Thèse
de doctorat
de l'UTT

Hadi EL ZEIN

Deep Learning for the Classification of Multidimensional Signals Applied to Human Activity Recognition

Champ disciplinaire :
Sciences pour l'Ingénieur

2025TROY0021 **Année 2025**

Thèse en cotutelle avec l'Université Libanaise – Beyrouth – Liban





THESE

pour l'obtention du grade de

DOCTEUR
de l'UNIVERSITE DE TECHNOLOGIE DE TROYES

en SCIENCES POUR L'INGENIEUR

Spécialité : OPTIMISATION ET SURETE DES SYSTEMES

présentée et soutenue par

Hadi EL ZEIN

le 27 juin 2025

Deep Learning for the Classification of Multidimensional Signals Applied to Human Activity Recognition

JURY

Mme Régine LE BOUQUIN JEANNES	PROFESSEURE DES UNIVERSITES	Présidente
M. Mehdi AMMI	PROFESSEUR DES UNIVERSITES	Rapporteur
M. Éric CAMPO	PROFESSEUR DES UNIVERSITES	Rapporteur
M. Amine NAIT-ALI	PROFESSEUR DES UNIVERSITES	Examinateur
M. Hassan AMOUD	PROFESSEUR	Directeur de thèse
Mme Farah MOURAD-CHEHADE	MAITRESSE DE CONFERENCES	Directrice de thèse

Dedicated to my family, friends, teachers, and supervisors,
whose unwavering encouragement, inspiration, and support
made this journey possible.

Remerciements

La réalisation de cette thèse n'aurait pas été possible sans l'apport et le soutien de nombreuses personnes, que je tiens ici à remercier sincèrement.

Je voudrais tout d'abord exprimer ma profonde gratitude envers mes directeurs de thèse, **Monsieur Hassan AMOUD** et **Madame Farah CHEHADE**, pour leur soutien inestimable, leurs conseils avisés et leur confiance tout au long de cette recherche.

J'adresse également mes sincères remerciements aux membres du jury pour le temps consacré à évaluer ce travail : **Monsieur Mehdi AMMI**, **Monsieur Eric CAMPO**, **Madame Régine LE BOUQUIN JEANNES** et **Monsieur Amine NAIT-ALI**.

Je remercie ensuite les membres de mon comité de suivi de thèse, **Madame Elnaz ASADOLLAHI**, **Monsieur Frédéric BERTRAND** et **Madame Estelle DELOUX**, pour leurs remarques pertinentes et leur accompagnement régulier.

Je remercie également **Monsieur Antoine GRALL**, directeur de l'école doctorale de l'Université de Technologie de Troyes, et **Monsieur Kassem HAMZE**, doyen de l'école doctorale de l'Université Libanaise.

J'exprime aussi ma gratitude à **Madame Bernadette ANDRÉ**, **Madame Véronique BANSE**, **Madame Fabienne BORNOT**, **Madame Aurélie CHAILLEY**, **Madame Annabelle CHEPY**, **Madame Julie CHIPAUX**, **Madame Pascale DENIS**, **Madame Jana EL HAJ**, **Madame Zainab IBRAHIM**, **Madame Thérèse KAZARIAN** et **Madame Isabelle LECLERCQ** pour leur assistance administrative précieuse.

Enfin, j'adresse mes vifs remerciements à **Monsieur Lionel AMODEO**, responsable du **Laboratoire LIST3N** de l'Université de Technologie de Troyes, et **Monsieur Mohamad KHALIL**, responsable du **Centre Azm pour la Recherche en Biotechnologie et ses Applications** de l'Université Libanaise, ainsi qu'à mes collègues de ces deux laboratoires, dont les échanges constructifs m'ont permis de beaucoup partager et apprendre tout au long de ce parcours doctoral.

Résumé

La reconnaissance d'activités humaines (HAR) via la Channel State Information (CSI) s'impose comme une solution non intrusive dans le secteur de la santé. Cette thèse aborde deux problématiques majeures liées aux données et aux modèles, à travers trois objectifs : (1) développer une méthodologie de HAR efficace pour la reconnaissance d'activités en temps réel, (2) améliorer la précision de la détection de chutes dans un contexte d'interventions médicales urgentes, et (3) permettre l'identification d'utilisateurs à partir de schémas comportementaux subtils. Pour y parvenir, des solutions d'apprentissage profond intègrent l'augmentation de données, des architectures convolutionnelles ou basées sur des transformers, ainsi qu'un mécanisme de décision avancé.

Trois contributions principales en découlent. D'abord, **HAR-LightCNN** traite les jeux de données limités et déséquilibrés grâce à un réseau de neurones convolutionnel allégé. Ensuite, **HFD-Adv** introduit une méthodologie en quatre étapes (fonction de coût optimisée par algorithmes génétiques, transfert d'apprentissage, post-traitement démographique) pour minimiser les faux négatifs. Enfin, **IdentNet** exploite l'auto-attention d'un transformer afin de distinguer des utilisateurs ayant des profils d'activité similaires.

En rendant ces systèmes plus fiables et adaptatifs, ce travail contribue à des applications concrètes dans le domaine de la santé, depuis la simple reconnaissance d'activités jusqu'à l'intervention personnalisée en cas de chute.

Mots clés : CSI (télécommunications), Reconnaissance de l'activité humaine (informatique), Apprentissage profond, Chutes (accidents), Suivi médical

Abstract

Human Activity Recognition (HAR) using Wi-Fi Channel State Information (CSI) enables privacy-preserving, device-free monitoring in healthcare settings. This thesis tackles data- and model-related concerns through three objectives: (1) developing an efficient real-time HAR methodology, (2) enhancing fall detection accuracy for urgent healthcare interventions, and (3) enabling user identification based on subtle behavioral patterns. To achieve this, novel deep learning solutions combine data augmentation, convolutional or transformer-based architectures, and advanced decision mechanisms.

Three core contributions follow. First, **HAR-LightCNN** addresses small, imbalanced datasets with a lightweight CNN. Second, **HFD-Adv** integrates genetic algorithm-optimized loss functions, transfer learning, and demographic-sensitive post-processing to enhance fall detection. Finally, **IdentNet** leverages transformer self-attention to distinguish users with similar activity patterns.

By increasing reliability and adaptability, these methodologies support real-world healthcare applications, from basic activity recognition to personalized emergency responses for fall incidents.

Keywords: CSI (Telecommunications), Human activity recognition, Deep learning (Machine learning), Falls (Accidents), Medical Monitoring

Table of Contents

Remerciements iii

Résumé iv

Abstract v

Table of Contents v

List of Figures xi

List of Tables xii

General Introduction 1
1 Context and Motivations . . . 1
2 Problems and Challenges . . . 2
3 Research Questions . . . 3
4 Contributions of the Thesis . . . 4
5 Organization of the Manuscript . . . 4

I STATE OF THE ART 6

1 Fundamentals of Wi-Fi Channel State Information in Human Activity Recognition 8
1 Introduction . . . 10
1.1 Context and Definition of our Research Domain . . . 10
1.2 Chapter Purpose . . . 11
1.3 Organization of the Chapter . . . 12
2 Definitions and Evolution of CSI Technology . . . 12
3 Temporal Dynamics in CSI and Its Role in HAR . . . 14
4 CSI Characteristics and Its Application in HAR . . . 15
4.1 Effects of Moving Objects Factors on CSI . . . 15
4.2 Distinctions between CSI and Other Activity Measurement Technologies . . . 17
5 Effects of Environmental Factors on CSI . . . 18
6 Integration of MIMO and OFDM Technologies in CSI for HAR . . . 19
6.1 MIMO Technology . . . 19
6.2 OFDM Technology . . . 20
6.3 Combined MIMO-OFDM Systems . . . 21
7 System Setup for CSI Data Collection . . . 24
7.1 Hardware Requirements . . . 24
7.2 Software Tools . . . 25
7.3 Environments and Parameters . . . 25
7.4 Configuration and Collection Process . . . 25
8 Publicly Available CSI Datasets for Evaluation . . . 26
8.1 First CSI Dataset . . . 26

8.2 Second CSI Dataset . . . 28
8.3 Third CSI Dataset . . . 28
9 Methodologies for Processing CSI Data in HAR . . . 30
9.1 Data Filtering and Normalization . . . 31
9.2 Feature Extraction . . . 33
9.3 Potential of Deep Learning Approaches . . . 35
10 Conclusion . . . 35

2 Systematic Literature Review of CSI-Based HAR for healthcare domain 37
1 Introduction . . . 38
1.1 Chapter Purpose . . . 38
1.2 Selecting a Suitable Literature Review Approach . . . 38
1.2.1 Overview of Literature Review Approaches . . . 38
1.2.2 Motivation for a Systematic Mapping Study . . . 38
1.3 Organization of the Chapter . . . 39
2 Systematic Mapping Approach: Search, Screening, and Classification . . . 40
3 Narrative Review of the Selected Articles . . . 46
3.1 Multi-activity Recognition–Focused Approaches . . . 47
3.2 Fall Detection–Focused Approaches . . . 47
3.3 Summary and Next Steps . . . 49
4 Key Challenges and Limitations in CSI-Based HAR . . . 50
4.1 Data-Related Concerns . . . 50
4.2 Model-Related Concerns . . . 51
4.2.1 Processing Technique Concerns . . . 51
4.2.2 Post-processing Mechanism Concerns . . . 51
4.2.3 Evaluation Process Concerns . . . 52
5 Performance Aspects for CSI-Based HAR Methodologies . . . 52
5.1 Effectiveness of the Methodology . . . 53
5.2 Efficiency of the Methodology . . . 53
5.3 Versatility of the Methodology . . . 54
6 Potential Healthcare Services Benefiting from HAR . . . 55
7 Synthesis and Interpretation of the Mapped Literature . . . 56
7.1 Classification Scheme Description . . . 56
7.2 Illustrative Example of Table Entry Filling . . . 57
7.3 Observed Trends and Common Limitations . . . 57
7.4 Bridging Identified Gaps to Our Contributions . . . 58
8 Conclusion . . . 59

II CONTRIBUTIONS 62

3 CSI-Based HAR Approach via a Lightweight CNN and Data Augmentation for Patient Monitoring Service 64
1 Introduction . . . 65
1.1 Background and Importance of HAR in Modern Healthcare . . . 65
1.2 Defining the Problem Space and Its Challenges . . . 65
1.3 Preliminaries . . . 66
1.3.1 Overview of Time-Series Data Augmentation Techniques . . . 66
1.3.2 Theoretical Background on Convolutional Neural Networks . . . 68
1.4 Contributions . . . 70
1.5 Organization of the Chapter . . . 71
2 HAR-LightCNN Methodology Overview . . . 72
2.1 Introduction to the Methodology . . . 72
2.2 Objectives and Design Principles . . . 73
3 Data Preparation and Preprocessing . . . 74
3.1 Datasets Overview and Initial Setup Considerations . . . 74

3.2 Overview of the Preprocessing Steps . . . 75
3.2.1 Testing Subset Preparation for Ensuring Fair Evaluation . . . 76
3.2.2 Validation Procedures for Ensuring Reliable Training . . . 77
3.2.3 Implementing Data Augmentation for Enhancing Generalization . . . 79
4 Lightweight CNN Activity Recognizer . . . 83
4.1 Selecting CNNs for Multidimensional Time-Series Analysis . . . 83
4.2 Architecture of the Recognizer . . . 84
4.2.1 Layers and Configurations . . . 84
4.2.2 Lightweight Design and Efficiency Optimizations . . . 87
4.3 Implementation Details . . . 89
5 Results and Evaluation . . . 89
5.1 Benchmarking with Standardized Testing Subsets and Evaluation Metrics . . . 89
5.2 Effectiveness and Efficiency of HAR-LightCNN . . . 91
5.2.1 Evaluation of HAR-LightCNN Without DA . . . 93
5.2.2 Evaluation of HAR-LightCNN With DA . . . 93
5.2.3 Benchmarking Against State-of-the-Art Models . . . 97
6 Conclusions . . . 98
6.1 Summary of Key Findings . . . 98
6.2 Benefits for Healthcare Services . . . 98
6.2.1 Enhancing Patient Monitoring and Care . . . 98
6.2.2 Advantages Over Traditional Monitoring Methods . . . 99
6.2.3 Scenario: Application in Elderly Care Facilities . . . 99
6.2.4 Advancing Healthcare Delivery . . . 100
6.2.5 Summary of Benefits . . . 100
6.3 Future Directions . . . 100

4 CSI-Based Human Fall Detection through a Comprehensive Four-Stage DL Approach for Emergency Response Service 102
1 Introduction . . . 104
1.1 Public Health Impact of Falling Accidents . . . 104
1.2 Significance of Rapid Intervention in Healthcare . . . 105
1.3 Defining the Problem Space and Its Challenges . . . 107
1.4 Contributions . . . 107
1.5 Organization of the Chapter . . . 109
2 HFD-Adv Methodology Overview . . . 109
2.1 Introduction to the Methodology . . . 109
2.2 Objectives and Design Principles . . . 110
3 Data Preparation and Signal Preprocessing . . . 114
3.1 Collection and Structuring of Activity CSI Data . . . 115
3.2 System Calibration with Empty Activity Data . . . 116
3.3 User-Specific Train-Test Split . . . 116
3.4 Dimensionality Reduction with PCA . . . 116
3.5 Handling Incomplete Frames with Threshold-Based Padding Strategy . . . 117
4 ActiNet Flexible CNN Classifier . . . 120
4.1 Architecture for Fall Detection . . . 121
4.2 Weighted Loss Function Optimization using Genetic Algorithm . . . 121
4.3 Learning Rate Optimization with LRFinder . . . 124
4.4 Hyperparameter Optimization and Training Strategies . . . 126
5 Transfer Learning for Model Adaptability . . . 127
5.1 Role of Transfer Learning in Enhancing Model Adaptability . . . 127
5.2 Transfer Learning Strategy and Implementation for Model Fine-Tuning . . . 127
6 Advanced Decision Mechanism for Classifier Outputs Reassessment . . . 128
6.1 Introduction to the Advanced Decision Mechanism . . . 128
6.2 Age-Sensitive Threshold Adjustment . . . 130
7 Results and Evaluation . . . 131
7.1 Evaluation Metrics and Benchmarking . . . 131

7.2 Effectiveness, Efficiency, and Versatility of HFD-Adv . . . 132
7.3 Benchmarking Against State-of-the-Art Models . . . 133
8 Conclusions . . . 134
8.1 Summary of Key Findings . . . 134
8.2 Benefits for Healthcare Services . . . 135
8.2.1 Improving Emergency Response to Falls . . . 135
8.2.2 Reducing Healthcare Costs and Burden . . . 135
8.2.3 Enhancing Patient Safety and Quality of Life . . . 136
8.2.4 Facilitating Integration into Existing Healthcare Frameworks . . . 136
8.2.5 Scenario: Application in Home-Based Care . . . 136
8.2.6 Advancing Public Health Outcomes . . . 137
8.2.7 Summary of Benefits . . . 137
8.3 Future Directions . . . 137
9 Appendix: Additional Validation on a Different Left-out User (user_2) . . . 138

5 CSI-Based Human Identity Recognition Approach Using a Transformer-Based Model for Personalized Healthcare 141
1 Introduction . . . 143
1.1 Background and Importance of HIDR in HAR for Personalized Healthcare . . . 143
1.2 Defining the Problem Space and Its Challenges . . . 143
1.3 Literature Review on CSI-Based Human Identity Recognition . . . 145
1.3.1 CSI-Based Human Identity Recognition Approaches . . . 145
1.3.2 Techniques for Robust CSI-Based Identity Recognition . . . 145
1.3.3 Applications of CSI-Based Identity Recognition . . . 146
1.3.4 Gaps and Limitations in Existing Methods . . . 146
1.3.5 Positioning the Current Work . . . 147
1.4 Contributions . . . 147
1.5 Organization of the Chapter . . . 148
2 Methodology Overview . . . 149
2.1 Introduction to the Methodology . . . 149
2.2 Objectives and Design Principles . . . 149
3 User CSI Data Preprocessing for User Identification . . . 151
4 IdentNet Transformer User Identifier . . . 152
4.1 Theoretical Background on Transformers and Their Use in Multidimensional Time-Series Analysis . . . 152
4.2 Architecture of the Identifier . . . 153
5 Results and Evaluation . . . 156
5.1 Evaluation Metrics and Benchmarking . . . 156
5.2 Effectiveness and Efficiency of IdentNet . . . 157
5.2.1 Comparison Against CNN-Based Model from Previous Contributions . . . 157
5.2.2 Benchmarking Against State-of-the-Art Models . . . 160
6 Conclusions . . . 161
6.1 Summary of Key Findings . . . 161
6.2 Benefits for Healthcare Services . . . 162
6.2.1 Enhancing Personalized Emergency Response . . . 162
6.2.2 Protocol for Integrated HFD-Adv and IdentNet Deployment . . . 163
6.2.3 Scenario: Application in Residential Healthcare Settings . . . 164
6.2.4 Advancing Personalized Healthcare . . . 164
6.2.5 Summary of Benefits . . . 164
6.3 Future Directions . . . 165
7 Appendix: Extending IdentNet to Other User Activities . . . 166
7.1 Introduction and Experimental Hypotheses . . . 166
7.2 User CSI Data Preprocessing . . . 166
7.2.1 Run-based Training Set Preparation . . . 166
7.2.2 Holdout Method for Robust Evaluation . . . 167
7.2.3 Three Testing Sets Preparation . . . 167

7.3 Experiments and Results . . . 167
7.3.1 Baseline: Identity Recognition through Running . . . 167
7.3.2 Direct Generalization to Falling User Data . . . 167
7.3.3 Fine-Tuning on Limited Falling User Data . . . 167
7.3.4 Discussion . . . 168
7.4 Conclusion and Outlook . . . 169

General Conclusion and Perspectives **170**
1 Conclusions and Discussions . . . 170
2 Perspectives . . . 172

Résumé long de la thèse **175**
1 Contexte et motivations . . . 175
2 Problématique . . . 175
3 Questions de recherche . . . 176
4 Objectifs et contributions . . . 176
5 Aperçu méthodologique . . . 177
6 Principaux résultats et évaluation . . . 178
7 Impact et perspectives . . . 179

List of Publications **180**

Bibliography **181**

List of Figures

1.1 Wi-Fi signal propagation and reflection patterns . . . 15
1.2 Bending CSI Amplitude Data Signal from 1st dataset [Fard Moshiri, 2021] . . . 27
1.3 Falling CSI Amplitude Data Signal from 2nd dataset [Yousefi, 2017] . . . 29
1.4 Empty and Standing Up CSI Amplitude Data Signals from 3rd dataset [Schäfer, 2021a] . . . 30
1.5 CSI Data Tensor Structure for HAR . . . 32

2.1 HAR research by subject area . . . 41
2.2 Evolution of HAR research . . . 42
2.3 PRISMA Flow Diagram . . . 45

3.1 HAR-LightCNN three-phase methodology schema . . . 75
3.2 Class Distribution in test data array of 1st dataset before balancing . . . 76
3.3 Class Distribution in test data array of 2nd dataset before balancing . . . 77
3.4 Class Distribution in train data array of 2nd dataset before balancing . . . 79
3.5 Dataset preprocessing steps . . . 80
3.6 Augmentation effect of random-transformation techniques . . . 81
3.7 Augmentation effect of RGW pattern-mixing technique . . . 82
3.8 Architecture of the lightweight CNN multi-activity classifier . . . 87
3.9 Confusion Matrix for Most Efficient Classifier (with DA) on 1st dataset . . . 95
3.10 Confusion Matrix for Most Efficient Classifier (with DA) on 2nd dataset . . . 96

4.1 Monthly Trends in Unintentional Older Adults Falls (Age 65 and Older) in the U.S. . . . 104
4.2 HFD-Adv four-stage methodology schema . . . 114
4.3 Steps of preprocessing stage . . . 115
4.4 Architecture of ActiNet fall detector . . . 122
4.5 Plot of loss values versus learning rates determined by LRFinder . . . 125
4.6 Impact of ADM on Frame-wise Prediction for a Fall Activity Trial . . . 139
4.7 Confusion Matrices for Experiments 1a–2 . . . 139
4.8 Confusion Matrices for Experiments 3–6 . . . 140

5.1 Class Distribution in test data array . . . 152
5.2 Architecture of Transformer user identifier . . . 155
5.3 Confusion Matrix for CNN-Based Classifier . . . 158
5.4 Confusion Matrix for Transformer-Based IdentNet . . . 159

