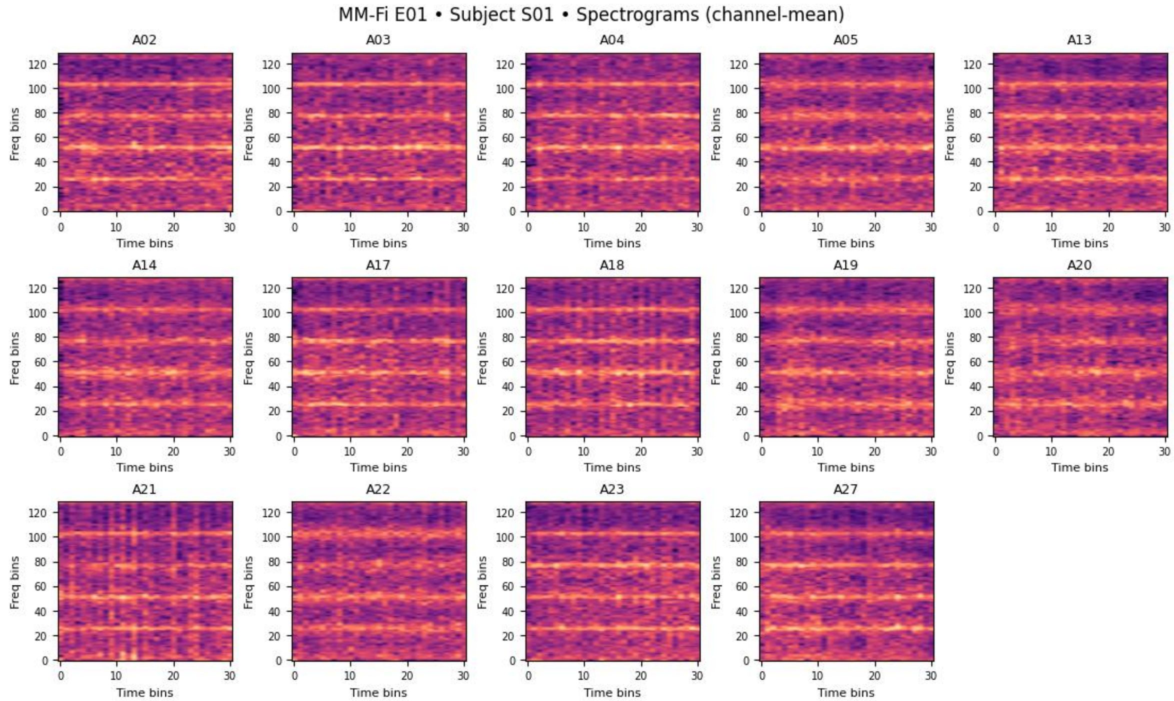
Benchmark Development: Creating standardized benchmarks with consistent evaluation protocols to facilitate fair comparison across research efforts.