List of Tables

2.1 State-of-the-art of CSI-based HAR for healthcare: ML and signal processing methodologies 61

3.1 Summary of Model Hyper-Parameters . . . 88
3.2 Efficiency Benchmarking of the Proposed CNN Against Established Lightweight CNN Architectures . . . 88
3.3 Testing Results on [Fard Moshiri, 2021] without DA . . . 93
3.4 Testing Results on [Yousefi, 2017] without DA . . . 93
3.5 Testing Results on [Fard Moshiri, 2021] with DA . . . 94
3.6 Classification Report of Most Efficient Classifier on [Fard Moshiri, 2021] . . . 94
3.7 Testing Results on [Yousefi, 2017] with DA . . . 95
3.8 Classification Report of Most Efficient Classifier on [Yousefi, 2017] . . . 96
3.9 F1-score Values of Best Classifier on [Yousefi, 2017] without DA . . . 96
3.10 Results on [Fard Moshiri, 2021] for HAR-LightCNN vs. State-of-the-art . . . 97
3.11 Results on [Yousefi, 2017] for HAR-LightCNN vs. State-of-the-art . . . 97

4.1 Summary of Hyperparameters and Training Strategies for ActiNet . . . 126
4.2 Performance of ActiNet Across Experiments . . . 133
4.3 Results on testing set derived from 1st and 3rd datasets for HFD-Adv vs. State-of-the-art 134
4.4 Performance of ActiNet Across Experiments for Left-out $user_2$ . . . 138

5.1 Summary of Hyperparameters for IdentNet . . . 156
5.2 Comparative Results of CNN-Based Classifier and Transformer-Based IdentNet . . . 159
5.3 Comparison of IdentNet with State-of-the-Art Models in CSI-Based HIDR . . . 160
5.4 Performance of IdentNet on Run-based Testing Set . . . 167
5.5 Performance of IdentNet on Fall-based Testing Set Before Fine-tuning . . . 168
5.6 Performance of IdentNet After Fine-tuning . . . 168

List of Abbreviations

ABLSTM *Attention-based Bi-directional Long Short-Term Memory*
ACF *Autocorrelation Function*
ADM *Advanced Decision Mechanism*
ANN *Artificial Neural Network*
BP *Band-Pass*
CFO *Carrier Frequency Offset*
CNN *Convolutional Neural Network*
CSI *Channel State Information*
CV *Cross-Validation*
DA *data augmentation*
DBA *DTW Barycenter Averaging*
DBSCAN *Density-Based Spatial Clustering of Applications with Noise*
DFS *Doppler Frequency Shift*
DL *deep learning*
DTW *Dynamic Time Warping*
DWT *Discrete Wavelet Transform*
EMD *Empirical Mode Decomposition*
FDR *Fall Detection Rate*
FLOPs *Floating-point Operations*
FNs *false negatives*
GAN *Generative Adversarial Network*
GW *Guided Warping*
HAR *Human Activity Recognition*
HFD *Human Fall Detection*
HFD-Adv *Advanced Human Fall Detection*
HIDR *Human Identity Recognition*
IoT *Internet of Things*
kNN *k-Nearest Neighbor*
LOS *Line-of-Sight*
LPF *Low-Pass*
LSR *Least Squares Regression*
MA *Moving Average*
MIMO *Multiple Input Multiple Output*
ML *machine learning*

MUSIC *Multiple Signal Classification*
MVS *Moving Variance Segmentation*
NB *Naive Bayes*
NICs *network interface cards*
OFDM *Orthogonal Frequency-Division Multiplexing*
PCA *Principal Component Analysis*
PDD *Packet Detection Delay*
PHY *physical layer*
RBF *Radial Basis Function*
RF *Random Forest*
RGW *Random Guided Warping*
RNNs *Recurrent Neural Networks*
RSSI *Received Signal Strength Indicator*
SFO *Sampling Frequency Offset*
SLN-DTW *Band-relaxed Segmental Locally Normalized DTW*
SLR *Systematic Literature Review*
SMS *Systematic Mapping Study*
SR *Systematic Reviews*
STFT *Short-Time Fourier Transform*
STO *Sampling Time Offset*
SVD *Singular Value Decomposition*
SVM *Support Vector Machine*
VAE *Variational Autoencoder*
WMA *Weighted Moving Average*
XGBoost *eXtreme Gradient Boosting*

General Introduction

Human Activity Recognition (HAR) has emerged as a pivotal field at the intersection of healthcare, security, and the *Internet of Things* (IoT), propelled by the growing capabilities of IoT and sensor technologies. This doctoral manuscript centers on advancing HAR through the innovative use of Wi-Fi *Channel State Information* (CSI), a promising but underexplored data source that offers privacy-preserving, device-free activity recognition capabilities. Wi-Fi CSI, with its sensitivity to environmental changes induced by human movements, provides a unique avenue for HAR without the privacy and practical limitations associated with cameras and wearable sensors. The manuscript is grounded in the development and evaluation of three novel contributions to the field, each aiming to address distinct challenges within CSI-based HAR: developing efficient and accurate activity recognizers, enhancing fall detection methodologies for emergency response healthcare service, and advancing identity recognition techniques leveraging unique behavioral patterns to facilitate personalized healthcare interventions.

1 Context and Motivations

HAR represents a dynamic and evolving field, intersecting with various disciplines to address the challenges of understanding and predicting human behaviors. These behaviors span from periodic movements indicative of vital signs, such as respiration and heart rates, to predefined actions and spontaneous activities observed in indoor and residential environments. The significance of HAR has burgeoned over the years, driven by its potential to revolutionize sectors like healthcare, security, robotics, and the IoT. Specifically, in residential healthcare, HAR solutions have emerged as pivotal tools for monitoring vulnerable groups, children, elderly, patients, and disabled, facilitating timely interventions in emergencies and aiding in accurate diagnosis with data-driven insights.

Traditionally, HAR research has been partitioned into vision-based, sensor-based, and wireless-technology-based approaches. Vision-based solutions, while powerful, face significant limitations in indoor settings due to privacy concerns, lighting conditions, and line-of-sight restrictions. Similarly, sensor-based approaches, despite their precision, often stumble upon practical challenges such as user compliance and cost considerations, especially with IoT-connected variants. These limitations underscore the need for non-intrusive, cost-effective solutions that respect user privacy and offer broad applicability.

To overcome these limitations, Wi-Fi-based HAR offers a device-free approach that leverages the omnipresence and connectivity range of Wi-Fi routers. Unlike other electromagnetic waves, Wi-Fi signals interact with the environment, absorbed, reflected, and scattered by obstacles, thereby encoding the movements of humans within the signal patterns. CSI, detailing these signal interactions, becomes a rich activity measurement technology for HAR. By capturing the nuanced changes in Wi-Fi signals caused by human activities, CSI facilitates the development of intelligent systems capable of recognizing and interpreting complex behaviors without intruding on personal privacy.

An emerging aspect of HAR that holds significant potential in healthcare applications is *Human Identity Recognition* (HIDR). While HAR focuses on recognizing and interpreting human activities, HIDR extends this capability by identifying individuals based on their unique behavioral patterns captured through sensor data. In the context of healthcare, integrating HIDR with HAR systems enables personalized patient monitoring and tailored interventions, which are crucial for improving patient outcomes and optimizing healthcare resources. By leveraging the same Wi-Fi CSI data used for activity recognition, HIDR can provide an additional layer of information without requiring intrusive biometric systems, thus enhancing the effectiveness and applicability of HAR solutions in personalized healthcare settings.

The advancement of HAR, particularly in emergency contexts like fall detection, has significant public health implications. Falls, especially among the elderly, constitute a major health risk, leading to high morbidity, mortality, and substantial economic costs. The World Health Organization identifies falls as a leading cause of accidental injuries and fatalities, emphasizing the critical need for effective fall detection and rapid intervention systems. Incorporating HAR into emergency response healthcare service can dramatically enhance the speed and efficacy of medical interventions, mitigating the severity of injuries and potentially saving lives. Furthermore, integrating HIDR can facilitate personalized interventions by identifying individuals during critical events, enhancing the effectiveness of emergency responses. The integration of sophisticated HAR methodologies, including Wi-Fi CSI and *deep learning* (DL), into fall detection systems presents a promising avenue for addressing this global health challenge.

Our research is motivated by the dual objectives of advancing HAR methodologies and addressing the pressing need for efficient, non-intrusive fall detection systems, while also facilitating personalized healthcare interventions through HIDR. By harnessing the potential of Wi-Fi CSI and DL, we aim to develop solutions that not only elevate the state of HAR in healthcare but also contribute significantly to improving patient outcomes and emergency response strategies. This endeavor underscores a commitment to technological innovation in service of societal well-being, bridging the gap between theoretical research and practical, life-saving applications.

2 Problems and Challenges

While CSI-based HAR offers significant potential for enhancing healthcare applications and improving quality of life, the field faces two major problems that hinder its practical implementation and widespread adoption: **Data-Related Concerns** and **Model-Related Concerns**. Addressing these problems is crucial for developing robust, efficient, and generalizable HAR systems capable of operating effectively in real-world settings.

Data-Related Concerns: The first major problem revolves around the availability, quality, and diversity of data used for training and evaluating HAR models. CSI-based HAR often suffers from limited and imbalanced datasets, which pose significant challenges for model training and generalization. Limited data quantities restrict the model's ability to learn diverse patterns, leading to poor performance on unseen activities or environments. Class imbalance, where certain activities are overrepresented while others are underrepresented, skews the learning process and affects the model's accuracy on minority classes. Additionally, issues such as noisy data, missing or incomplete data frames, and lack of representativeness across various demographics and environments further complicate the training process. These data-related limitations impact the reliability, validity, and generalizability of HAR models, making it difficult to deploy them in diverse real-world healthcare settings where data conditions vary significantly.

Model-Related Concerns: The second major problem pertains to the design, development, and evaluation of models and algorithms capable of effectively processing CSI data for HAR tasks. CSI data is

high-dimensional and complex, requiring models that can handle its intricacies without incurring prohibitive computational costs. Challenges include hyper-parameter sensitivity, where models heavily depend on carefully tuned parameters to achieve optimal performance, making them less robust to variations in data or environments. Overfitting is another concern, where models perform well on training data but fail to generalize to new, unseen data. High model complexity and long training periods hinder the practical deployment of HAR systems, especially in resource-constrained or real-time applications. Additionally, the absence of effective post-processing mechanisms and inadequate evaluation processes can affect the accuracy and reliability of HAR methodologies. Limitations in processing techniques, such as difficulty handling long-term dependencies in data and reliance on sub-optimal algorithms, further impact the effectiveness and efficiency of HAR models. These model-related concerns necessitate the development of advanced models and algorithms that are both accurate and efficient, capable of generalizing across different users and environments while maintaining practicality for real-world deployment.

Addressing these two major problems is critical for advancing CSI-based HAR and realizing its potential in healthcare applications. Each contribution in this thesis is designed to mitigate specific limitations within these broader problems. Through innovative approaches in *data augmentation* (DA), model design, optimization techniques, and evaluation frameworks, this thesis aims to enhance the performance, reliability, and applicability of HAR systems in real-world healthcare settings.

3 Research Questions

The exploration of CSI-based HAR within the healthcare sector is set to revolutionize how we monitor and interact with vulnerable populations. However, to fully realize this potential, we must navigate through a landscape full of challenges, including the necessity of data abundance, ensuring real-time responsiveness, accommodating environmental and user variabilities, and capturing subtle behavioral differences for accurate identity recognition. Drawing inspiration from the outlined challenges, we articulate a set of research questions aimed at guiding this dissertation's inquiry and contributing to the advancement of CSI-based HAR.

Question 1: What strategies can be employed to enhance the generalization capabilities of HAR models in the face of limited and imbalanced datasets, and to capture subtle, user-specific variations in activities, particularly for HIDR?

Question 2: How can DL models be adapted to effectively generalize across the diverse environments and users encountered in CSI-based HAR, particularly in fall detection applications?

Question 3: In the context of real-time HAR applications, such as emergency fall detection, what model architectures and computational strategies ensure swift, accurate activity recognition without compromising performance?

Question 4: How can personalized HAR systems, capable of recognizing and distinguishing between users based on activity- and user-specific patterns, be developed and integrated within healthcare applications to facilitate tailored interventions?

By addressing these research questions, this dissertation aims to advance the field of CSI-based HAR, overcoming existing challenges to harness the full potential of this technology for healthcare. The goal is to contribute to the development of effective, efficient, and versatile HAR solutions that improve health outcomes and the quality of life for individuals globally.

4 Contributions of the Thesis

This thesis presents three principal contributions to the field of CSI-based HAR, with the third contribution extending into HIDR. Collectively, these contributions advance the development of lightweight models, sophisticated fall detection mechanisms, and innovative approaches to identity recognition utilizing CSI data. Each contribution is designed to mitigate specific limitations within the major data-related and model-related concerns in CSI-based HAR, thereby enhancing healthcare services.

The first contribution introduces HAR-LightCNN, a pioneering solution for HAR that capitalizes on the unique properties of CSI data to detect user activities in real-time. This approach is underpinned by a lightweight *Convolutional Neural Network* (CNN) architecture, specifically designed to balance computational efficiency with the need for effective activity recognition. Recognizing the limitations posed by small-sized datasets and class imbalance, a common challenge in CSI-based HAR, this contribution employs time-series DA techniques to enhance the model's generalization capabilities. Through rigorous evaluation on unseen testing data, HAR-LightCNN has demonstrated superior accuracy in recognizing single-user activities, establishing a new benchmark in the field and surpassing existing state-of-the-art methods.

The second contribution delves into the critical area of fall detection, proposing an *Advanced Human Fall Detection* (HFD-Adv) methodology that leverages Wi-Fi CSI with unprecedented accuracy and efficiency. The methodology is distinguished by its holistic four-stage approach, preprocessing, processing, enhanced-processing, and post-processing, each stage meticulously crafted to address the nuances of fall detection across varied scenarios. At the heart of this system lies ActiNet, a flexible CNN designed for adaptive HAR, augmented by cutting-edge strategies such as genetic algorithm-optimized weighted loss function, refined transfer learning for enhanced adaptability to new users, and an *Advanced Decision Mechanism* (ADM) sensitive to demographic variations. This comprehensive approach has been extensively validated, demonstrating a flawless *Fall Detection Rate* (FDR) and significantly outperforming existing fall detection methodologies.

The third contribution focuses on the emergent domain of HIDR, presenting an innovative methodology that leverages a transformer-based model, IdentNet, for identifying users based on their unique fall patterns captured through Wi-Fi CSI data. This approach addresses the challenges of intra-class similarity and long-term dependencies inherent in CSI data from falling activity. By utilizing the transformer's self-attention mechanism, IdentNet effectively captures subtle user-specific variations, enabling accurate user identification during falls. Extensive evaluations demonstrate that IdentNet outperforms CNN-based models and other state-of-the-art HIDR methods, showcasing its potential in applications demanding precise user identification. This contribution marks a significant advancement in personalized healthcare, facilitating tailored interventions and improving patient outcomes.

Collectively, these contributions significantly advance the use of CSI for HAR by mitigating critical data-related and model-related limitations through innovative solutions that enhance accuracy, efficiency, and versatility. Our research not only pushes the boundaries of what is technically possible with CSI-based recognition systems but also lays the groundwork for future research and development in this promising field.

5 Organization of the Manuscript

The rest of this manuscript is organized as follows:

Chapters 1 and 2 are dedicated to the state of the art, presenting the contextual aspect of our thesis: HAR through CSI and DL for healthcare services. These chapters lay the groundwork by introducing the foundational concepts of Wi-Fi CSI and its pivotal role in HAR, as well as providing a systematic review of the existing literature. Initially, in Chapter 1, we explore the fundamental principles of CSI, its characteristics, and its distinction from other activity measurement technologies in HAR, thereby framing the integration of CSI with DL techniques for enhanced healthcare services. Then, in Chapter 2, we present a systematic review of the existing literature, which enables us to identify and discuss the challenges and limitations that currently impede the broader adoption of CSI-based HAR solutions, while highlighting areas ripe for innovation and ultimately setting the stage for our contributions.

Chapter 3 details the first contribution of the thesis: the development of HAR-LightCNN. It articulates the architectural design and rationale behind creating a lightweight, efficient CNN tailored for real-time HAR using CSI data. Emphasis is placed on addressing the dual challenges of computational efficiency and the need for high accuracy in diverse environments. We outline the DA techniques employed to enrich training datasets and mitigate class imbalance issues. An extensive evaluation section showcases the model's performance, underscoring its superiority in recognizing a wide range of activities with minimal computational resources.

Chapter 4 focuses on the second contribution, introducing the comprehensive four-stage HFD-Adv methodology. From preprocessing to post-processing, each stage is meticulously designed to enhance the accuracy and efficiency of fall detection in varied settings. The chapter delves into the innovative aspects of the ActiNet CNN, including its adaptability through genetic algorithms and transfer learning. Detailed evaluations demonstrate the methodology's effectiveness, with discussions on its potential impact on emergency response service and patient outcomes.

Chapter 5 presents our final contribution, introducing a novel transformer-based approach to HIDR using Wi-Fi CSI data during falling activity. We detail the development of IdentNet, a transformer-based model that leverages self-attention mechanisms to capture subtle, user-specific variations in falling patterns. This approach addresses challenges such as intra-class similarity and long-term dependencies in CSI data, enhancing the accuracy of user identification. Through rigorous testing and benchmarking against state-of-the-art models, the chapter highlights IdentNet's superior performance and its potential for enhancing personalized healthcare services by facilitating tailored emergency interventions responding to fall incidents.

Finally, the General Conclusion and Perspectives synthesizes the research work described across the chapters, reflecting on the cumulative advancements made in the use of CSI for HAR within healthcare contexts. It outlines promising research perspectives that build on the foundations laid by this thesis, pointing to the future of CSI-based HAR in enhancing healthcare services.

Part I

STATE OF THE ART

Chapter 1

Fundamentals of Wi-Fi Channel State Information in Human Activity Recognition

Objectives

This chapter establishes the foundational elements of *Wi-Fi CSI-based HAR*, particularly for healthcare applications. By defining core HAR building blocks, examining why Wi-Fi CSI is chosen over other sensing technologies, and illustrating the basic principles of CSI signal propagation and measurement, it provides the essential groundwork that underpins the advanced methodologies developed in this thesis.

Chapter Outline

1 Introduction . . . **10**
1.1 Context and Definition of our Research Domain . . . 10
1.2 Chapter Purpose . . . 11
1.3 Organization of the Chapter . . . 12
2 Definitions and Evolution of CSI Technology . . . **12**
3 Temporal Dynamics in CSI and Its Role in HAR . . . **14**
4 CSI Characteristics and Its Application in HAR . . . **15**
4.1 Effects of Moving Objects Factors on CSI . . . 15
4.2 Distinctions between CSI and Other Activity Measurement Technologies . . . 17
5 Effects of Environmental Factors on CSI . . . **18**
6 Integration of MIMO and OFDM Technologies in CSI for HAR . . . **19**
6.1 MIMO Technology . . . 19
6.2 OFDM Technology . . . 20
6.3 Combined MIMO-OFDM Systems . . . 21
7 System Setup for CSI Data Collection . . . **24**
7.1 Hardware Requirements . . . 24
7.2 Software Tools . . . 25
7.3 Environments and Parameters . . . 25
7.4 Configuration and Collection Process . . . 25
8 Publicly Available CSI Datasets for Evaluation . . . **26**
8.1 First CSI Dataset . . . 26
8.2 Second CSI Dataset . . . 28
8.3 Third CSI Dataset . . . 28
9 Methodologies for Processing CSI Data in HAR . . . **30**
9.1 Data Filtering and Normalization . . . 31

9.2 Feature Extraction 33
9.3 Potential of Deep Learning Approaches 35
10 Conclusion **35**

1 Introduction

1.1 Context and Definition of our Research Domain

The **research field** of interest for this thesis is *Human Activity Recognition (HAR)*, which encompasses a broad spectrum of studies aimed at identifying and analyzing human movements through technological means. HAR is a crucial field that provides significant benefits across various applications, enhancing both technological advancements and quality of life. HAR has become a significant research field due to its diverse applications and the exponential growth in technology. In 2022 alone, over 500 research papers were published on HAR in top scientific journals, reflecting the field's rapid development. HAR solutions are extensively used in healthcare for monitoring patient activities, enabling early detection of health issues, and providing timely interventions [Arshad, 2022]. In smart home environments, HAR technology enables automation that adapts to the resident's behavior, thereby increasing energy efficiency and convenience. The global market for gesture recognition, a subset of HAR, was valued at USD 19.37 billion in 2023 and is projected to grow to USD 169.26 billion by 2032, with a compound annual growth rate of 27.1% [Insights, 2024]. This growth is driven by advancements in sensor technology and *machine learning* (ML). Additionally, the rising demand for smart healthcare solutions, enhanced user interfaces in consumer electronics, and improved safety and surveillance systems contribute significantly [Arshad, 2022; Insights, 2024]. Furthermore, HAR is pivotal in enhancing security through surveillance systems that can identify unusual behaviors or potential threats in real-time, significantly improving public safety measures [Arshad, 2022]. These statistics and applications underscore the transformative impact of HAR, making it an essential research field in the advancement of smart, responsive, and safe digital environments.

Through rigorous investigation and innovation, this study aims to extend the boundaries of HAR, improving accuracy and efficiency in activity detection and analysis. As we delve into the complexities of HAR, we focus on developing robust models that can effectively handle the diverse and dynamic nature of human activities, ensuring that our findings contribute significantly to both the academic field and practical applications. At the core of the HAR field lie **recognizable components** that detail specific human activities, including *spontaneous activities* like walking and sitting, *periodic movements* such as breathing and heartbeats, and *predefined actions*, for example, gestures used in sign language. These activities are essential for fulfilling diverse recognition purposes across various application fields. In healthcare, HAR solutions enable crucial applications like fall detection for enhancing the quality of emergency response service [Tan, 2018; Ge, 2023], often within *indoor* and *residential* care settings. In smart home automation, gesture recognition technologies allow for seamless control of home environments. Furthermore, communication assistance through HAR, such as recognizing sign language, supports individuals with speech and hearing impairments [Wang, 2018; Zou, 2017]. These detailed activities and their **recognition purposes** are defined as the **core objectives** of HAR, guiding the technological development and application strategies.

The development of HAR systems relies heavily on advanced *measurement and recognition technologies*. **Measurement technologies** employed include *cameras*, which provide detailed visual insights but may raise privacy concerns; *wearable sensors* like accelerometers and gyroscopes, which require user compliance and can be intrusive; and *environmental self-powered sensors* using Wi-Fi Channel State Information (CSI), which preserves privacy and ensures non-intrusiveness. **Recognition technologies** involve *ML algorithms*, including both supervised models like support vector machines and deep neural networks, and unsupervised methods for anomaly detection, as well as traditional *pattern recognition techniques* that identify structured patterns in activity data [Ge, 2023].

These *enabling technologies* are pivotal in operationalizing the recognition of activities and are chosen

based on the requirements set by the recognizable components and their intended purposes. The system's complexity and applicability are further defined by the **number of recognizable components** it can identify, ranging from *multi-activity recognition* in dynamic environments to *single-activity recognition* in specialized settings, and the **number of moving objects** it can track, from *single individual monitoring* to recognizing activities from *multiple individuals* simultaneously.

The practical application of HAR solutions, driven by these core objectives alongside the enabling technologies, manifests in scenarios where the impact of research is most tangible. In **application fields** such as *healthcare*, *smart home automation*, and *communication assistance*, these HAR solutions are strategically implemented, translating technological capabilities into practical, real-world benefits. This direct application highlights the profound effect of HAR systems in enhancing everyday life [Arshad, 2022 ; Ge, 2023 ; He, 2020].

Collectively, these foundational elements—recognizable components and their purposes (core objectives), measurement and recognition technologies (enabling technologies), and application fields—serve as the essential building blocks of the HAR research field. They illustrate how researchers progress from the overarching HAR domain to a more specific focus, thereby defining the domain of study for their work.

Building upon the profound relationship between HAR and healthcare highlighted earlier, this thesis pinpoints healthcare as the primary field of application. By directing development efforts towards addressing its specific needs and challenges, the focus shifts towards **"enhancing healthcare services"** through comprehensive patient *monitoring, emergency response, diagnostic support, and an extensive suite of services including telehealth, rehabilitative services, palliative care, health education, mental health support, and nutritional guidance* [Tan, 2018 ; Ge, 2023 ; Wang, 2018]. These applications not only enable real-time, data-driven interventions but also pave the way for predictive healthcare, offering a proactive approach to managing and mitigating health risks. Within this context, the choice to concentrate on recognizing **"spontaneous activities"** is driven by their direct relevance to the healthcare field, particularly those occurring *indoors* or in *residential* environments. Further refining our technological toolkit, **"CSI"** is selected as the measurement technology due to its capacity to sensitively and accurately capture the dynamics of human movement without the need for invasive or wearable sensors, thus promising a non-intrusive solution for continuous monitoring. Additionally, **"ML"** is chosen for data processing and classification because of its proven efficacy in extracting meaningful patterns from complex datasets, thereby enhancing the precision and utility of healthcare interventions. Our research, therefore, stands at the confluence of technology and healthcare, proposing CSI-based HAR methodologies utilizing ML to offer innovative solutions for healthcare. This alignment not only reflects current technological advancements but also meets the growing demand for sophisticated, non-intrusive healthcare monitoring systems, marking a significant step forward in the integration of technology into everyday health management practices.

By outlining these foundational elements, we reach the domain of **"Wi-Fi CSI-based HAR for healthcare"**. The subsequent sections of this chapter delve into the technical specifics of CSI, illustrating its capability to sense and interpret human activities accurately, thereby benefiting a diverse range of healthcare services.

1.2 Chapter Purpose

In this first chapter of the state of the art, we focus on the **fundamentals of Wi-Fi CSI** and its direct connection to HAR tasks, particularly in healthcare and often within *indoor* or *residential* settings, while also **defining and structuring the specialized domain** of our research. We begin by explaining the key building blocks of HAR research field, such as recognizable components (e.g., spontaneous

activities, periodic movements), recognition purposes, and enabling technologies (measurement and recognition), and illustrate how these elements guide us in determining our domain scope. We then define CSI by describing its amplitude and phase components, followed by an examination of how Wi-Fi signals propagate, experience environmental influences, and interact with human activities. This foundation is crucial for understanding how CSI can be leveraged to design robust, privacy-friendly, and non-intrusive HAR solutions.

1.3 Organization of the Chapter

The remainder of this chapter is structured as follows:

— Section 2 introduces the basic principles of Wi-Fi signals and illustrates how CSI captures both amplitude and phase changes.

— Section 3 explains how time-varying factors—like user motion or environmental shifts—impact CSI, providing avenues for HAR.

— Section 4 covers how human activities alter CSI, distinguishing it from other measurement technologies such as cameras or wearables.

— Section 5 highlights external influences (e.g., path loss, fading) and why amplitude typically exhibits the most noticeable activity-induced changes.

— Section 6 outlines how combining MIMO and OFDM refines our ability to capture subcarrier-level CSI, increasing data granularity for HAR tasks.

— Section 7 identifies practical devices and software tools needed to acquire CSI information.

— Section 8 describes well-known open datasets.

— Section 9 discusses best practices for filtering, feature extraction, and the potential of deep learning to handle raw CSI signals without manual intervention.

— Section 10 concludes the chapter with a brief summary of key insights and a lead-in to Chapter 2, which offers a comprehensive *systematic review* of existing CSI-based HAR methodologies.

Having established the foundational elements that define our Wi-Fi CSI-based HAR research domain and laid out these fundamental principles, the next chapter will systematically examine the state-of-the-art literature, linking the technical underpinnings of CSI to real-world methodologies and uncovering current research gaps in healthcare-oriented HAR.

2 Definitions and Evolution of CSI Technology

A Wi-Fi signal is an electromagnetic wave that functions as a medium to carry data between devices (transmitters and receivers) through the process of modulation of radio waves. This modulation involves altering properties (characteristics) of a carrier signal, represented mathematically as:

$$s(\tau) = A \cdot \cos(2\pi f_c \tau + \varphi)$$

where $s(\tau)$ is the signal at time τ, A is the amplitude of the signal, f_c is the carrier frequency, and φ is the phase of the signal. The amplitude, A, refers to the maximum height of the wave and determines the signal's strength or power; higher amplitudes mean stronger signals. In an ideal scenario, this amplitude remains constant; however, in real-world conditions, factors associated with both the surrounding environment and moving objects (e.g., path loss, shadowing, or scattering) can cause it to fluctuate. The phase, φ, represents the shift of the wave relative to a reference point in time and is crucial in defining

the wave's position at the initial moment of time $\tau_1 = 0$. Alterations in these properties (amplitude, frequency, or phase) encode the data intended for transmission. The encapsulated data is organized into packets, which are transmitted over a selected Wi-Fi channel within the chosen frequency band. These packets not only contain the payload or the actual data but also headers and footers with the necessary information for data reconstruction and error correction at the receiver's end. In essence, the Wi-Fi signal itself does not constitute the data but rather carries the data. It acts as a tool through which data, converted into digital format (bits), is modulated onto the carrier wave, transmitted across distances, and then demodulated back into its original form at the destination, ensuring the integrity and reliability of data communication over wireless networks [Proakis, 2008 ; Rappaport, 2024].

Wi-Fi operates on two primary frequency bands: 2.4 GHz and 5 GHz, each comprising multiple channels that facilitate communication between devices. In Wi-Fi technology, each channel is characterized by its bandwidth (B_c, also known as channel width) and center frequency. The bandwidth, B_c, denotes the range of frequencies the channel encompasses, reflecting the channel's capacity for data transmission. The center frequency, derived from the Wi-Fi operation bands (2.4 GHz or 5 GHz), serves as the focal point around which the channel's frequencies are allocated, ensuring precise and efficient communication. Specifically, the 2.4 GHz band typically includes a total of $N_c = 14$ channels, each spaced 5 MHz apart, with a standard channel bandwidth of 20 MHz. The frequency of each channel c in the 2.4 GHz band can be calculated as:

$$f_c = 2.4\,\text{GHz} + (c-1) \times 5\,\text{MHz}, \quad \text{for } c \in C = \{1, 2, \ldots, N_c\},$$

where f_c represents the center frequency of channel c. The 5 GHz band offers a broader spectrum, typically supporting up to 23 channels in many regions, with some regulatory domains permitting additional channels. These channels feature bandwidths (B_c) that can be 20, 40, 80, or even 160 MHz wide, facilitating higher data throughput rates and enabling more diverse and flexible network configurations. In the 5 GHz band, the channel frequency calculation is more complex due to the variable channel widths and non-linear channel numbering. However, the calculation generally follows a similar principle, using a base frequency and a channel spacing that varies based on regulatory domains. The choice of frequency band and specific channel directly impacts Wi-Fi network performance, influencing factors such as signal propagation, penetration through obstacles, and susceptibility to interference. While the 2.4 GHz band is more prone to congestion due to its fewer channels, it provides better range because of its longer wavelengths, making it suitable for broader coverage areas. In contrast, the 5 GHz band, with its higher number of channels, offers less interference and more bandwidth options but tends to have a shorter range due to its shorter wavelengths. This distinction underscores the importance of selecting the appropriate band and channel based on the specific needs and environment of the Wi-Fi network. Different Wi-Fi protocols operate on these bands, each offering variations in data transmission rates (R_t), range, and reliability. The 2.4 GHz band is primarily used by protocols like 802.11b/g/n, whereas the 5 GHz band is utilized by 802.11a/n/ac/ax protocols [Proakis, 2008 ; Rappaport, 2024 ; IEEE, 2021].

The Channel State Information (CSI) is a critical component in understanding the characteristics and performance of Wi-Fi communication links (channels), providing detailed insights into how signals propagate between a transmitter and receiver. CSI encompasses two fundamental components, amplitude and phase, each offering unique information about the channel's behavior and conditions [Zou, 2017]:

— **Amplitude Component**: The amplitude of CSI captures changes in the signal's strength or power as it traverses the channel, providing valuable information on various aspects such as signal attenuation, path loss, shadowing effects, and the impact of environmental obstacles. These insights are pivotal for assessing channel quality, estimating capacity, and adapting transmission parameters to optimize performance. Signal attenuation reflects the weakening of the signal due to distance,

materials, and interference, while path loss and shadowing reveal reductions in signal strength caused by obstructions. Additionally, changes in amplitude can indicate the presence and movement of objects within the environment, affecting signal propagation and, consequently, the overall channel conditions.

— **Phase Component**: The phase information within CSI is crucial for understanding the temporal dynamics of the channel, which refer to changes in channel characteristics over time due to factors such as the movement of the transmitter, receiver, or obstacles. The phase component captures the signal's phase change as it propagates, providing insights into the signal's time of flight and frequency response. This information is key for applications requiring precise location tracking, motion detection, and adaptive communication systems, which adjust parameters in real-time based on the channel's changing conditions.

While the amplitude (A) and phase (φ) of the Wi-Fi signal describe its fundamental wave properties, the CSI amplitude and phase components reflect how these properties are altered by the channel. In other words, the real-world fluctuations or shifts that occur as the signal propagates (due to various influencers, such as the surrounding environment and moving objects) are encapsulated in CSI, providing a more detailed view of how the influencers affect the signal. Moreover, the amplitude component offers a direct measure of how signal strength fluctuates over time, forming a vital part of the channel's temporal dynamics and serving as a practical indicator for detecting changes due to both environmental conditions and human activities.

3 Temporal Dynamics in CSI and Its Role in HAR

Temporal dynamics in the context described refer to the changes in the characteristics of a Wi-Fi communication channel over time. These changes can result from various influencers, including the movement of the transmitter or receiver, the presence of moving objects within the communication path, or changes in the environment surrounding the communication path. The movement of the transmitter or receiver involves changes in the physical location or orientation of either the transmitting or receiving devices. Key factors contributing to temporal dynamics due to this movement encompass path loss, fading, shadowing, scattering, and power decay with distance. The presence of moving objects within the communication path includes humans performing activities like walking, sitting, or falling. The primary factors in this scenario are obstruction, reflection, diffraction, and scattering. The changes in the environment surrounding the communication path comprise weather conditions, alterations in physical infrastructure, or the introduction of new obstacles. The specific factors associated with these environmental changes include path loss, fading, shadowing, scattering, and power decay with distance [Proakis, 2008 ; Rappaport, 2024 ; Chen, 2023a ; Ma, 2019].

Temporal dynamics, as captured by both amplitude and phase components, are central to several critical applications:

1. **Adaptive Systems**: These systems leverage knowledge of temporal dynamics to adjust communication parameters dynamically, enhancing efficiency and reliability.
2. **Motion Detection and Tracking**: By analyzing temporal changes, it is possible to detect and track movement within the communication path, useful in surveillance and navigation applications.
3. **Location Tracking**: Understanding signal propagation and its changes enables the determination of device or object locations, essential for indoor navigation and asset tracking.
4. **Improved Communication Reliability**: Awareness of temporal dynamics helps maintain reliable communication in dynamic environments by adapting to obstacles and changing conditions.

5. **Network Optimization**: In larger networks, insights into temporal dynamics facilitate optimized resource allocation, ensuring stable performance amidst environmental changes.

By integrating both amplitude and phase data, CSI offers a comprehensive view of the Wi-Fi channel's condition, enabling systems to adapt and optimize performance based on a detailed understanding of signal propagation, strength, and the impact of the various influencers. While phase often provides high-precision insights for fine-grained motion tracking, amplitude data is typically more robust and straightforward to interpret for general HAR applications [Yousefi, 2017; Diaz, 2023]. This integrated approach is essential for developing robust, efficient, and adaptive communication systems capable of operating effectively in diverse and changing conditions [Proakis, 2008; Rappaport, 2024; Chen, 2023a; Ma, 2019].

4 CSI Characteristics and Its Application in HAR

4.1 Effects of Moving Objects Factors on CSI

Using Wi-Fi CSI for HAR is a cutting-edge application. The rich data obtained from CSI allows systems to recognize a wide range of human activities without the need for wearable sensors. This has potential applications in healthcare (e.g., monitoring elderly or patients), security, smart home systems, and even in retail to track customer movement. The challenge lies in accurately interpreting the CSI data, which can be affected by various environmental factors, and in ensuring privacy and security in the use of this technology [Wang, 2018; He, 2020].

The intricate relationship between Wi-Fi technology and human activities, especially in detecting events such as falls, underscores the sophisticated interplay between wireless communication technologies and the monitoring of human behavior without the need for wearable sensors. This dynamic interaction is facilitated by analyzing characteristic changes in Wi-Fi signal patterns, which are significantly influenced by the presence and movement of humans. These changes are primarily captured through CSI, providing detailed insights into the amplitude and phase of Wi-Fi signals [Wang, 2018; He, 2020].

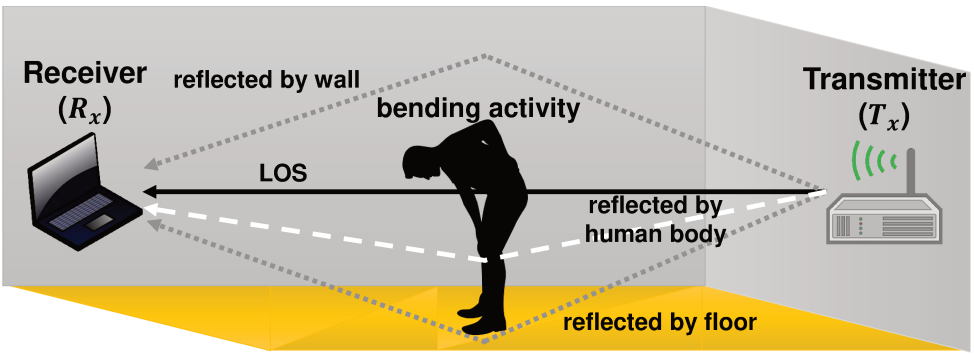


FIGURE 1.1 – Illustration of Wi-Fi signal propagation and reflection patterns in the presence of human activity. The figure shows how a Wi-Fi signal transmitted from a generic transmitter (T_x) to a receiver (R_x) is affected by reflections from walls, the floor, and the human body performing a bending activity, in addition to the direct *Line-of-Sight* (LOS) path. Here, T_x and R_x denote generic transmitter and receiver roles.

As illustrated in FIGURE 1.1, CSI represents the status of a communication channel in a Wi-Fi network, providing a detailed analysis of how a Wi-Fi signal propagates from the transmitter to the receiver and how it is affected by the presence of moving objects. The presence of moving objects, such as humans walking, sitting, or falling, induces specific alterations in the Wi-Fi signal's propagation

environment through factors like obstruction, reflection, diffraction, and scattering. These factors affect the signal's behavior in distinct ways when interacting with dynamic elements within the communication path, altering both phase and amplitude, the key characteristics of a Wi-Fi signal. Specifically, these interactions introduce phase shifts ($\Delta\varphi$) and amplitude changes (ΔA) in the signal, which together form the basis of the CSI phase and amplitude components. The amplitude of the signal can undergo attenuation or, in rare cases due to constructive interference, amplification. This variability leads to fluctuations in amplitude values. Similarly, the phase of the signal experiences shifts as a result of these interactions. Describing the effects of these factors is crucial in understanding how Wi-Fi signals are affected during propagation. Below is an overview describing their effects:

— **Obstruction**: The presence of moving objects can temporarily or permanently obstruct the signal path. This effect can be described as a reduction in signal strength determined by the object's opacity, where a higher opacity represents greater blockage (e.g., a wall) and a lower opacity represents minimal obstruction (e.g., clear air). The resulting signal loss or the need to find alternate paths underscores the direct impact of human presence on signal availability and quality, primarily affecting signal amplitude.

— **Reflection**: Reflection occurs when the signal encounters surfaces (e.g., walls, furniture) altered by human activity, leading to multipath effects. This phenomenon modifies both the signal's phase and amplitude due to the superposition of multiple signal paths, highlighting how human movements change the environment's reflective properties and thus influence signal behavior.

— **Diffraction**: The bending of the signal around the edges of objects, such as when humans move within the environment, introduces phase shifts ($\Delta\varphi$) and amplitude changes (ΔA). This can be approximated by the Huygens-Fresnel principle, where each point on a wavefront generates secondary wavelets that interfere to form the new wavefront. Diffraction's effect on the signal can thus be modeled as a function of the object's shape and the wavelength of the Wi-Fi signal, affecting both phase and amplitude in ways indicative of specific activities or spatial arrangements.

— **Scattering (for Moving Objects)**: Scattering becomes more complex with movement, as it dynamically alters signal paths and introduces variability in signal reception. For moving objects, this results in a continuous change in propagation characteristics, enriching the analysis of environmental and activity-related dynamics. It primarily affects signal amplitude and, to a lesser extent, phase.

Therefore, while both characteristics, amplitude and phase, of Wi-Fi signals are impacted, it is the amplitude that often shows more pronounced variations due to the presence of moving objects, such as humans. This prominence arises because human activities directly influence signal strength through various mechanisms. Obstruction and reflection can cause significant changes in signal amplitude by either blocking the signal path or redirecting it, thereby altering the signal strength received. Scattering and diffraction also affect amplitude, though in more complex and dynamic ways, especially with moving objects. The impact of these mechanisms on amplitude is typically more consistent and straightforward to observe, making it a critical factor for analyzing signal variations caused by human activities. Consequently, the amplitude of Wi-Fi signals undergoes substantial alterations as they interact with the environment, often more noticeably than phase changes, due to factors such as obstruction, reflection, and, to a lesser extent, scattering and diffraction. Nevertheless, phase information can provide finer-grained motion details, albeit requiring more careful calibration [Yousefi, 2017 ; Diaz, 2023]. This highlights the significant role of amplitude in the context of Wi-Fi-based HAR [Proakis, 2008 ; Rappaport, 2024 ; Chen, 2023a ; Ma, 2019].