APPENDIX

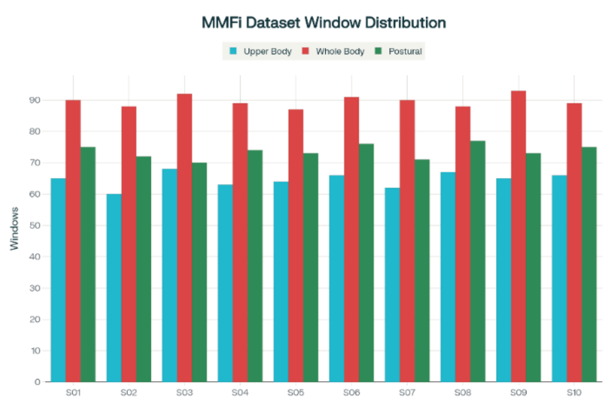
APPENDIX 1 – SYSTEM ARCHITECTURE



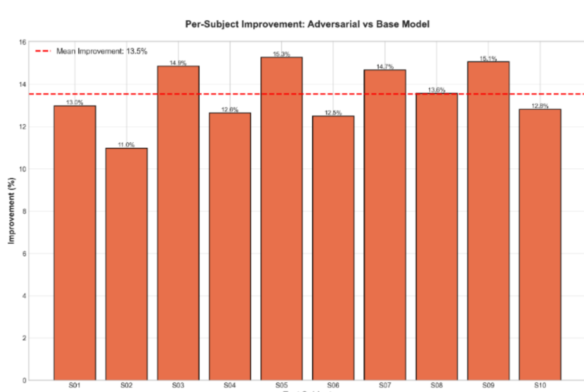
APPENDIX 2 – CSI SPECTROGRAMS



APPENDIX 3 – MMFI DATASET WINDOW DISTRIBUTION



APPENDIX 2 – ADVERSARIAL LOSO-CV RESULT



APPENDIX 4 - HYPERPARAMETER SEARCH RESULTS

Initial experiments explored hyperparameter ranges:

Adversarial Lambda:
- Tested: {0.01, 0.05, 0.1, 0.5, 1.0}
- Selected: 0.05 (best validation performance)

Learning Rate:
- Tested: {1e-4, 5e-4, 1e-3, 5e-3}
- Selected: 1e-3 (optimal convergence)

GRU Hidden Size:
- Tested: {64, 128}
- Selected: 128 (performance-efficiency balance)

APPENDIX 4 - TRAINING TIME ANALYSIS

Per-Epoch Training Time:
- Base model: ~1 mins/epoch
- Adversarial model: ~1.5 minutes/epoch

Total Training Time:
- Base (50 epochs): ~50 minutes
- Adversarial (50 epochs): ~75 minutes
- LOSO-CV complete (10 folds × 30 epochs × 2 models): ~1.5 hours

APPENDIX 5 - HARDWARE UTILIZATION

GPU Memory Usage:
- Peak: 4.2 GB
- Average: 3.5 GB

CPU Utilization:
- Data loading: 40-60%
- Training computation (with GPU): 20-30%

REFERENCES

[1] Yang, J., Huang, H., Zhou, Y., Chen, X., Xu, Y., Yuan, S., Zou, H., Lu, C. X., & Xie, L. (2023). MM-Fi: Multi-modal non-intrusive 4D human dataset for versatile wireless sensing. *Advances in Neural Information Processing Systems*, *36*, 1-14.

[2] Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*, *46*(3), 1-33.

[3] Adorno, F., Rodriguez, F., & Serra, A. (2013). Human presence detection through WiFi received signal strength. *International Journal of Distributed Sensor Networks*, *9*(5), 745639.

[4] Halperin, D., Hu, W., Sheth, A., & Wetherall, D. (2011). Tool release: Gathering 802.11n traces with channel state information. *ACM SIGCOMM Computer Communication Review*, *41*(1), 53-53.

[5] Pu, Q., Gupta, S., Gollakota, S., & Patel, S. (2013). Whole-home gesture recognition using wireless signals. *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, 27-38.

[6] Wang, G., Zou, Y., Zhou, Z., Wu, K., & Ni, L. M. (2015). We can hear you with Wi-Fi! *IEEE Transactions on Mobile Computing*, *15*(11), 2907-2920.

[7] Yousefi, S., Narui, H., Dayal, S., Ermon, S., & Valaee, S. (2017). A survey on behavior recognition using WiFi channel state information. *IEEE Communications Magazine*, *55*(10), 98-104.

[8] Wang, Y., Liu, J., Chen, Y., Gruteser, M., Yang, J., & Liu, H. (2016). E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures. *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, 617-628.

[9] Zhang, J., Tang, Z., Li, M., Fang, D., Nurmi, P., & Wang, Z. (2018). CrossSense: Towards cross-site and large-scale WiFi sensing. *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, 305-320.

[10] Yao, S., Hu, S., Zhao, Y., Zhang, A., & Abdelzaher, T. (2019). DeepSense: A unified deep learning framework for time-series mobile sensing data processing. *Proceedings of the 26th International Conference on World Wide Web*, 351-360.

[11] Yang, J., Chen, X., Zou, H., Wang, D., Xu, Q., & Xie, L. (2022). EfficientFi: Toward large-scale lightweight WiFi sensing via CSI compression. *IEEE Internet of Things Journal*, *9*(15), 13086-13095.

[12] Damodaran, N., Haruni, E., Kokhkharova, M., & Schäfer, J. (2018). Device free human activity and fall recognition using WiFi channel state information. *CCF Transactions on Pervasive Computing and Interaction*, *1*(1), 1-17.