4.2 Distinctions between CSI and Other Activity Measurement Technologies

The granular data provided by CSI enables a nuanced analysis of signal perturbations, facilitating the differentiation between subtle signal pattern variations that are essential for accurately identifying specific human activities. In contrast, other activity measurement technologies, such as cameras and sensors, have notable limitations that CSI overcomes, making it a superior choice for HAR.

Vision-based systems [Gupta, 2024; Zhang, 2018a], which use cameras to monitor human activities, can provide detailed visual insights and are powerful for motion tracking and recognition tasks. However, they come with significant drawbacks [Ge, 2023; Zou, 2017]:

— **LOS Requirement:** Vision-based systems require clear LOS, making them ineffective in occluded environments.

— **Lighting Conditions:** They depend heavily on good lighting conditions to function accurately.

— **Privacy Concerns:** Cameras are intrusive and raise substantial privacy issues, making users uncomfortable and less likely to accept such systems in personal or sensitive environments.

Sensor-based systems [Demrozi, 2020] employ various sensors, such as accelerometers, gyroscopes, and physiological sensors, often embedded in wearable devices. These sensors can provide precise data on body movements and are widely used in HAR [Ge, 2023; Zou, 2017]:

— **User Compliance:** Wearable sensors require users to wear them consistently, which can be inconvenient and affect data accuracy if users forget or choose not to wear them.

— **Intrusiveness:** The need to wear devices can be intrusive and uncomfortable for long-term use.

— **Environmental Constraints:** Data accuracy can be affected by external environmental factors and the physical positioning of the sensors.

Wireless technology-based systems [He, 2020], particularly those utilizing Wi-Fi signals, offer a compelling alternative for HAR. These methods leverage the ubiquitous presence of Wi-Fi routers and their capability to cover large areas without requiring additional infrastructure [Ge, 2023; Zou, 2017]:

— **Non-intrusive Monitoring:** Wi-Fi-based systems do not require users to wear any devices, ensuring a non-intrusive monitoring solution.

— **Privacy-friendly:** Compared to cameras, Wi-Fi-based solutions pose fewer privacy concerns, as they do not capture visual information.

— **Environmental Independence:** Wi-Fi signals are not dependent on lighting conditions and can penetrate walls, enabling through-wall detection and monitoring.

While *Received Signal Strength Indicator* (RSSI) measures the power level detected by a receiver, it has several limitations compared to CSI [Tan, 2018; Ge, 2023; Zou, 2017]:

— **Coarse Metric:** RSSI provides a coarse metric of signal strength, susceptible to environmental noise, and lacks the granularity needed for detailed activity recognition.

— **Limited Insight:** RSSI offers limited insight into the dynamic signal changes induced by moving objects, making it less reliable for complex activity recognition.

On the other hand, CSI offers several superior features that make it more effective for HAR [Tan, 2018; Ge, 2023; Zou, 2017]:

— **Detailed Channel Information:** CSI captures detailed characteristics of the wireless channel, including amplitude and phase information, which are sensitive to even slight movements.

— **Multipath Effects:** CSI can accurately capture multipath effects, where signals reach the receiver through multiple altered paths due to human movement. This capability allows for precise monitoring and recognition of human activities.

— **Robustness to Noise:** CSI demonstrates robustness to environmental noise, maintaining reliability in various real-world.

— **Comprehensive Data Analysis:** By providing fine-grained data about the wireless channel's state, CSI enables a deeper understanding of how human movements impact signal propagation, which is crucial for developing advanced wireless communication technologies for activity monitoring.

In summary, while Vision-based and Sensor-based systems have their applications, they face significant limitations in terms of privacy, user compliance, and environmental dependency. Wireless technology-based approaches, especially those utilizing CSI, offer a non-intrusive, privacy-friendly, and robust solution for HAR. CSI's ability to capture detailed channel information and multipath effects, along with its robustness to noise, underscores its effectiveness in accurately monitoring and understanding human activities through wireless communication technology.

5 Effects of Environmental Factors on CSI

In addition to the temporal dynamics induced by human activities, the surrounding environment also imposes key changes on the channel. CSI represents the status of a communication channel in a Wi-Fi network, providing a detailed analysis of how a Wi-Fi signal propagates from the transmitter to the receiver and how it is affected by the surrounding environment, through factors such as path loss, fading, shadowing, scattering, and power decay with distance. Describing the effects of these factors is crucial for understanding how the amplitude and phase of Wi-Fi signals are impacted during propagation. Below is an overview describing their effects:

— **Scattering:** Scattering occurs when a signal encounters irregularities in the propagation environment, causing the signal to scatter in different directions. This can lead to multipath propagation where the signal reaches the receiver through multiple paths, leading to interference and potential signal degradation [Proakis, 2008 ; Rappaport, 2024].

— **Power Decay with Distance:** Power decay with distance describes the reduction in signal strength as the distance between the transmitter and receiver increases. This phenomenon is a fundamental principle of physics, where signal strength decreases with distance, though real-world environments introduce additional complexities such as obstructions, reflections, and diffractions [Proakis, 2008 ; Rappaport, 2024].

— **Path Loss:** Path loss quantifies the reduction in signal strength as the signal travels through space. This factor, influenced by environmental conditions and distance, predominantly decreases the amplitude of the wireless signal, indicating that as the distance increases, the power received at the receiver becomes significantly lower [Proakis, 2008 ; Rappaport, 2024].

— **Shadowing:** Shadowing, also known as log-normal shadowing, represents the signal power variations caused by obstacles obstructing the signal path. These variations depend on the environment, such as urban or rural settings, and are influenced by the extent of obstacles in the signal's path. Shadowing causes unpredictable attenuation levels, leading to a non-uniform reduction in signal strength across different locations [Proakis, 2008 ; Rappaport, 2024].

— **Fading:** Fading refers to the variation in the strength of a radio signal caused by factors such as multipath propagation or the movement of objects. It affects both the amplitude and phase of the wireless signal. Fading can be classified into different types, such as fast fading, slow fading, Rayleigh fading, and Rician fading, each characterized by its unique environmental conditions and propagation effects [Proakis, 2008 ; Rappaport, 2024].

— **Rayleigh Fading:** Rayleigh fading occurs when there is no direct LOS path between the transmitter and receiver, and the signal reaches the receiver through multiple paths. This results in significant fluctuations in the amplitude and introduces random phase variations due to the interaction of multipath components.

— **Rician Fading:** Rician fading occurs when there is a dominant direct signal path (LOS) along with multiple scattered paths. The relative strength of the direct path compared to the scattered components significantly influences the variability in the signal's amplitude and phase. A stronger LOS component stabilizes the amplitude while still allowing fluctuations from the scattered paths.

The factors of path loss, shadowing, and fading, including both Rayleigh and Rician modes, have a pronounced effect on the amplitude of wireless signals in a Wi-Fi network. While the primary impact is observed in the amplitude, indicating how power levels fluctuate over distance and through various environmental conditions, phase alterations are also significant, especially under fading conditions. Rayleigh and Rician fading notably influence both amplitude and phase due to the complex interactions of multipath signals and, in the case of Rician fading, the presence of a direct LOS component. These effects underscore the challenges in maintaining signal integrity over wireless channels and highlight the importance of comprehensive CSI for optimizing Wi-Fi communication systems [Proakis, 2008; Rappaport, 2024; Chen, 2023a; Ma, 2019].

6 Integration of MIMO and OFDM Technologies in CSI for HAR

6.1 MIMO Technology

The advent of *Multiple Input Multiple Output* (MIMO) technology marks a significant leap in enhancing Wi-Fi performance. MIMO utilizes multiple antennas at both the transmitter and receiver ends, enabling simultaneous transmission and reception of multiple data streams. This innovation significantly improves throughput and signal range, effectively leveraging spatial diversity for better communication efficiency. The impact of MIMO on communication system capacity is quantitatively expressed by the following formula:

$$\text{Capacity (bps)} = \min(N_t, N_r) \log_2(1 + \text{SNR})$$

where N_t and N_r represent the number of transmitter and receiver antennas, respectively. The term $\min(N_t, N_r)$ indicates the minimum value between N_t and N_r, effectively determining the maximum number of simultaneous independent data streams that can be supported by the system. The Signal-to-Noise Ratio (SNR) quantifies the level of the desired signal relative to the level of background noise, with a higher SNR indicating better signal quality. SNR is crucial for evaluating and optimizing the performance of Wi-Fi communication systems. The SNR is mathematically expressed as:

$$\text{SNR (dB)} = 10 \log_{10} \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right)$$

where P_{signal} is the power of the signal and P_{noise} is the power of the background noise. Capacity, in this context, refers to the maximum data rate (in bits per second) that can be transmitted over a communication channel without error. It is a critical metric for assessing the efficiency and performance of wireless networks, directly influencing the quality and speed of data transmission. The formula encapsulates the essence of MIMO technology, showcasing how the utilization of multiple antennas at both

ends of the communication link can substantially increase the capacity of wireless networks by enabling parallel data transmissions, thereby enhancing the overall user experience with faster and more reliable connections [Proakis, 2008; Rappaport, 2024; IEEE, 2021; Chen, 2023a; Ma, 2019].

In a MIMO system with N_t transmitters and N_r receivers, the matrix H of dimensions $N_r \times N_t$ represents the channel gains between each transmitter and receiver pair. The matrix H at time τ, can be detailed as:

$$H(\tau) = \begin{bmatrix} h_{11}(\tau) & \cdots & h_{1N_t}(\tau) \\ \vdots & \ddots & \vdots \\ h_{N_r1}(\tau) & \cdots & h_{N_rN_t}(\tau) \end{bmatrix}$$

Here, each element $h_{mn}(\tau)$ in $H(\tau)$ corresponds to the channel gain from the nth transmitter (T_n) to the mth receiver (R_m). Each row in $H(\tau)$ corresponds to a different receiver (R_m) and each column in $H(\tau)$ corresponds to a different transmitter (T_n). This matrix $H(\tau)$ captures the complex interaction between multiple transmitters and receivers at time τ, including aspects like signal fading, path loss, shadowing, obstruction, and scattering, each represented by the $h_{mn}(\tau)$ terms. The matrix form allows for a concise representation of these interactions in the calculation of the received signal vector $\mathbf{r}(\tau)$ at time τ, as shown in the below formula:

$$\mathbf{r}(\tau) = H(\tau) \cdot \mathbf{s}(\tau) + \mathbf{n}(\tau)$$

where $\mathbf{s}(\tau)$ is the transmitted signal vector, representing the signals sent from all transmitters at time τ, and $\mathbf{n}(\tau)$ is the noise vector at the receivers, representing the noise present at each receiver. This type of noise is commonly referred to as "additive white Gaussian noise" (AWGN). AWGN is a model used in communications to mimic the effect of many random processes that occur in nature. The term "white" refers to the noise having a flat spectral density (i.e., it has the same intensity at different frequencies, similar to white light which contains all frequencies of the visible spectrum). The term "Gaussian" refers to the statistical properties of the noise; specifically, the amplitude of the noise follows a Gaussian (or normal) distribution [Proakis, 2008; Rappaport, 2024; IEEE, 2021; Chen, 2023a; Ma, 2019].

6.2 OFDM Technology

Orthogonal Frequency-Division Multiplexing (OFDM) technology enhances Wi-Fi efficiency through a digital multi-carrier modulation technique, dividing a wide frequency channel into numerous narrower sub-channels or subcarriers. This division allows OFDM to mitigate interference, enhance spectrum utilization, and is particularly effective in high-density settings [Proakis, 2008; Rappaport, 2024; IEEE, 2021; Chen, 2023a; Ma, 2019].

OFDM's ability to cope with severe channel conditions like multipath propagation and interference, without complex equalizers, is a significant advantage. The data stream in an OFDM system is distributed across N_s orthogonal subcarriers, each modulated by a low-rate data stream. This setup reduces inter-symbol interference (ISI) and optimizes the efficiency of spectrum usage by allowing the parallel transmission of data. Each subcarrier can be modulated with a conventional scheme (e.g., QAM or PSK), enhancing the signal's resilience against time dispersion effects caused by multipath. The orthogonality of subcarriers ensures minimal interference despite their close spacing, a crucial factor in maintaining high data rates in environments with limited bandwidth. This orthogonality is achieved by spacing subcarriers such that their frequency separation Δf is the reciprocal of the symbol duration, T_s, as expressed by:

$$\Delta f = \frac{1}{T_s}$$

Here, T_s represents the symbol duration, which corresponds to the time required to transmit one OFDM symbol, including the cyclic prefix. This duration ensures the subcarriers remain orthogonal, preventing overlap and inter-symbol interference [Proakis, 2008; Rappaport, 2024; IEEE, 2021; Chen, 2023a; Ma, 2019].

The total number of subcarriers, N_s, is a critical parameter in OFDM systems, influencing spectral efficiency and the ability to handle multipath interference and channel frequency selectivity. This number is determined by the channel's bandwidth and the desired subcarrier spacing, Δf, and is calculated using:

$$N_s = \frac{B_c}{\Delta f},$$

where B_c is the channel's total available bandwidth. For example, in a channel with $B_c = 20$ MHz and $\Delta f = 312.5$ kHz, $N_s = \frac{20\,\text{MHz}}{312.5\,\text{kHz}} = 64$. Similarly, in a channel with $B_c = 160$ MHz and $\Delta f = 78.125$ kHz, $N_s = \frac{160\,\text{MHz}}{78.125\,\text{kHz}} = 2048$. In Wi-Fi (IEEE 802.11 standards), N_s ranges from 64 in older standards to 2048 in newer ones like 802.11ax (Wi-Fi 6) [Proakis, 2008; Rappaport, 2024; IEEE, 2021; Chen, 2023a; Ma, 2019].

OFDM subcarriers are categorized into data subcarriers, carrying the actual user data; pilot subcarriers, used for channel estimation and synchronization; and null subcarriers, acting as buffers at the spectrum edges and between used bands. The exact count of data, pilot, and null subcarriers is defined by the Wi-Fi protocol specification (e.g., 802.11n) and by the CSI-extraction tool implementation (e.g., Linux 802.11n CSI Tool or Nexmon). This categorization facilitates optimal data decoding and easier signal processing. To calculate the frequency of each subcarrier (f_k) in relation to the channel's center frequency (f_c), we use:

$$f_k = f_c + \left(k - \frac{N_s}{2}\right) \cdot \Delta f, \quad \text{for } k \in K = \{1, 2, \ldots, N_s\},$$

ensuring subcarriers are symmetrically aligned around f_c, optimizing bandwidth utilization. The transmitted OFDM signal, $s(\tau)$, can be represented as:

$$s(\tau) = \sum_{k=0}^{N-1} X_k \cdot e^{i2\pi f_k \tau},$$

where X_k is the data symbol carried by the kth subcarrier, and τ is time. This expression highlights the efficient parallel data transmission and spectrum usage in OFDM systems. By transmitting data across multiple subcarriers at lower rates, OFDM maximizes spectral efficiency and minimizes the impact of multipath interference. This approach is especially beneficial in high-density areas, ensuring high data rates and effective spectrum utilization [Proakis, 2008; Rappaport, 2024; IEEE, 2021; Chen, 2023a; Ma, 2019].

6.3 Combined MIMO-OFDM Systems

MIMO and OFDM are crucial technologies in wireless communication, each enhancing performance and efficiency through their specific functions. MIMO uses multiple antennas at both the transmitter and receiver ends to increase system capacity and efficiency by enabling the simultaneous transmission and reception of multiple data streams. This exploits spatial diversity to boost throughput without needing more bandwidth or transmit power, making MIMO essential in Wi-Fi and cellular networks, including 4G LTE and 5G. On the other hand, OFDM, a digital multi-carrier modulation technique, improves spectral efficiency and reduces interference from multipath propagation by dividing a broad frequency channel into narrower sub-channels. These sub-channels, arranged orthogonally, carry parts of the data

stream, minimizing interference and optimizing bandwidth use. OFDM supports various standards, such as Wi-Fi, digital television, and cellular networks. The technologies operate on different layers; MIMO focuses on spatial signal transmission, while OFDM deals with signal modulation and organization within the frequency domain. When combined in wireless communication protocols like 802.11n/ac/ax (Wi-Fi 4/5/6), they maximize efficiency, capacity, and performance, underscoring their importance in advancing wireless communication infrastructure. Delving into their synergistic operation, OFDM's division of bandwidth into narrowband channels or subcarriers enables parallel data transmission, enhancing efficiency and reducing multipath interference. Meanwhile, MIMO leverages spatial paths, varied trajectories that radio waves take due to obstacles, to send and receive different data streams across the same frequency band but through distinct paths, significantly increasing data throughput. In MIMO-OFDM systems, the interaction between subcarriers and spatial paths highlights how parallel subcarrier transmission and multipath exploitation via MIMO greatly enhance the capacity, robustness, and efficiency of Wi-Fi networks, ensuring high data rates and resistance against interference [Proakis, 2008; Rappaport, 2024; IEEE, 2021; Chen, 2023a; Ma, 2019].

Initially, CSI estimation might have been performed for the entire channel. However, with the advent of OFDM, CSI can now be calculated for each subcarrier, enhancing the adaptability and efficiency of the network. Unlike the RSSI, which only measures signal strength, CSI offers a more granular view by detailing the amplitude and phase of the signal at different subcarriers. This includes capturing the impact of the influencers on the signal, making it a powerful tool for understanding the wireless environment [Proakis, 2008; Rappaport, 2024; IEEE, 2021; Chen, 2023a; Ma, 2019].

In a MIMO-OFDM system, the integration of spatial (multiple transmitters and receivers) and frequency (multiple subcarriers) domains significantly boosts communication efficiency and reliability. The system leverages a detailed CSI matrix for each subcarrier, enabling a nuanced representation of the channel characteristics across all spatial paths and frequencies. This rich representation is crucial for adapting transmission strategies in real-time, optimizing the system's performance under varying conditions [Proakis, 2008; Rappaport, 2024; IEEE, 2021; Chen, 2023a; Ma, 2019].

In a MIMO-OFDM system, the channel matrix H encapsulates the CSI across all transmitters and receivers, extended across the frequency domain to include all N_s subcarriers. Each matrix H_k for subcarrier k^{th} represents the channel gains between all pairs of transmitters and receivers, illustrating the intricate interaction within the system. The matrix for a specific subcarrier k, at time τ, can be detailed as:

$$H_k(\tau) = \begin{bmatrix} h_{11,k}(\tau) & \cdots & h_{1N_t,k}(\tau) \\ \vdots & \ddots & \vdots \\ h_{N_r1,k}(\tau) & \cdots & h_{N_rN_t,k}(\tau) \end{bmatrix}$$

Here, each element $h_{mn,k}(\tau)$ represents the CSI for the path from the nth transmitter (T_n) to the mth receiver (R_m) on subcarrier k^{th}, encapsulating both amplitude and phase information:

$$\text{CSI}_{mn,k}(\tau) = |h_{mn,k}(\tau)|e^{i\phi_{mn,k}(\tau)}$$

where i is the imaginary unit, $\text{CSI}_{mn,k}(\tau)$ denotes the CSI for the kth subcarrier from T_n to R_m, expressed in terms of the channel gain $h_{mn,k}(\tau)$ as a complex value. The amplitude $|h_{mn,k}(\tau)|$, the magnitude of the complex number $h_{mn,k}(\tau)$, represents the amplitude change (ΔA) of this specific subcarrier, quantifying how much the signal has been attenuated or amplified by the channel. This magnitude is a real and non-negative value, reflecting the signal's strength or power level as it is always measurable and quantifiable in the physical realm. This magnitude is calculated as:

$$|h_{mn,k}(\tau)| = \sqrt{(\text{Re}(h_{mn,k}(\tau)))^2 + (\text{Im}(h_{mn,k}(\tau)))^2}$$

where $\text{Re}(h_{mn,k}(\tau))$ and $\text{Im}(h_{mn,k}(\tau))$ are the real (in-phase component) and imaginary (quadrature component) parts of $h_{mn,k}(\tau)$, respectively. The phase $\phi_{mn,k}(\tau)$ is the phase shift ($\Delta\phi$) introduced by the channel for this subcarrier, representing the phase shift that the signal undergoes as it travels from the transmitter to the receiver. This phase shift is a real value, typically measured in radians, and falls within the range from $-\pi$ to π, indicating the angle of rotation in the complex plane. This phase shift, influenced by the propagation delay, reflections, and other channel effects, is given by:

$$\phi_{mn,k}(\tau) = \arctan\left(\frac{\text{Im}(h_{mn,k}(\tau))}{\text{Re}(h_{mn,k}(\tau))}\right)$$

This phase angle is typically measured in radians. This phase angle is essential for understanding the signal's behavior in the presence of noise and interference, utilizing the concepts of in-phase and out-of-phase components. By analyzing both the in-phase (real) and quadrature (imaginary) components, more data can be transmitted over the same frequency band, enhancing the efficiency of modern digital Wi-Fi communication systems and ensuring efficient use of bandwidth [Proakis, 2008 ; Rappaport, 2024 ; IEEE, 2021 ; Chen, 2023a ; Ma, 2019].

Consider a simple scenario to illustrate the above concepts: A wireless signal is transmitted over a channel with minimal interference and arrives at the receiver with a certain strength and phase shift. If the transmitted signal for a specific subcarrier has a frequency f_k and the environment causes an attenuation of 3dB and a phase shift of $\frac{\pi}{4}$ radians, the CSI for this subcarrier can be exemplified as follows:

— **Amplitude change (ΔA) Calculation:** The amplitude $|h_{mn,k}(\tau)|$ is calculated as 0.5. This is because an attenuation of 3 dB corresponds to a reduction in signal strength by a factor of 2. The logarithmic relationship of decibels to power ratios explains this phenomenon: a 3 dB decrease represents a halving of the signal's power, and since the amplitude of a signal is proportional to the square root of its power, the amplitude is thus reduced by a factor of $\sqrt{0.5}$, which is 0.5. This precise application of logarithmic principles aligns with the theoretical framework, where the magnitude of the channel gain reflects signal attenuation.

— **Phase Shift ($\Delta\phi$) Calculation:** The phase shift $\phi_{mn,k}(\tau)$ is $\frac{\pi}{4}$ radians. This shift is a direct measure of the phase alteration the signal undergoes due to channel effects, such as propagation delays and reflections, and is a real value within the typical range of phase shifts, representing the angle of the complex number describing the CSI for the subcarrier.

Thus, integrating both the reduction in signal strength and the phase shift caused by the channel, the CSI for this subcarrier could be represented as $\text{CSI}_{mn,k}(\tau) = 0.5 \cdot e^{i\left(\frac{\pi}{4}\right)}$. This example not only serves as a concrete instance of the theoretical framework described, demonstrating how CSI encapsulates both amplitude and phase information in a real-world scenario, but also elucidates the practical implications of signal attenuation expressed in decibels and its effect on the computed amplitude of the CSI [Proakis, 2008 ; Rappaport, 2024 ; IEEE, 2021 ; Chen, 2023a ; Ma, 2019].

Given N_t transmitters, N_r receivers, and N_s subcarriers, the overall channel information can be represented as a collection of matrices, one for each subcarrier:

$$\{H_1, H_2, \ldots, H_{N_s}\}$$

where each H_k for subcarrier k^{th} is an $N_r \times N_t$ matrix, emphasizing that each $h_{mn,k}$ within H_k corresponds to the channel gain from the nth transmitter to the mth receiver on the kth subcarrier. The dimensions

of the overall channel representation for a single subcarrier are $N_r \times N_t$. However, considering all N_s subcarriers, the complete channel information could be conceptualized as a three-dimensional structure of dimensions $N_r \times N_t \times N_s$, although it's more common to treat each H_k independently for analysis and processing due to the different channel conditions on each subcarrier [Proakis, 2008; Rappaport, 2024; IEEE, 2021; Chen, 2023a; Ma, 2019].

Channel conditions on each subcarrier critically influence the performance of Wi-Fi communication systems. Understanding the impact of these conditions, including amplitude attenuation and phase shift, is vital for applying appropriate strategies to enhance signal quality and system efficiency. One key metric for assessing signal quality is the SNR. The goal in Wi-Fi communication systems is to maximize SNR, which directly correlates to improved data transmission quality and efficiency. Phase and amplitude adjustments are instrumental in maximizing SNR. The signal's amplitude affects P_{signal} directly, while the phase alignment or correction can reduce interference effects, indirectly improving SNR. Mathematically, if a signal can be represented as $s(\tau) = A \cdot \cos(w\tau + \phi)$, where A is the amplitude, w is the angular frequency, and ϕ is the phase, optimizing both A and ϕ can enhance P_{signal} and thus SNR. With detailed CSI across all subcarriers and spatial paths, the system can adaptively apply:

— **Adaptive Modulation and Coding:** By adjusting modulation schemes and coding rates based on real-time channel conditions, systems can maximize throughput while maintaining a desired error rate. This adaptability is directly linked to optimizing SNR, as it allows for adjusting signal power and resilience based on noise and interference levels.

— **Beamforming:** This technique directs signal power towards the intended receiver, minimizing interference to others. Beamforming effectively increases P_{signal} at the receiver's end, which enhances SNR, especially in dense environments where interference from other sources is significant.

— **MIMO Spatial Multiplexing:** Utilizing multiple spatial streams increases data rates without requiring additional spectrum. By efficiently managing spatial resources, MIMO technology can improve P_{signal} for each stream, contributing to a higher overall SNR.

This adaptive approach, leveraging granular CSI, ensures that each subcarrier and transmitter-receiver pair is optimized for maximum SNR. By understanding and mitigating the channel's impact through strategic amplitude and phase adjustments, and applying advanced modulation, coding, and transmission techniques, the system optimizes the use of spectral and spatial resources, enhancing overall communication quality and efficiency [Proakis, 2008; Rappaport, 2024; IEEE, 2021; Chen, 2023a; Ma, 2019].

7 System Setup for CSI Data Collection

7.1 Hardware Requirements

The collection of CSI data necessitates specialized hardware capable of capturing detailed information about the wireless channel's state. This hardware setup consists of a Wi-Fi Access Point (AP) operating as the *transmitter (T_x)*, and one or more Monitoring Point (MP) devices equipped with CSI-capable *network interface cards* (NICss), functioning as the *receiver(s) (R_x)*. These devices are equipped with capabilities to monitor and report the state of the wireless medium, providing insights into the *physical layer* (PHY) characteristics of Wi-Fi communication.

Devices compliant with IEEE 802.11 standards are often preferred, as they can provide high-resolution CSI information, including amplitude and phase data for signals at different subcarriers. Examples of commonly used hardware include specific router models with custom firmware (e.g., OpenWrt) and NICs (e.g., Intel Wireless Cards) that enable CSI extraction.

7.2 Software Tools

The extraction and processing of CSI data are facilitated by specialized software tools designed to interface with the hardware. These tools enable the capture, decoding, and analysis of CSI data, providing valuable insights into wireless channel behavior. One widely used example is the Linux 802.11n CSI Tool, which operates at the PHY layer to extract detailed CSI information. This tool offers functionalities such as data logging, real-time analysis, and post-processing, supporting detailed evaluation of the wireless environment.

At the receiver side of a communication link, the hardware (router and NIC) serves as the interface between the device hosting the CSI extractor (laptop, desktop, or single-board computer like a Raspberry Pi) and the wireless network, while software tools like the 802.11n CSI Tool enable the extraction of detailed PHY information. Common hardware and software combinations for CSI data collection include:

— **Laptop/Desktop with Intel 5300 Wi-Fi NIC and Linux 802.11n CSI Tool:** This combination uses the Intel 5300 NIC, renowned for its ability to capture detailed CSI data [Halperin, 2011].

— **Laptop/Desktop with Atheros Wi-Fi NICs and Atheros-CSI-Tool:** This setup employs Atheros NICs with the Atheros-CSI-Tool for CSI extraction [Xie, 2015].

— **Raspberry Pi with compatible NIC (that supports CSI extraction) and Nexmon CSI Extractor:** A cost-effective and flexible solution using Raspberry Pi with a compatible NIC capable of extracting CSI data, alongside the Nexmon CSI Extractor [Gringoli, 2019].

These hardware/software combinations facilitate the collection of CSI data for various applications, such as activity recognition during daily activities like running, walking, and bending. The resulting raw CSI data is typically represented as multidimensional time series [Ma, 2019].

7.3 Environments and Parameters

CSI data collection can be conducted in diverse environments, ranging from controlled laboratory settings to real-world scenarios. Here, *environment* refers to the physical setting and surrounding conditions—such as layout, obstacles, interference sources, and human movement—within which CSI data is collected. The choice of environment significantly impacts the data collection process and parameter configurations.

— **Controlled Environments:** Laboratory settings allow for precise control of environmental variables and minimize external interference, enabling systematic data collection. Typical controlled spaces include signal-isolation *labs*, small *closed rooms*, or purpose-built *testbeds* equipped with fixed transmitter–receiver layouts. Parameters such as the distance between transmitter and receiver, the presence of obstacles, and subject movement patterns can be meticulously controlled [Zhuravchak, 2022].

— **Real-World Environments:** Deploying the system in real-world scenarios introduces variability, such as dynamic interference and non-stationary subjects. Examples include a *home* living room, a *residential hallway*, an *office* workspace, or a *hospital ward*—all representative of indoor healthcare and daily-living contexts. Robust hardware and software configurations are necessary to capture high-quality data despite environmental complexities [Demrozi, 2023].

7.4 Configuration and Collection Process

The configuration and collection process involves several critical steps to ensure effective CSI data gathering and analysis:

— **Hardware and Software Setup:** The initial step involves setting up the hardware, including router(s) and NIC(s), and installing appropriate software tools (CSI extractor) on the receiving device(s) (e.g., laptop, desktop, or Raspberry Pi). This setup establishes the foundation for effective data collection.

— **Parameter Configuration:** Key parameters such as channel frequency (f_c), bandwidth (B_c), and transmission rate (R_t) are configured to optimize data quality. This includes:

  — **Router and Receiver Pairing on a Specific Channel:** Ensuring that the router and receiver operate on the same channel for proper communication. For example, pairing a TP-Link Archer C20 router with a Raspberry Pi receiver on channel 36 in the 5 GHz band.

  — **Selection of a Frequency Band:** Choosing an appropriate frequency band (e.g., 2.4 GHz or 5 GHz) based on the application. The 5 GHz band, used by IEEE 802.11a, 802.11n, and 802.11ac standards, offers higher data transmission rates but shorter range.

  — **Subcarrier Distribution within the Channel Bandwidth:** Configuring the distribution of a channel's bandwidth across orthogonal subcarriers, a key aspect of OFDM-based communication. For instance, an IEEE 802.11n 20 MHz channel contains 48 data + 4 pilot + 12 null subcarriers, whereas an 80 MHz 802.11ac channel holds 234 data + 8 pilot + 14 null.

— **Data Collection:** After configuring hardware and software, the system records CSI data, capturing amplitude and phase information across subcarriers. This data reflects the impact of environmental factors and activities on the wireless channel.

Subsequent analysis involves processing the collected data to extract meaningful features that can be used for activity recognition, aligning with the primary purpose of CSI, which is to understand and adapt to the channel's current state as it affects signal propagation.

8 Publicly Available CSI Datasets for Evaluation

To evaluate the contributions presented in this thesis, we utilized three publicly available CSI datasets, referenced in the following sources: [Fard Moshiri, 2021], [Yousefi, 2017], and [Schäfer, 2021a]. These datasets comprise spontaneous activities performed in indoor environments, providing a robust foundation for assessing our proposed methodologies.

8.1 First CSI Dataset

For the first dataset described in [Fard Moshiri, 2021], the authors employed a combination of hardware and software specifically designed for CSI data collection. The setup consisted of a single-antenna receiver and a single-antenna transmitter ($N_r = 1$ and $N_t = 1$):

— **Receiver (MP)**: Raspberry Pi equipped with compatible NIC (that supports CSI extraction) and Nexmon CSI Extractor.

— **Transmitter (AP)**: TP-Link Archer C20 router paired on channel 36 with the Raspberry Pi over the 5 GHz band.

The specific configuration for this dataset involved several critical steps:

1. **Frequency Band and Channel Width**:
   — The 5 GHz band, employed by the IEEE 802.11a, 802.11n, and 802.11ac standards, was used.
   — Wi-Fi channels within this frequency band were ($B_c = 20$ MHz) wide.

2. **OFDM Subcarrier Distribution**: OFDM distributed the $B_c = 20\,\text{MHz}$ channel across 56 orthogonal subcarriers, of which 52 are *usable* (48 data + 4 pilot) and 4 are guard/null subcarriers, as captured by the Nexmon CSI extractor.
3. **Experimental Setup**: AP was placed 3 meters apart from MP, with each device positioned 1 meter above the ground in an indoor environment.

Using the Nexmon CSI Extractor, MP collected CSI amplitude data for seven daily human activities: bending, falling, lying down, running, sitting down, standing up, and walking. Three users of different ages performed each activity 20 times, with each trial lasting 20 seconds. The collected data was exported as CSV files. The dataset is structured as follows:

— **Number of Files**: 420 files in total (60 files per activity).

— **Sampling Rate**: 200 Hz, resulting in 4000 samples per 20-second signal.

— **Activity Duration**: Users remained inactive at the beginning and end of the period, with the activity occurring in the middle for a duration of 3 to 6 seconds. Based on video recordings and a stopwatch, the activity period was extracted by selecting only the CSI amplitude data corresponding to the performed activity, discarding inactivity (non-activity) periods. The extracted segment ranges between 600 to 1100 samples.

— **Data Format**: Each CSV file contained 52 columns and 600 to 1100 rows. The 52 columns represented the CSI amplitude data of the 52 data subcarriers captured by a single-antenna receiver (R_1), while the number of rows depended on the activity duration.

The dataset is publicly available on GitHub for further analysis [1]. An instance of the CSI amplitude data from the bending activity in this dataset is presented in FIGURE 1.2.

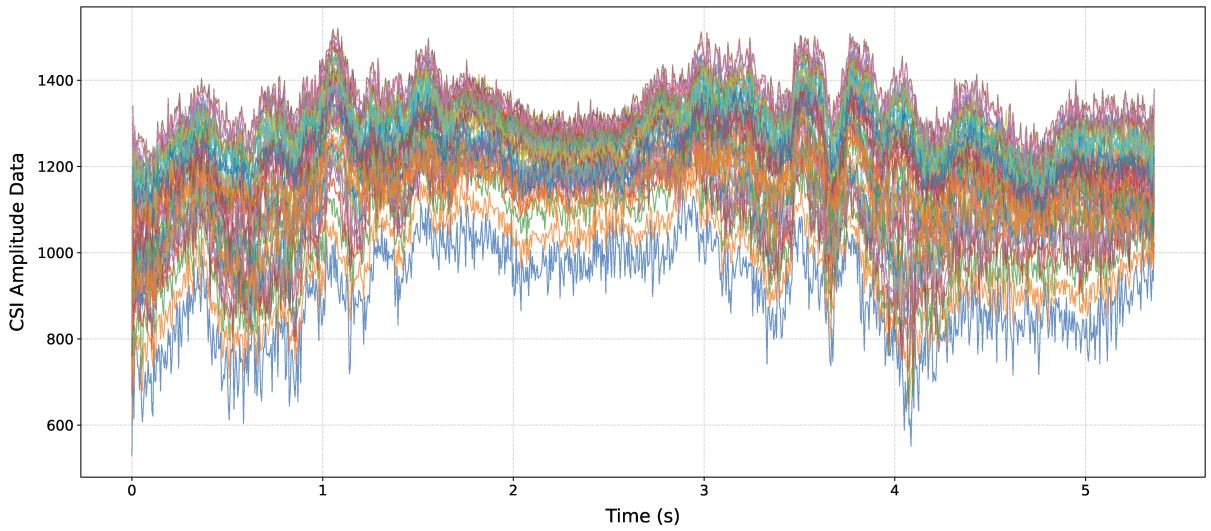


FIGURE 1.2 – This figure illustrates an instance of the CSI amplitude data collected from the bending activity in the 1st dataset [Fard Moshiri, 2021]. The data shows amplitude variations across 52 subcarriers over time, captured at a sampling rate of 200 Hz. The x-axis represents the time in seconds, while the y-axis indicates the amplitude of the CSI data. The figure demonstrates the multidimensional nature of the CSI data used for HAR.

1. Available at: https://github.com/parisafm/CSI-HAR-Dataset [dataset]

8.2 Second CSI Dataset

For the second dataset described in [Yousefi, 2017], the authors utilized a specific combination of hardware and software tailored for CSI data collection. The setup consisted of a three-antenna receiver and a single-antenna transmitter ($N_r = 3$ and $N_t = 1$):

— **Receiver (MP)**: A laptop equipped with an Intel 5300 Wi-Fi NIC and the Linux 802.11n CSI Tool. The NIC was equipped with three antennas ($N_r = 3$).

— **Transmitter (AP)**: A single-antenna transmitter ($N_t = 1$).

The configuration for this dataset involved several key steps to ensure effective data collection:

1. **OFDM Subcarrier Distribution**: OFDM distributed the link between each transmitter–receiver antenna pair across 56 subcarriers internally ; the Linux 802.11n CSI Tool down-samples these to 30 usable subcarriers by default for efficiency [Halperin, 2011].
2. **Experimental Setup**: The experimental environment was an indoor office area. AP was positioned 3 meters apart from MP, with each device appropriately placed to facilitate data collection.

Using the Linux 802.11n CSI Tool, MP collected CSI data for seven daily human activities: bed, falling, picking up, running, sitting down, standing up, and walking. Six users performed each activity between 10 to 20 times, with each trial lasting approximately 20 seconds. The collected data was exported as CSV files.

— **Number of Files**: 557 files in total, averaging about 80 files per activity.

— **Sampling Rate**: 1 kHz (1000 Hz), resulting in approximately 20,000 samples per 20-second signal.

— **Activity Duration**: Users remained inactive at the beginning and end of the period, with the activities occurring in the middle for a variable duration. Video recordings were used to label the signal's samples as either no activity or activity, and these labels were exported into separate annotation CSV files.

— **Data Format**: Each activity trial CSV file contained 180 columns and approximately 20,000 rows. The 180 columns represented the CSI amplitude and phase data of the 30 data subcarriers from the three receiver antennas (R_1, R_2, and R_3), while the number of rows depended on the trial's duration.

This dataset is also publicly available on GitHub for further analysis[2]. FIGURE 1.3 presents an instance of the CSI amplitude data from the falling activity in this dataset, illustrating both subcarrier-level variations (FIGURE 1.3a) and inter-receiver amplitude differences (FIGURE 1.3b).

8.3 Third CSI Dataset

For the dataset described in [Schäfer, 2021a], the authors utilized a comprehensive combination of hardware and software specifically tailored for CSI data collection. The setup included various configurations with different devices equipped with compatible NIC (that supports CSI extraction) and Nexmon CSI Extractor:

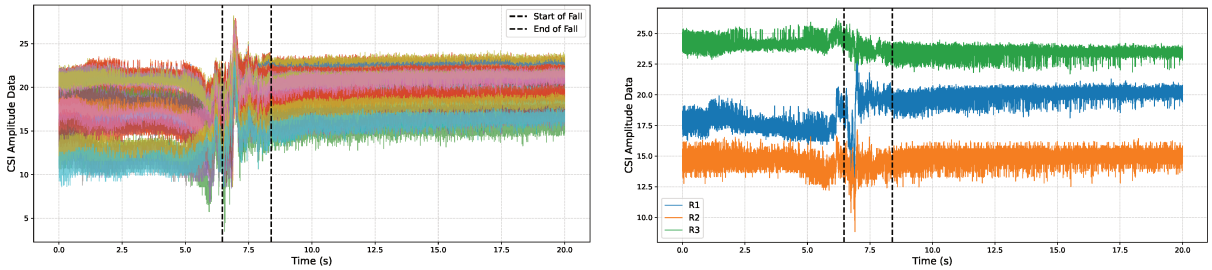
— **Receivers (MP)**: Raspberry Pi 3B+ and Raspberry Pi 4B, each equipped with compatible NIC (that supports CSI extraction) and Nexmon CSI Extractor.

— **Transmitters (AP)**: Fritzbox dual-band router (2.4 GHz and 5 GHz) and Asus RT-AC86U routers.

— Additional Devices: Intel NUC acting as a traffic generator.

The configuration for this dataset involved several key steps to ensure effective data collection:

2. Available at: https://github.com/ermongroup/Wifi_Activity_Recognition [dataset]



(a) (b)

FIGURE 1.3 – This figure illustrates an instance of the CSI amplitude data collected from the falling activity in the 2nd dataset [Yousefi, 2017]. The figure demonstrates the multidimensional nature of the CSI data used for HAR. (a) CSI amplitude variations across all 30 subcarriers for R_1 over time. (b) CSI amplitude variations for subcarrier 6 across all three receivers (R_1, R_2, R_3) to demonstrate inter-receiver amplitude level differences. The fall start and end points are marked by vertical dashed lines. The amplitude scale differs from other datasets due to hardware variations, as this dataset was collected using Intel 5300 NIC receivers, whereas the other datasets used Raspberry Pi-based receivers with compatible NIC (that supports CSI extraction) and Nexmon CSI Extractor.