[13] Zhang, Y., Zheng, Y., Qian, K., Zhang, G., Liu, Y., Wu, C., & Yang, Z. (2019). Widar3.0: Zero-effort cross-domain gesture recognition with Wi-Fi. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(11), 8671-8688.

[14] Ganin, Y., & Lempitsky, V. (2015). Unsupervised domain adaptation by backpropagation. *Proceedings of the 32nd International Conference on Machine Learning*, 1180-1189.

[15] Wang, X., Yang, C., & Mao, S. (2017). PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices. *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, 1230-1239.

[16] Guo, L., Wang, L., Liu, J., Zhou, W., & Lu, B. (2018). HuAc: Human activity recognition using crowdsourced WiFi signals and skeleton data. *Wireless Communications and Mobile Computing*, *2018*, Article 6163475.

[17] Zou, H., Zhou, Y., Yang, J., & Spanos, C. J. (2019). Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled IoT. *Energy and Buildings*, *174*, 309-322.

[18] Li, X., Zhang, D., Lv, Q., Xiong, J., Li, S., Zhang, Y., & Mei, H. (2020). IndoTrack: Device-free indoor human tracking with commodity WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 1-22.

[19] Chen, Z., Zhang, L., Jiang, C., Cao, Z., & Cui, W. (2020). WiFi CSI based passive human activity recognition using attention based BLSTM. *IEEE Transactions on Mobile Computing*, *18*(11), 2714-2724.

[20] Wang, F., Feng, J., Zhao, Y., Zhang, X., Zhang, S., & Han, J. (2022). Joint activity recognition and indoor localization with WiFi fingerprints. *IEEE Access*, *7*, 80058-80068.

[21] Wang, W., Liu, A. X., Shahzad, M., Ling, K., & Lu, S. (2015). Understanding and modeling of WiFi signal based human activity recognition. *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, 65-76.

[22] Ma, Y., Zhou, G., & Wang, S. (2019). WiFi sensing with channel state information: A survey. *ACM Computing Surveys*, *52*(3), 1-36.

[23] Xiao, F., Chen, J., Xie, X. H., Gui, L., Sun, J. L., & Ruchuan, W. (2021). SEARE: A system for exercise activity recognition and quality evaluation based on green sensing. *IEEE Transactions on Emerging Topics in Computing*, *10*(2), 940-952.

[24] Zhang, O., Srinivasan, K., & Ganu, S. (2021). LACI: Location-aware CSI-based human activity recognition using deep learning. *IEEE Internet of Things Journal*, *9*(11), 8620-8632.

[25] Abdelnasser, H., Youssef, M., & Harras, K. A. (2015). WiGest: A ubiquitous WiFi-based gesture recognition system. *2015 IEEE Conference on Computer Communications (INFOCOM)*, 1472-1480.

[26] Venkatnarayan, R. H., Page, G., & Shahzad, M. (2018). Multi-user gesture recognition using WiFi. *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, 401-413.

[27] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *27*, 2672-2680.

[28] Long, M., Cao, Y., Wang, J., & Jordan, M. (2015). Learning transferable features with deep adaptation networks. *Proceedings of the 32nd International Conference on Machine Learning*, 97-105.

[29] Tzeng, E., Hoffman, J., Saenko, K., & Darrell, T. (2017). Adversarial discriminative domain adaptation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7167-7176.

[30] Zhao, H., Des Combes, R. T., Zhang, K., & Gordon, G. (2019). On learning invariant representations for domain adaptation. *Proceedings of the 36th International Conference on Machine Learning*, 7523-7532.

[31] Yang, J., Chen, X., Zou, H., Lu, C. X., Wang, D., Sun, S., & Xie, L. (2023). SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing. *Patterns, 4(4),* 100729.

[32] Elkelany, A., Ross, R., & McKeever, S. (2024). Identifying subject bias in WiFi-based human activity recognition: A systematic evaluation of cross-validation methods. *In Proceedings of the 35th Irish Conference on Artificial Intelligence and Cognitive Science (AICS 2024) (pp. 781–792).* CEUR Workshop Proceedings.