1. **Frequency Band and Channel Width:**
   - The 5 GHz frequency band was primarily used due to its lower interference compared to the 2.4 GHz band.
   - Different channel widths were employed: $B_c = 20\,\text{MHz}, 40\,\text{MHz}$, or $80\,\text{MHz}$.
2. **OFDM Subcarrier Distribution:**
   - For $B_c = 40\,\text{MHz}$, OFDM distributed the channel across 128 subcarriers (108 data + 6 pilot + 14 null).
   - For $B_c = 80\,\text{MHz}$, OFDM distributed the channel across 256 subcarriers (234 data + 8 pilot + 14 null) ; theoretically 242 subcarriers (data + pilot) are usable, while the Nexmon CSI extractor records CSI for the full 256 in this dataset.
3. **Experimental Setup:**
   - Experiment 1: Conducted in a 4 m × 4.5 m room with four activities (EMPTY, STAND, SIT, WALK) in an LOS scenario.
   - Experiment 2: Conducted in a 3.5 m × 4.5 m room with six activities (WALK, SIT, STAND, SIT-DOWN, STAND-UP, FALL) performed in different parts of the room to avoid location dependence.
   - Experiment 3: Conducted in a 4.5 m × 5 m room with five activities (EMPTY, LYING, SIT, STAND, WALK) in an LOS scenario.
4. **Firmware Patching and Monitoring Mode:**
   - Raspberry Pi: The Nexmon patch for the BCM43455c0 NIC was used to enable monitoring mode. Due to instabilities in the latest patch, an older stable version (7_45_189) was used.
   - Asus RT-AC86U Router: The Nexmon CSI tool was used to extract CSI data from the BCM4366c0 Wi-Fi chip.

The data collection involved capturing CSI data for various human activities using the configured hardware and software:

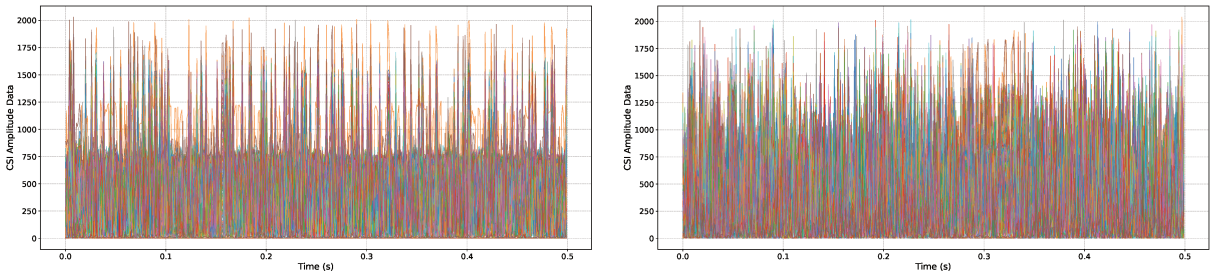
- **Activities**: Four to six different activities were captured in each experimental setup, including EMPTY, STAND, SIT, WALK, SIT-DOWN, STAND-UP, FALL, and LYING.

— **Sampling Rate**: Data was captured at different sampling rates depending on the setup, with some experiments capturing data at 1 kHz (1000 Hz).

— **Data Format**: Each captured packet contained CSI data for all subcarriers. For instance, at 40 MHz, data for 128 subcarriers was captured, with 114 subcarriers being useful for HAR on the Pi NIC.

— **File Format**: Collected data was exported into CSV files. Each CSV file contained columns representing the CSI amplitude and phase data of the subcarriers, and rows corresponding to the samples over time.

The dataset is publicly available on IEEE DataPort for further analysis[3]. FIGURE 1.4 presents instances of the CSI amplitude data from each of the empty non-activity and the standing up activity in this dataset, illustrating variations in CSI amplitude data across time and subcarriers.



(a) (b)

FIGURE 1.4 – This figure illustrates instances of the CSI amplitude data from the 3rd dataset [Schäfer, 2021a], representing: (a) CSI amplitude variations for the empty (non-activity) scenario, showing the baseline signal in the absence of human motion. (b) CSI amplitude variations for the standing up activity, capturing the signal changes induced by movement. The data shows amplitude variations across 256 subcarriers over time, captured at a sampling rate of 1000 Hz. Both signals demonstrate the multidimensional nature of CSI data for HAR, with the x-axis representing time in seconds and the y-axis indicating the amplitude of the CSI data.

9 Methodologies for Processing CSI Data in HAR

Wi-Fi-based HAR systems are designed to detect and monitor human activities, such as falls, by leveraging the subtle variations in Wi-Fi signals. These variations result from human activity influencing the signal's characteristics, namely its amplitude and phase. The key to capturing these minute changes lies in the CSI, which provides a granular view of how Wi-Fi signals' amplitude and phase change across individual subcarriers. This level of detail in CSI offers a comprehensive snapshot of the signal dynamics altered by human presence and activities. The raw CSI data can be structured to represent these variations in a detailed manner. Upon understanding how human activities can alter the intrinsic properties of Wi-Fi signals, it becomes crucial to analyze this information for recognition purposes. This is where CSI data becomes invaluable. The CSI does not merely reflect the amplitude and phase of Wi-Fi signals; instead, it encapsulates the variations in these parameters induced by human activities. It is this refined, activity-induced variation in the CSI data, spanning both amplitude changes (ΔA) and phase shifts ($\Delta\phi$), that we seek to explore for recognition tasks.

To effectively harness the insights contained within CSI data for activity recognition, we must extract pertinent features that are indicative of specific activities. Features refer to individual measurable

3. Available at: https://ieee-dataport.org/open-access/csi-human-activity [dataset]

properties or characteristics extracted from the data. In the context of CSI data, features might include statistical properties, frequency components, or changes over time that are indicative of human activity. It is important to distinguish that the features of interest are not the basic Wi-Fi signal characteristics (amplitude and phase) themselves, but rather the features extracted from the raw CSI data signals that reflect human actions' influence.

Given a MIMO-OFDM system with N_t transmitters, N_r receivers, and N_s subcarriers, the overall channel information can be conceptualized as a collection of matrices, each representing the CSI at a specific time for a specific activity across all subcarriers. This can be denoted as:

$$\mathrm{CSI}^{(\mathrm{act})}(\tau) = \{\mathrm{CSI}_1^{(\mathrm{act})}(\tau), \mathrm{CSI}_2^{(\mathrm{act})}(\tau), \ldots, \mathrm{CSI}_{N_s}^{(\mathrm{act})}(\tau)\}$$

where $\mathrm{CSI}_k^{(\mathrm{act})}(\tau)$ represents the CSI for subcarrier $k \in K = \{1, 2, \ldots, N_s\}$ at time τ during activity act, structured as an $N_r \times N_t$ matrix:

$$\mathrm{CSI}_k^{(\mathrm{act})}(\tau) = \begin{bmatrix} \mathrm{CSI}_{11,k}^{(\mathrm{act})}(\tau) & \cdots & \mathrm{CSI}_{1N_t,k}^{(\mathrm{act})}(\tau) \\ \vdots & \ddots & \vdots \\ \mathrm{CSI}_{N_r1,k}^{(\mathrm{act})}(\tau) & \cdots & \mathrm{CSI}_{N_rN_t,k}^{(\mathrm{act})}(\tau) \end{bmatrix}$$

Each element $\mathrm{CSI}_{mn,k}^{(\mathrm{act})}(\tau)$ encapsulates the CSI for the path from transmitter T_n to receiver R_m on subcarrier k, capturing the amplitude $|\mathrm{CSI}_{mn,k}^{(\mathrm{act})}(\tau)|$ and phase $\phi_{mn,k}^{(\mathrm{act})}(\tau)$ changes due to human activity act over time. The complete channel information for an activity act over a duration with discrete time steps $\tau_1, \tau_2, \ldots, \tau_{\mathcal{T}}$ can be visualized as a four-dimensional tensor:

$$\mathrm{CSI}^{(\mathrm{act})} = [\mathrm{CSI}^{(\mathrm{act})}(\tau_1), \mathrm{CSI}^{(\mathrm{act})}(\tau_2), \ldots, \mathrm{CSI}^{(\mathrm{act})}(\tau_{\mathcal{T}})]$$

with dimensions $\mathcal{T} \times N_r \times N_t \times N_s$, where $\mathcal{T}$ represents the total number of time steps recorded over the duration of the activity [Ma, 2019; Chen, 2023a] (see FIGURE 1.5).

To effectively recognize specific activities, the raw CSI data signal $\mathrm{CSI}^{(\mathrm{act})}$, derived from the CSI for each subcarrier over time and corresponding to each activity, is meticulously analyzed. Given the multi-dimensional nature of the data, with dimensions spanning time, receivers, transmitters, and subcarriers, this complexity necessitates deep temporal and spatial analysis. This involves extracting features from $\mathrm{CSI}^{(\mathrm{act})}$ that are indicative of the performed activity. The feature extraction process from the raw CSI data signals can be approached through two principal approaches: ML and its subset deep learning (DL). Both ML and DL approaches provide diverse strategies and sophisticated tools for processing CSI data, efficiently isolating meaningful features, and then conducting detailed temporal and spatial analysis on these features to accurately recognize a wide array of activities [Ge, 2023; Ma, 2019; Chen, 2023a].

The ML approach necessitates manual feature engineering to transform the raw CSI data signals into a more informative and interpretable format for the ML algorithms. This transformation process highlights the signal characteristics (features) essential for accurately recognizing activities such as falls, walking, or sitting. The process involves several key steps designed to refine the data and extract meaningful features that directly correlate with human movements and their impact on the Wi-Fi signal propagation. These steps include filtering, normalization, and, features extraction [Ge, 2023; Ma, 2019; Chen, 2023a].

9.1 Data Filtering and Normalization

Filtering is a crucial initial step in processing raw CSI data, aiming to remove noise and irrelevant information that could hinder the detection and monitoring of human activities. By applying digital filters, we can isolate signal components significantly influenced by human movement, making subsequent

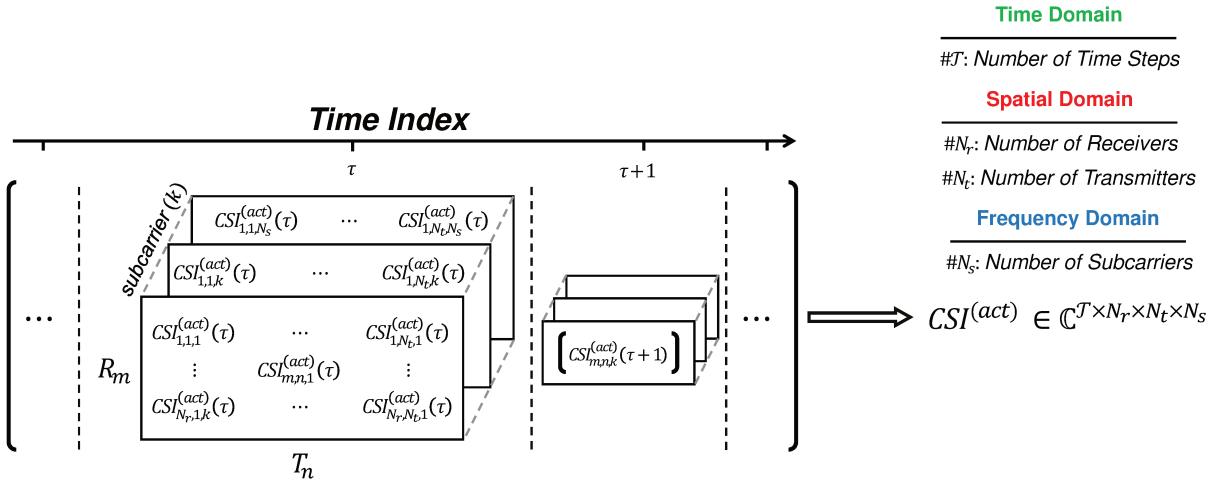


FIGURE 1.5 – CSI Data Tensor Structure for HAR: The 4D CSI tensor represents a time series of CSI matrices for MIMO-OFDM channels, capturing activity-induced multi-path channel variations, including amplitude attenuation (ΔA) and phase shifts ($\Delta\phi$). The tensor is structured across time ($\mathcal{T}$: Number of Time Steps), spatial (N_r: Number of Receivers, N_t: Number of Transmitters), and frequency (N_s: Number of Subcarriers) domains.

analysis more precise and effective. Various types of filters can be employed, each with specific use cases and benefits, tailored to enhance the signal's features relevant to activity recognition. Examples of these filters include:

— **Low-Pass Filters (LPF):** Ideal for removing high-frequency noise while preserving the baseline signal changes caused by human movements. In the context of CSI data, LPFs help in focusing on slow, continuous movements by filtering out rapid fluctuations in the amplitude and phase components that are not related to the intended activities.

— **High-Pass Filters (HPF):** Used to eliminate low-frequency drift or background noise, emphasizing sudden movements or changes in the signal that are more likely to correspond to activities. For CSI data, HPFs enhance the detection of quick, dynamic activities by removing slow-changing or static components in both amplitude and phase signals.

— **Band-Pass Filters (BPF):** Combines LPF and HPF to isolate a specific frequency range, filtering out noise and irrelevant information outside this range. In CSI-based activity recognition, BPFs are ideal for focusing on a particular band of frequencies that most accurately reflects human activities, balancing the removal of high-frequency noise and low-frequency drift in the amplitude and phase data.

— **Band-Stop Filters (BSF):** Used to remove frequencies within a certain range while keeping the rest of the spectrum intact. For CSI data, BSFs are useful for eliminating specific types of interference or noise known to occur within a certain frequency band, such as equipment-induced noise, while preserving the relevant activity-induced changes in the signal.

— **Moving Average (MA) Filter:** Smoothens the signal by averaging a number of points from the input signal to produce each point in the output signal. For CSI data, MA filters are good for reducing random noise in the amplitude and phase components while retaining the general trend of human activity signals.

— **Weighted Moving Average (WMA) Filter:** Similar to the MA filter but assigns different weights to the points being averaged, often giving more importance to recent points. In CSI-based

activity recognition, WMA filters can be more effective in reducing noise while still being responsive to more recent changes in the signal, which are critical for accurately capturing human activities.

— **Gaussian Filter:** Applies a Gaussian function to smooth the signal, useful for reducing noise while minimizing the loss of information in signal features. For CSI data, Gaussian filters provide a smooth, natural attenuation of frequencies, preserving the shape of activity-induced signal features better than a simple moving average.

— **Hampel Filter:** A robust filter designed to identify and correct outliers in the signal. In the context of CSI data, the Hampel filter replaces each point in the amplitude or phase signal with the median of neighboring points within a specified window, effectively removing spikes and sudden noise that are not related to human activities while preserving signal integrity.

— **Median Filter:** A non-linear filter that replaces each point in the signal with the median of neighboring points. For CSI data, the median filter is particularly effective at removing salt-and-pepper noise (sudden disturbances or outliers) in the amplitude and phase components while preserving the edges of the signal, which are important for detecting precise human activities.

— **Wavelet Filter:** Uses wavelet transforms to decompose the signal into different frequency components. In CSI-based activity recognition, wavelet filters allow for both time and frequency localization, making them particularly useful for analyzing non-stationary signals with varying frequency content. They are ideal for capturing transient features and filtering noise without losing significant signal information related to human activities.

Choosing the right filter depends on the specific characteristics of the noise and the activities being monitored. The objective is to enhance the signal components that are indicative of human activity while minimizing distortion and preserving the integrity of the signal for accurate activity recognition. In the context of CSI data from a MIMO-OFDM system, filtering can indeed be applied separately to amplitude $|CSI^{(act)}_{mn,k}(\tau)|$ and phase $\phi^{(act)}_{mn,k}(\tau)$ components to address specific types of noise or distortions that affect each differently. This distinction is crucial since amplitude and phase carry different information about the channel's characteristics and are affected by different environmental and system factors. Amplitude-specific filtering focuses on mitigating variations in signal strength that are not attributed to human activities. This can include static or dynamic changes in the environment, such as the presence of obstacles or varying transmission power. By employing these filters, the processing of raw CSI data becomes more accurate, enhancing the detection and monitoring of human activities through improved signal clarity and focus [He, 2020 ; Ge, 2023 ; Proakis, 2007 ; Vetterli, 2014].

Given the wide range of signal strengths and variations encountered in different environments, normalization or standardization of the CSI data is crucial. Normalization rescales the data to a common range, ensuring that the feature values have comparable magnitudes, which is vital for algorithms that rely on distance measures. This step helps in dealing with the variability in signal strength due to distance, transmitter power, and environmental factors, making the features more robust to external changes. On the other hand, standardization transforms the data to have a mean of zero and a standard deviation of one, aligning the statistical properties of the features. This is particularly necessary when dealing with data from different sources or when the phase and amplitude components have different distributions. By applying these techniques, we enhance the ML algorithm's ability to learn from the data effectively, improving the accuracy and reliability of activity recognition [He, 2020 ; Ge, 2023 ; Proakis, 2007].

9.2 Feature Extraction

The feature extraction step is crucial for identifying and isolating the most informative features (characteristics, attributes) of the CSI data that are directly related to human movements. By extracting

these features, we can enhance the ML algorithm's ability to accurately recognize activities such as falls, walking, or sitting. Here is a list of useful features commonly extracted from CSI data:

- **Energy Content:** This feature measures the total power of the signal over a period. It is calculated as the sum of the squared amplitudes of the CSI data across all subcarriers. Higher energy content often indicates more significant or intense movements, making it useful for detecting activities like falls.
- **Normalized Standard Deviation (STD):** This feature represents the variability of the signal relative to its mean. By normalizing the standard deviation, we can compare the variability across different signals, helping to identify activities that cause distinct variations in signal strength.
- **Median Absolute Deviation (MAD):** MAD measures the average absolute deviation of the signal from its median. This robust statistic is less affected by outliers and provides a reliable measure of the signal's dispersion, useful for distinguishing between different types of movements.
- **Interquartile Range (IQR):** IQR is the range within which the central 50% of the data lies. It is a measure of statistical dispersion and helps in identifying the spread of the signal, which can be indicative of the movement's intensity and variability.
- **Signal Shannon Entropy (SSE):** Entropy quantifies the randomness or complexity of the signal. Higher entropy values suggest more complex movements, such as falls, whereas lower values may correspond to simpler activities like sitting or walking.
- **Signal Strength Offset (SSO):** SSO represents the difference between the maximum and minimum signal strengths over a period. This feature helps in identifying abrupt changes in the signal, which can be associated with sudden movements like falls.
- **Signal Change Velocity (SCV):** SCV measures the rate of change in signal strength over time. Rapid changes in signal strength can indicate dynamic activities, making this feature useful for detecting quick movements or transitions.
- **Peak-to-Peak (PtP):** PtP measures the difference between the highest and lowest values of the signal within a given time window. This feature helps in detecting significant movements that cause large fluctuations in signal strength.
- **Time Duration (TD):** TD represents the duration of a specific activity or event. By measuring how long a particular pattern in the CSI data lasts, we can distinguish between activities of different lengths, such as a fall versus a brief movement.
- **Power Decline Ratio (PDR):** PDR measures the rate at which the signal power decreases over time. This feature can help in identifying activities that cause a rapid drop in signal strength, such as falls.
- **Volatility Number:** This feature quantifies the frequency and intensity of fluctuations in the signal. Higher volatility numbers indicate more dynamic and irregular movements, which can be useful for detecting complex activities.
- **Activity Fading Rate:** This feature measures how quickly the signal strength diminishes after a movement. A higher fading rate can be associated with abrupt stops, which are typical in fall events.
- **Median Percentage:** This feature calculates the median value of the signal as a percentage of its range. It provides a measure of central tendency that is less affected by outliers, helping to identify typical signal levels during different activities.
- **Cross-Correlation:** This feature measures the similarity between the observed CSI patterns and known patterns of specific activities, such as falls. By calculating the cross-correlation, we can match the real-time signal data with predefined templates to detect specific movements.

By extracting these features, the CSI data is transformed into a more informative and interpretable format for the ML algorithms. This enhances the sensitivity and specificity of the models in recognizing human activities, leading to more accurate and reliable detection and monitoring. Each feature contributes uniquely to capturing the nuances of human movement, thereby facilitating the detection of complex activities such as falls with greater precision [He, 2020 ; Ge, 2023 ; Chen, 2023a].

9.3 Potential of Deep Learning Approaches

The integration of sophisticated processing techniques in ML algorithms for Wi-Fi-based HAR systems underscores the pivotal role of manual feature engineering in refining signal features pertinent to HAR. This approach, involving meticulous preparation and transformation of CSI data, is designed to enhance the sensitivity and specificity of ML models to the subtle nuances of human movement, thereby facilitating the detection of complex activities such as falls with greater accuracy and reliability. Through the above steps, this methodology underscores the significance of transforming raw CSI data into a format that is both informative and interpretable for ML algorithms [He, 2020 ; Ge, 2023 ; Chen, 2023a].

However, the reliance on manual feature engineering within ML frameworks can introduce several limitations potentially affecting their efficacy for activity recognition. Firstly, the dependency on manually extracted features can lead to a potential overlook of complex data patterns not immediately discernible through traditional engineering techniques. This limitation is particularly pronounced in scenarios where human activities manifest in highly subtle or nuanced signal variations, which may escape the predefined criteria set by manual feature selection. Moreover, the effectiveness of manual feature engineering is intrinsically tied to the domain-specific knowledge of the engineers, necessitating a deep understanding of both the technical aspects of CSI data and the physical context of human movements. This requirement can constrain the adaptability of ML models to new or unforeseen types of activities, as the initial set of features may not fully encapsulate the diversity of human behaviors. Additionally, the labor-intensive nature of manual feature engineering poses scalability challenges, especially as the volume and complexity of data escalate. This approach can be time-consuming and may not be feasible for real-time or large-scale applications, where swift and dynamic adaptation to new data patterns is crucial [He, 2020 ; Ge, 2023 ; Chen, 2023a].

The DL approach offers a compelling alternative by leveraging the raw CSI data directly, enabling the autonomous learning of features that are indicative of specific activities without the need for manual intervention. DL models, through their deep neural network architectures, are capable of extracting high-level, complex features from raw data, which can reveal intricate patterns associated with different types of human movements. This ability not only overcomes the limitations posed by manual feature engineering but also enhances the model's capacity to generalize across diverse scenarios without the need for extensive domain-specific knowledge. Consequently, DL methodologies can adapt more fluidly to new data, making them particularly suited for the dynamic and varied nature of HAR [He, 2020 ; Ge, 2023 ; Chen, 2023a ; Goodfellow, 2016].

While manual feature engineering in ML models has been instrumental in advancing Wi-Fi-based HAR systems, its limitations underscore the growing importance of DL approaches. By directly analyzing raw CSI data, DL models offer a scalable, adaptable, and potentially more accurate framework for detecting human activities, heralding a new era in the development of activity recognition technologies.

10 Conclusion

This chapter has introduced the **fundamental principles** of Wi-Fi CSI and illustrated how amplitude and phase data, combined with technologies such as MIMO and OFDM, provide a powerful basis

for HAR. We have highlighted how human activities induce measurable fluctuations in Wi-Fi signals and examined the key mechanisms, temporal dynamics, environmental impacts, and subcarrier-level granularity, that make CSI well-suited for non-intrusive monitoring in indoor or residential environments and healthcare applications. Additionally, we have outlined the **foundational elements** of our research domain, clarifying how these fundamentals align with broader HAR objectives and ultimately guide our focus on healthcare solutions.

Key Insights and Importance:

- *Granular View of Wi-Fi Signals :* By capturing fine-grained amplitude and phase shifts across multiple subcarriers, CSI offers deeper insights than simpler metrics like RSSI.
- *Non-intrusiveness and Privacy:* CSI-based systems can sense human movement without wearable sensors or cameras, allowing for *continuous* yet privacy-preserving healthcare monitoring.
- *Adaptability to Complex Environments:* Integration with MIMO and OFDM technologies helps isolate the influences of environmental factors and user motion, maintaining robust activity detection despite multipath propagation or noise.

Roadmap to Chapter 2: While these fundamentals lay a strong technical foundation, implementing CSI-based HAR in practice requires deeper analysis of existing methodologies. Therefore, the following chapter (Chapter 2) will *systematically review the literature on CSI-based HAR*, identifying key challenges, limitations, and performance aspects. By bridging these fundamental insights to contemporary research, we will uncover the gaps that motivate this thesis's contributions and guide the design of robust HAR solutions for healthcare scenarios.

Chapter 2

Systematic Literature Review of CSI-Based HAR for healthcare domain

Objectives

This chapter systematically surveys the state of the art in Wi-Fi CSI-based HAR, focusing on healthcare-oriented scenarios. Through a structured mapping of existing works, it classifies the key limitations and research gaps—both data-related and model-related—thereby guiding the design and evaluation of the novel contributions proposed.

Chapter Outline

1 Introduction . . . **38**
1.1 Chapter Purpose . . . 38
1.2 Selecting a Suitable Literature Review Approach . . . 38
1.3 Organization of the Chapter . . . 39
2 Systematic Mapping Approach: Search, Screening, and Classification . . . **40**
3 Narrative Review of the Selected Articles . . . **46**
3.1 Multi-activity Recognition–Focused Approaches . . . 47
3.2 Fall Detection–Focused Approaches . . . 47
3.3 Summary and Next Steps . . . 49
4 Key Challenges and Limitations in CSI-Based HAR . . . **50**
4.1 Data-Related Concerns . . . 50
4.2 Model-Related Concerns . . . 51
5 Performance Aspects for CSI-Based HAR Methodologies . . . **52**
5.1 Effectiveness of the Methodology . . . 53
5.2 Efficiency of the Methodology . . . 53
5.3 Versatility of the Methodology . . . 54
6 Potential Healthcare Services Benefiting from HAR . . . **55**
7 Synthesis and Interpretation of the Mapped Literature . . . **56**
7.1 Classification Scheme Description . . . 56
7.2 Illustrative Example of Table Entry Filling . . . 57
7.3 Observed Trends and Common Limitations . . . 57
7.4 Bridging Identified Gaps to Our Contributions . . . 58
8 Conclusion . . . **59**

1 Introduction

1.1 Chapter Purpose

In the previous chapter, we introduced the *fundamental principles* of Wi-Fi Channel State Information (CSI) in Human Activity Recognition (HAR), defined our *specialized research domain*, and highlighted its potential, particularly in healthcare applications, often within indoor or residential settings.

Establishing a solid state of the art is fundamental to any research endeavor [Snyder, 2019]. To deepen our understanding of the defined domain of study, this second chapter (of the state of the art part I) systematically examines relevant literature, identifying key trends and research gaps that guide the design and evaluation of robust CSI-based HAR solutions.

1.2 Selecting a Suitable Literature Review Approach

This subsection explains our choice of literature review approach, specifically a *Systematic Mapping Study* (SMS), and how it aligns with our objective of gaining a broad overview of CSI-based HAR solutions, especially those that could be adapted to healthcare scenarios.

1.2.1 Overview of Literature Review Approaches

Various approaches exist for conducting literature reviews, each with specific goals and methodological rigor [Grant, 2009]. Generally, these approaches can be grouped into three broad categories:

- **Narrative Reviews (Traditional Reviews)**: Provide a descriptive summary of prior studies without enforcing strict screening or selection criteria. They are useful for offering a general (but not necessarily comprehensive) overview of a field. However, due to limited systematic guidelines, narrative reviews can be prone to selection bias and may not guarantee reproducibility [Ferrari, 2015 ; Grant, 2009].
- ***Systematic Reviews* (SR)**: Follow structured guidelines or "protocols" to search for, select, analyze, and synthesize literature in a transparent and reproducible manner. These reviews aim to answer specific research questions in depth, making them highly valued in disciplines like evidence-based medicine and software engineering. Within SR, several research methods exist (e.g., *Systematic Literature Review* (SLR), meta-analysis, meta-synthesis), each focusing on detailed quantitative or qualitative synthesis [Chakraborty, 2024 ; Grant, 2009].
- **Scoping Reviews**: Aim to map out the extent, range, and nature of available research on a broad topic without necessarily appraising study quality or delving into in-depth synthesis. Scoping reviews are often used to identify gaps in the literature or clarify complex concepts. In some references, systematic mapping is considered a variant or extension of scoping reviews, given its broad exploratory perspective but more formal classification procedures [Arksey, 2005 ; Grant, 2009].

While all these approaches offer value, we seek to establish a structured, high-level map of CSI-based HAR before zooming in on narrower, specific questions. As explained next, this motivation aligns strongly with what a SMS can provide.

1.2.2 Motivation for a Systematic Mapping Study

SMS is a systematic but broad review approach that categorizes research, identifies trends, and highlights gaps, often with less emphasis on deeply synthesizing each study. In some classifications, SMS is considered separate from (yet conceptually related to) scoping reviews, because it employs methodical steps (e.g., well-defined search, inclusion/exclusion criteria, classification schemes) to ensure clarity and reproducibility [Petersen, 2015 ; Grant, 2009 ; Munn, 2018].

Broad Landscape, Not Narrow Questions. In this thesis, our initial goal is to understand the domain of CSI-based HAR in healthcare from a wide-angle perspective:

— We do not yet have a singular, tightly framed research question that warrants deep, point-by-point synthesis.

— We are interested in mapping the landscape of existing methodologies, pinpointing common challenges and possible underexplored areas, particularly spontaneous activities like falling or walking.

A full-fledged SLR (one of the more detailed methods under the SR umbrella) generally focuses on thoroughly answering specific, pre-defined questions (e.g., "Which ML models achieve the highest accuracy for CSI-based fall detection?") [Petersen, 2015 ; Grant, 2009 ; Munn, 2018]. Conducting such an in-depth synthesis can be premature when the overall structure and scope of the domain is not yet fully mapped.

Categorizing Research for Healthcare Adaptation. The healthcare field encompasses diverse applications, from emergency response to patient monitoring. By employing an SMS, we can [Petersen, 2015 ; Grant, 2009 ; Munn, 2018]:

— **Catalog existing studies** in Wi-Fi CSI-based HAR, even when healthcare is not their primary focus, and observe how they might be adapted to health-oriented use cases.

— Identify both *technical innovations* and *methodological frameworks*, clarifying their respective strengths and limitations in a broad sense (rather than deeply evaluating each one's empirical outcomes).

— **Spot underrepresented topics** or *challenges* (for instance, lack of diverse datasets or limited post-processing mechanisms) which guides where our research can contribute most effectively.

Establishing a Research Agenda. An SMS approach allows us to structure the research field into high-level categories—for example, multi-activity recognition vs. fall detection—and note where and how healthcare use cases arise across these categories. Once this mapping is complete, we can decide if a follow-up review method (e.g., a targeted SLR) is needed for a specific gap [Petersen, 2015 ; Grant, 2009 ; Munn, 2018].Consequently, the SMS provides a strategic foundation for the subsequent contributions in this thesis, ensuring that our interventions or solutions address genuine, documented gaps in the literature.

Integrating Systematic Mapping with Narrative Synthesis. While our primary method is an SMS to categorize research comprehensively, we also adopt a narrative-style review of the final set of articles (Section 3) to qualitatively synthesize each study's focus. Hence, our approach can be seen as a *hybrid research method*, combining broad systematic mapping with narrative interpretation. In addition to this hybrid literature review approach, we also provide a more in-depth, interpretative analysis of the mapped articles (Section 7), thereby contextualizing their contributions, limitations, and emerging research gaps.

1.3 Organization of the Chapter

Having justified the selection of an SMS, the remainder of this chapter describes how the approach was carried out and how the resulting mapping aids our research focus:

— Section 2 applies our systematic mapping approach, outlining the search strategy, inclusion/exclusion criteria, and classification scheme used to identify and organize the final set of articles.

— Section 3 offers a narrative overview of the selected articles, categorizing each by datasets, recognizable components, and recognition methodologies. This establishes a baseline for understanding how different approaches handle data collection, preprocessing, processing, and (where applicable) post-processing.

— Section 4 compiles key challenges and limitations that we observe across the surveyed works.

— To guide our comparative analysis in later chapters, Section 5 introduces three essential performance aspects (*Effectiveness*, *Efficiency*, and *Versatility*) that capture a methodology's overall suitability for healthcare-oriented applications.

— Section 6 highlights a range of healthcare services that stand to benefit from robust HAR systems using CSI, illustrating the far-reaching potential of these solutions.

— Section 7 provides a synthesis of the mapped literature, presenting a detailed classification scheme, interpreting trends, and identifying gaps in current state-of-the-art methodologies.

— Section 8 concludes the chapter by summarizing the major takeaways from our systematic review, emphasizing the key limitations and open challenges in CSI-based HAR for healthcare. It also establishes a direct connection to the forthcoming contributions of this thesis, outlining how *Chapters 3, 4, and 5* will address select data- and model-related challenges identified in our review.

Together, these sections form a comprehensive understanding of the state-of-the-art literature and identify potential avenues for future research in the vital **Wi-Fi CSI-based HAR for healthcare** domain. By the close of this chapter, we gain a clear vantage point for directing our original contributions addressed in the following part II of this manuscript.

2 Systematic Mapping Approach: Search, Screening, and Classification

In the evolving field of HAR, the use of Wi-Fi CSI is increasingly recognized as a versatile and powerful domain. While HAR applications span various sectors, this research emphasizes the potential integration of CSI-based HAR methodologies in healthcare, particularly for applications like emergency response and patient monitoring. To capture the broad spectrum of advancements and potential applications, we adopted a Systematic Mapping Study (SMS) literature review approach. This approach not only identifies literature directly related to healthcare but also examines studies focused on the broader capabilities of CSI-based HAR, with a particular focus on spontaneous activities such as falling. Each study is analyzed to discern how the innovations and methodologies might be adapted to enhance healthcare services, even when the primary focus is not explicitly healthcare-oriented. This comprehensive mapping has facilitated the development of a detailed categorization table (TABLE 2.1), which includes potential healthcare applications derived from Wi-Fi CSI-based HAR advancements. The SMS approach allowed us to gain a systematic overview of the domain, identify key trends, and highlight areas requiring further research [Petersen, 2015]. The steps undertaken in our SMS are as follows:

1. **Research Questions:**
   — **RQ1:** What are the state-of-the-art methodologies used in "Wi-Fi CSI-based HAR" domain of study and might be adapted for enhancing healthcare services?
   — **RQ2:** What are the common limitations and challenges faced in these studies?
   — **RQ3:** How do the methodologies perform in terms of effectiveness, efficiency, and versatility?

2. **Search Strategy:** The development of a comprehensive search strategy is crucial for identifying relevant studies within the domain of Wi-Fi CSI-based HAR. This strategy involved selecting databases, defining search strings, and establishing preliminary inclusion/exclusion criteria to manage the scope of the search results effectively.

   For our literature search, we selected Scopus due to its extensive coverage of academic literature across various disciplines. In the beginning, an initial broad search was conducted using HAR terms to capture the broad domain of HAR. This search was carried out using Scopus's TITLE-ABS-KEY search function, covering the period from 2010 to 2023. The following keywords were used:

   > "Human Activity Recognition" OR "HAR" OR "Activity Recognition" OR "Activity Inference" OR "Human Motion Detection" OR "Motion Detection" OR "Human Activity Detection" OR "Device-Free Human Behavior Recognition" OR "Human Behavior Recognition" OR "Behavior Recognition" OR "Motion Recognition" OR "Human Fall Detection" OR "HFD" OR "Fall Detection" OR "Fall Prevention" OR "Fall Monitoring".

   This broad search yielded 56,572 results. These terms ensured the inclusion of relevant studies across various sectors. FIGURE 2.1 represents the various sectors (application fields) HAR applications span between 2010-2023.

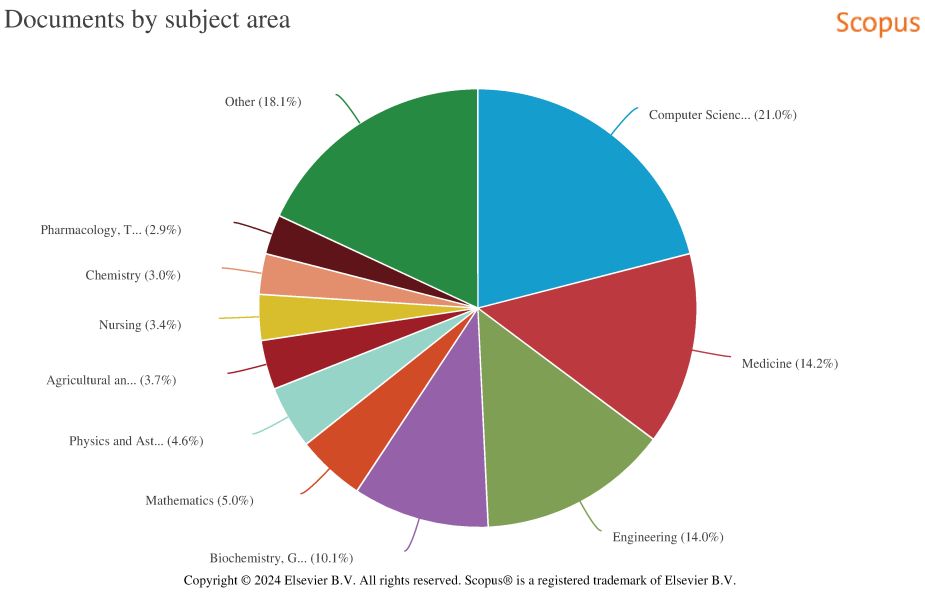


FIGURE 2.1 – Distribution of HAR research documents by subject area between 2010-2023. The pie chart shows the percentage contribution of various sectors to the HAR research field, with notable contributions from computer science, engineering, and health-related fields.

Scopus generated a pie chart illustrating the distribution of documents by subject area, showing the contribution percentage of each sector. Health fields notably span a remarkable percentage. FIGURE 2.2 depicts the evolution of the HAR research field between 2010-2023.

Scopus generated this curve showing the number of documents published each year, highlighting the increasing volume of research contributions in this field over the years.

To refine the initial broad search, a comprehensive query was structured to cover different aspects of the Wi-Fi CSI-based HAR domain, particularly focusing on spontaneous activities like falling. The query components (keywords) were divided into five categories:

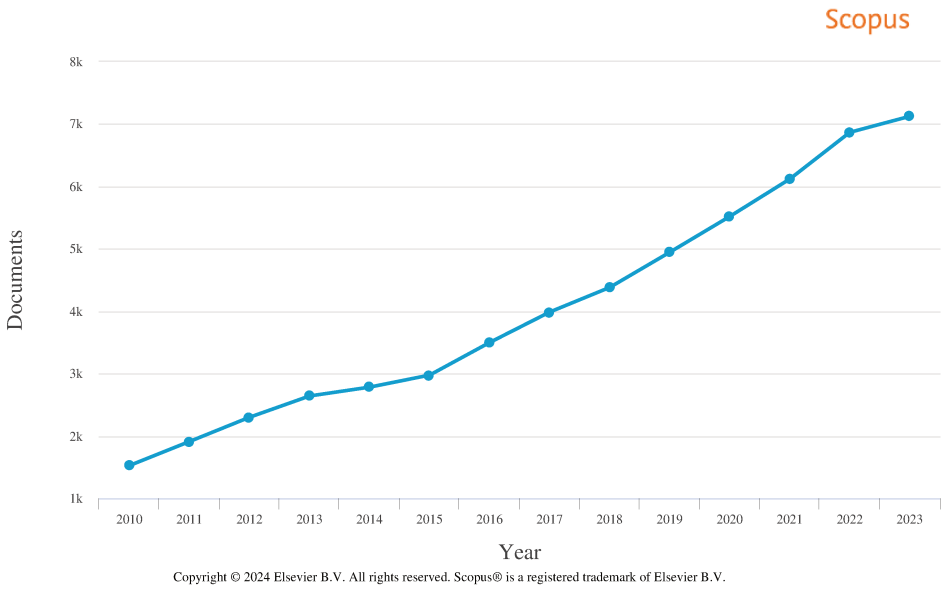


FIGURE 2.2 – Evolution of the HAR research field between 2010-2023. The line chart depicts the increasing number of documents published each year, highlighting the growing interest and research contributions in the domain of HAR.

— **Wi-Fi / Wireless Terms:** "Wi-Fi" OR "Wireless" OR "IEEE 802.11" OR "WiFi" OR "Wireless LAN" OR "WLAN" OR "Wi-Fi Networks" OR "WiFi Networks" OR "Wi-Fi Signals" OR "WiFi Signals" OR "Wi-Fi Devices" OR "WiFi Devices" OR "Device-Free Wi-Fi Sensing" OR "Device-Free WiFi Sensing" OR "Wi-Fi Sensing" OR "WiFi Sensing"

— **Channel State Information Terms:** "Channel State Information" OR "Channel State Information-Based" OR "Channel-State-Information" OR "CSI" OR "CSI-based" OR "Wi-Fi Channel State Information" OR "WiFi Channel State Information" OR "Wi-Fi CSI" OR "WiFi CSI" OR "Wireless Channel State Information" OR "Wireless CSI" OR "CSI Information" OR "CSI Correlation" OR "CSI Enhancement"

— **Human Activity Recognition Terms:** "Human Activity Recognition" OR "HAR" OR "Activity Recognition" OR "Activity Inference" OR "Human Motion Detection" OR "Motion Detection" OR "Human Activity Detection" OR "Device-Free Human Behavior Recognition" OR "Human Behavior Recognition" OR "Behavior Recognition" OR "Motion Recognition" OR "Human Fall Detection" OR "HFD" OR "Fall Detection" OR "Fall Prevention" OR "Fall Monitoring"

— **Falling Activity Terms:** "Falling Activity" OR "Fall Activity" OR "Fall Event" OR "Falling Event" OR "Fall Incident" OR "Falling Incident" OR "Fall Detection"

— **Spontaneous Activities:** "Spontaneous Activities" OR "Daily Activities" OR "Residential Activities" OR "Indoor Activities" OR "Home Activities" OR "Common Activities" OR "Spontaneous Movements" OR "Daily Movements" OR "Residential Movements" OR "Indoor Movements" OR "Home Movements" OR "Common Movements" OR "Human Activities"

The search initially resulted in 1,628 documents. To narrow down the results, the following filters were applied:

— **Publication Year Filter:** AND PUBYEAR $>$ 2016 AND PUBYEAR $<$ 2024
Focus on articles published in the last five years to ensure relevance and currency of the methodologies reviewed, reducing the number to 1,055 documents.

— **Language Filter:** AND (LIMIT-TO (LANGUAGE, "English"))
Include only articles written in English, further narrowing the results to 1,041 documents.

— **Document Type Filter:** AND (LIMIT-TO (DOCTYPE, "ar"))
Restrict results to peer-reviewed journal articles, ensuring high relevance and quality, resulting in 388 documents.

A total of 388 articles remained for screening and will undergo a detailed screening process. By employing this search strategy, we ensure a structured and thorough review of the literature, aligning with the objectives of the SMS methodology.

3. **Screening and Selection:**

The identified studies were screened based on the detailed inclusion/exclusion criteria, represented below. These criteria include all relevant CSI-based HAR advancements, with notes on potential healthcare applications. This multi-step process included title, abstract, and full-text screening to ensure the relevance and quality of the selected studies. The detailed criteria were applied to assess the studies' relevance to the specific research questions and their methodological rigor.

Inclusion Criteria:

— **Accessibility:** Full-text of the articles must be accessible.

— **Focus on HAR using Wi-Fi CSI:** The study must primarily focus on HAR using Wi-Fi CSI.

— **Relevant to healthcare:** The methodology applied or discussed must have potential applications or adaptations for enhancing healthcare services, such as patient monitoring and emergency response, etc.

— **Recognizable components are spontaneous activities:** The study must focus on spontaneous activities, including common daily movements such as walking, sitting, or falling, particularly in residential or indoor settings.

— **Measurement technology is Wi-Fi CSI:** The study must employ Wi-Fi CSI to capture data on human activities.

— **Document type is original research:** Only original research articles are included.

— **English language:** Articles must be written in English.

Exclusion Criteria:

— **Duplicate articles:** Any duplicate articles within the collection should be excluded.

— **Inaccessible full-text:** Articles for which the full text is not accessible should be excluded.

— **Non-relevant recognizable components:** Studies focusing on activities other than spontaneous activities (e.g., predefined or non-daily activities) should be excluded.

— **HAR not a key focus:** Studies where HAR is not the primary focus should be excluded.

— **Measurement technology is not Wi-Fi CSI:** Studies capturing data on human activities using technologies other than Wi-Fi CSI (e.g., camera or wearable sensors) should be excluded.

— **Document type other than original research:** Review articles or any other document types that are not original research articles should be excluded.

— **Out of Scope:** Articles that do not fall within the scope of the research objectives or do not align with the defined research questions should be excluded.

For screening, we used Rayyan [1], a free, web-based tool designed specifically for systematic reviews, which offers features for tagging, notes, and conflict resolution to help ensure a transparent and reproducible screening process. Before proceeding to screening, detecting and handling duplicates is

1. Rayyan, a tool for systematic reviews, available at: https://rayyan.qcri.org

an essential step. Rayyan detected and deleted 2 duplicate articles from the list of identified articles from our previous search strategy. The list of articles, represented by their metadata (structured information describing the characteristics of these articles and exported from a database like Scopus), forms our dataset. Rayyan thus ensures our dataset is clean before we begin the screening process. After removing the duplicates, a total of 386 articles remained for screening. Following this, our next steps involve title screening, followed by abstract and full-text screening. This clean dataset will ensure an efficient and effective screening process.

(a) Title Screening:
   — **Objective:** Scan the titles of all identified studies and exclude those that are obviously irrelevant to the research question.
   — **Outcome:** Articles Screened (N= 190) and Excluded Titles (N= 196).

(b) Abstract Screening:
   — **Objective:** Read the abstracts of the studies that passed the title screening stage. Studies that do not meet the inclusion criteria are excluded.
   — **Outcome:** Articles Screened (N=160) and Excluded Abstracts (N=30).

(c) Full-Text Screening:
   — **Objective:** Full-text articles should be assessed for detailed discussions on the methodological soundness of the CSI-based HAR solutions, their focus on spontaneous activities, and their ability to be adapted for enhancing healthcare services.
   — **Outcome:** Articles Screened (N=19) and Excluded Full-Texts (N=141)

To comprehensively assess the relevance, quality, and applicability of studies based on the full text, we followed a systematic approach. The primary objective of full-text screening is to ensure the methodological soundness, relevance, and quality of studies in detail, while also verifying that the studies aligned with our predefined inclusion criteria and addressed our research questions.

During the full-text screening process, we continued to use the predefined inclusion and exclusion criteria to systematically assess each study. The steps included:

(a) Accessibility Check: Ensuring the full text of each article was accessible, excluding any articles for which we could not obtain the full text.

(b) Deciding on the Number of Articles: While there is no fixed number of articles required for a systematic review, it is essential to maintain a manageable and comprehensive number. Ideally, having 15-30 high-quality, relevant articles is sufficient for a comprehensive review. In our case, we aimed to select 19 articles, which fell within this range.

(c) Selection Between Articles:
   — Methodological Rigor: Preference was given to studies with robust methodological designs and sound research practices.
   — Relevance to Healthcare Applications: Studies providing direct insights or applications to healthcare settings were favored.
   — Recency and Innovation: More recent studies reflecting the latest advancements and innovations were prioritized.
   — Citation and Impact: We considered the citation count and impact factor of the journal, indicating the influence and recognition of the study in the field.

Both objectives of full-text screening were crucial and applied concurrently:

— **Objective 1: Methodological Soundness -** Ensuring each study was methodologically sound by focusing on the robustness of the CSI-based HAR solutions. This included checking the study's design, data collection, analysis methods, and validation procedures, and ensuring that the study's conclusions were supported by the data and analysis presented.

— **Objective 2: Adaptation to Healthcare Services -** Assessing the potential of each study to be adapted for healthcare services, such as patient monitoring and emergency response. This involved evaluating whether the study's findings could be adapted or had potential applications in healthcare and looking for discussions or implications related to healthcare services.

Following the rigorous screening process, a total of 19 articles remained for critical analysis. By employing this process, we ensure that only the most relevant and high-quality studies are selected for our systematic literature review, aligning with the objectives of our research. FIGURE 2.3 presents the PRISMA Flow Diagram, documenting and illustrating the screening and selection process.

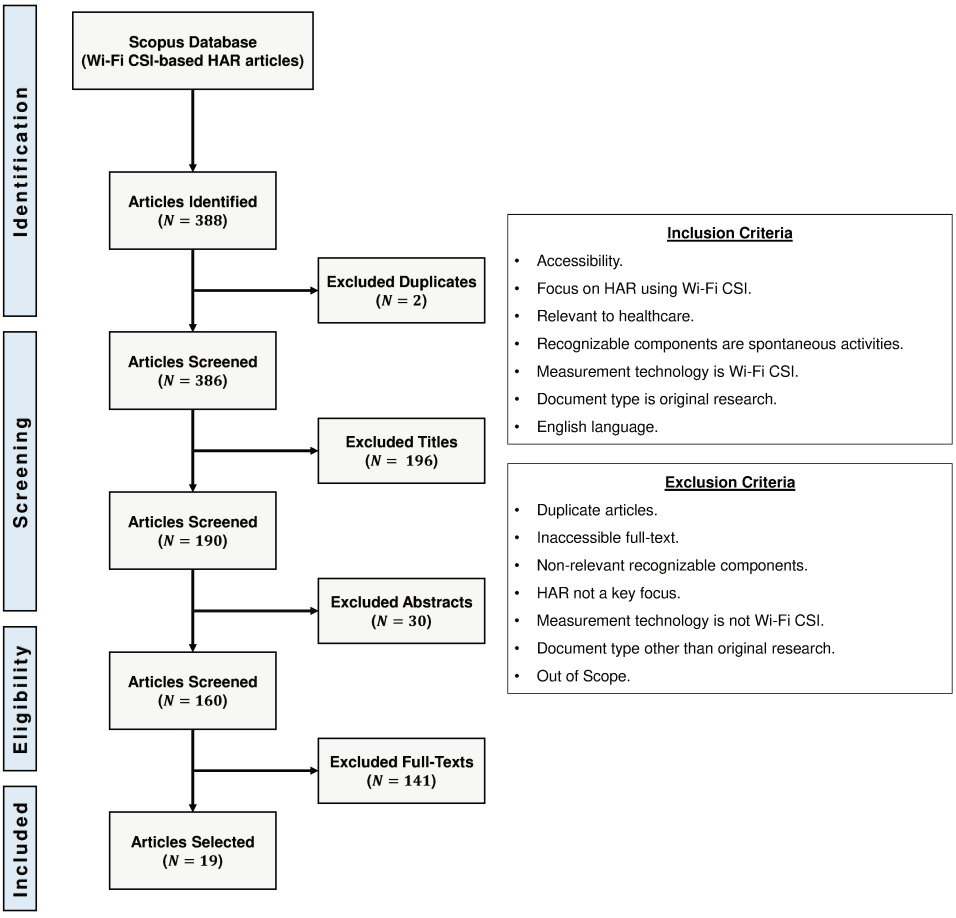


FIGURE 2.3 – PRISMA Flow Diagram illustrating the screening and selection process for identifying relevant Wi-Fi CSI-based HAR articles. The diagram shows the steps of identification, screening, eligibility, and inclusion, along with the number of articles at each stage.

4. **Classification Scheme:**

   We developed a detailed classification scheme to organize the 19 selected articles. Key columns in TABLE 2.1 include:

   — **Reference:** Identifies each study by combining the year and the name of the first author (e.g., [Wang, 2017b]). This column traces the evolution of recognition methodologies over time and highlights the contributors to the domain of study.

   — **Limitations:** Explores each study's constraints in data, techniques, and evaluation processes.

   — **Data:** Provides essential context about the scope and purpose of data utilized in each study. It breaks down into Healthcare Service(s) (potential applications of the CSI-based HAR solution), Dataset (details of the CSI dataset, including activities and environment), and binary indicators FD (Fall Detection) and AR (Activity Recognition) for the recognizable component of the HAR application.

   — **Recognition Methodology:** Covers the phases of CSI-based HAR solution development, from preprocessing of CSI data, through processing techniques for classification, to post-processing for outcome refinement.

   — **Methodology Performance:** Evaluates the recognition methodologies' effectiveness (activity recognition ability), efficiency (resource utilization proficiency), and versatility (reliability and adaptability across settings and users).

5. **Data Extraction and Mapping:** Data were extracted from the selected studies and categorized according to the classification scheme. This mapping provided a visual representation of the distribution of research across different categories.

6. **Analysis and Interpretation:** The mapped data were analyzed to identify trends, gaps, and common limitations. The detailed analysis is given in Section 7.

7. **Reporting:** The results of our analysis were documented, providing insights into the state of Wi-Fi CSI-based HAR in healthcare. Our findings identify key trends, common limitations, and areas for future research, offering a comprehensive understanding of the domain.

By employing the SMS approach, we ensured a structured and comprehensive review of the literature, enabling us to identify key trends, common limitations, and areas for future research in the domain of **Wi-Fi CSI-based HAR for healthcare**.

3 Narrative Review of the Selected Articles

In this section, we adopt a narrative-style review to summarize the final 19 selected articles within the domain of Wi-Fi CSI-based HAR. This complements the systematic mapping in Section 2 by shifting from structured classification to a more qualitative discussion of each study's dataset, objectives (i.e., multi-activity recognition and/or fall detection), and recognition methodology (including preprocessing, processing, and post-processing steps).

Beyond summarizing individual studies, this section lays the groundwork for subsequent classification and analysis. The collected insights establish a foundation for understanding how existing work addresses key issues in CSI-based HAR and, in some cases, falls short of providing adequate solutions. This sets the stage for the upcoming discussion on challenges (Section 4), performance aspects (Section 5), potential healthcare services (Section 6), and a more detailed analysis of trends and gaps (Section 7).

To enhance clarity and facilitate later discussions, we group the studies based on their primary focus: **Multi-activity Recognition–Focused Approaches** and **Fall Detection–Focused Approaches**. Note that evaluation metrics (e.g., accuracy or Latency) are not explicitly cited in this section to avoid

repetition; they will be discussed more thoroughly in the classification scheme and final analysis (Section 7).

3.1 Multi-activity Recognition–Focused Approaches

2019: Xiao, C. et al. [Xiao, 2019]: The authors introduced *CsiGAN*, an activity recognition approach leveraging a modified semi-supervised *Generative Adversarial Network* (GAN) model. CsiGAN addresses limited unlabeled activity CSI data from left-out (new) users. A novel complement generator produces diverse synthetic samples from scarce unlabeled data, refining a modified discriminator for multi-activity classification. The system was tested on raw CSI data from two datasets: one with several predefined actions and another containing six spontaneous activities. By designating one user as the left-out user, CsiGAN aims to enhance model performance for unseen participants.

2019: Wang, F. et al. [Wang, 2019]: This study proposed a joint activity recognition and indoor localization approach, collecting CSI data at 16 indoor experimental locations. The dataset captured variations in Wi-Fi channels induced by six predefined single-user gestures, gathered via one Etuss N210 transmitter and one Etuss N210 receiver. A one-dimensional Residual net (ResNet1D) model was trained exclusively on the raw CSI amplitude data.

2019: Chen, Z. et al. [Chen, 2019]: Chen and colleagues proposed an activity recognition approach using CSI data capturing seven different spontaneous activities in indoor settings (one transmitter, one receiver). Their methodology involved an *Attention-based Bi-directional Long Short-Term Memory* (ABLSTM) model trained on raw CSI data collected from both a public dataset [Fard Moshiri, 2021] and their own environment.

2021: Fard Moshiri, P. et al. [Fard Moshiri, 2021]: The authors proposed an activity recognition approach centered on seven single-user spontaneous activities. Data were collected via a Nexmon CSI Extractor setup (indoor environment, one transmitter, one receiver). Two distinct models were explored: a 2D Convolutional Neural Network (2D-CNN) that converts CSI amplitude data into RGB images (pseudocolor plots), and ABLSTM trained directly on raw CSI amplitude signals.

2022: Shi, Z. et al. [Shi, 2022]: Shi et al. introduced a recognition scheme named *MatNet-eCSI*, which spans three stages: CSI Collection, CSI Preprocessing, and Activity Recognition. First, using the Linux 802.11n CSI Tool, CSI data were collected from one transmitter and three receivers. After removing activity-irrelevant signals and compressing the high-dimensional CSI matrix, the authors employed MatNet, a hybrid CNN-BLSTM, to perform multi-activity classification. Training combined data from a previously seen environment with at least one sample from a new testing environment, enabling the model to learn robust hidden features.

3.2 Fall Detection–Focused Approaches

2017: Wang, Yuxi et al. [Wang, 2017b] (*WiFall*): *WiFall* was designed to detect falls and recognize other activities (e.g., walking, sitting down). After preprocessing with a *Weighted Moving Average* (WMA) filter and *Singular Value Decomposition* (SVD) for dimension reduction, the authors used a one-class *Support Vector Machine* (SVM) to distinguish falls from other movements and *Random Forest* (RF) for enhanced accuracy. A post-processing stage updated the profile database with false alarms, incrementally improving the system. The study also included a comparison with the previous version of WiFall [Han, 2014].

2017: Wang, H. et al. [Wang, 2017a] (*RT-Fall*): *RT-Fall* is a real-time fall detection system relying on an indoor fall CSI dataset. Preprocessing included 1-D linear interpolation, *Band-Pass* (BP) filtering (5–10 Hz), and CSI phase difference variance for activity segmentation. The authors extracted eight features (time- and frequency-domain), then employed an SVM with a Gaussian *Radial Basis Function* (RBF) kernel. Post-processing used real-time feedback to relabel misclassifications and update the model. RT-Fall was benchmarked against WiFall [Wang, 2017b].

2018: Dong, Z. et al. [Dong, 2018]: Dong and colleagues gathered an indoor motion CSI dataset encompassing fall detection, intruder detection, and activity recognition. They performed various preprocessing steps (Hampel and *Low-Pass* (LPF) filtering, subcarrier fusion, PCA, *Moving Variance Segmentation* (MVS) segmentation, Doppler Spread Spectrum analysis) to handle noise and extract relevant features. A RF classifier was then used for motion detection and activity recognition without any post-processing. The authors compared their approach against SVM and *k-Nearest Neighbor* (kNN).

2021: Wang, Yongchuan et al. [Wang, 2021a] (*FallViewer*): *FallViewer* targets elderly individuals at risk of falling. After assembling an indoor fall CSI dataset, the authors performed a rich suite of preprocessing (Wavelet Thresholding Denoising, Phase *Sampling Frequency Offset* (SFO) calibration, *Least Squares Regression* (LSR) for time lag correction, PCA, *Empirical Mode Decomposition* (EMD), etc.) to extract conventional and novel features. An SVM classifier followed, with no post-processing. FallViewer was benchmarked against *Naive Bayes* (NB), RF, KNN, RT-Fall [Wang, 2017a], and FallDeFi [Palipana, 2018].

2021: Schäfer, J. et al. [Schäfer, 2021b]: Schäfer and colleagues addressed fall detection and activity recognition. They used several preprocessing techniques (Hampel filter, *Discrete Wavelet Transform* (DWT), PCA, normalization). For processing, the authors compared an SVM to an LSTM network; no post-processing was involved.

2022: Hu, Y. et al. [Hu, 2022] (*DeFall*): *DeFall* focused on detecting falls among the elderly, working with both dummy-based and real human fall CSI data. The methodology utilized *Dynamic Time Warping* (DTW) (and variants like *Band-relaxed Segmental Locally Normalized DTW* (SLN-DTW)) for preprocessing and alignment, plus *DTW Barycenter Averaging* (DBA) for constructing a 2-D template (speed and acceleration). A DTW-driven pattern matching system detected falls by comparing real-time CSI streams to this 2-D template. No post-processing was employed. DeFall was benchmarked against WiFall [Wang, 2017b], FallDeFi [Palipana, 2018], and WiSpeed [Zhang, 2018b].

2022: Nakamura, T. et al. [Nakamura, 2022]: Nakamura et al. tackled fall detection for elderly populations. After preprocessing via 1-D linear interpolation and PCA (to highlight motion-relevant signals), the authors used *Short-Time Fourier Transform* (STFT) to convert CSI streams into spectrogram images. For classification, a ResNet34 model performed binary detection of falls. Post-processing was not used. Comparisons included lightweight CNNs (e.g., ResNet18) and FallDeFi [Palipana, 2018] to illustrate effectiveness.

2022: Wang, C. et al. [Wang, 2022]: Focusing on fall detection using an indoor fall CSI dataset, Wang, C. and colleagues preprocessed CSI data by extracting amplitude in a linear voltage space. Then they tested SVM and *eXtreme Gradient Boosting* (XGBoost) for binary classification of falls. No post-processing was applied. The authors compared both classifiers to gauge their suitability for real-world fall detection.

2023: Chen, S. et al. [Chen, 2023c] (*AFall*): *AFall* employed CSI Phase Unwrapping, Signal Sanitation (to mitigate *Sampling Time Offset* (STO), *Carrier Frequency Offset* (CFO), *Packet Detection Delay* (PDD)), and a Hampel filter for outlier removal. A *Moving Average* (MA) filter smoothed the AoA series further. The *Multiple Signal Classification* (MUSIC) algorithm, combined with statistical threshold methods, identified falls by changes in AoA rates. The authors discussed WiFall [Wang, 2017b], RT-Fall [Wang, 2017a], Anti-Fall [Zhang, 2015], and FallDeFi [Palipana, 2018] to contextualize AFall's performance.

2023: Cardenas, J.D. et al. [Cardenas, 2023]: Here, the authors utilized a large indoor activity dataset, emphasizing fall detection while also examining activity recognition. After generating spectrograms (via Gaussian windows) and filtering out noise, they relied on an LSTM alongside a CNN. No post-processing was employed. Comparative analysis of LSTM and CNN, along with referencing prior works offered additional perspective on evaluation metrics.

2023: Xia, Z. et al. [Xia, 2023]: This study focused on fall detection for older adults, using indoor CSI data. Preprocessing removed irrelevant subcarriers, employed Cubic Spline Interpolation to align timestamps, and performed *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) based clustering for noise removal. The core classifier was DeepConvLSTM, merging convolutional and LSTM layers. No post-processing was included. Performance references included RT-Fall [Wang, 2017a], EldTec [Limpanadusadee, 2018], DeFall [Hu, 2022], Sensing-Fi [Ramezani, 2018], and WiFall [Wang, 2017b].

2023: Chu, Y. et al. [Chu, 2023]: Chu and colleagues addressed fall detection by applying sliding windows to CSI data, optionally denoising via wavelet filtering, and downsampling. They then reshaped and normalized the data before feeding it into a modified EfficientNet for binary classification, without any post-processing. Benchmarking used FallDeFi [Palipana, 2018] and another Nakamura system [Nakamura, 2020] for comparison.

2023: Guo, R. et al. [Guo, 2023]: Guo et al. examined fall detection on a construction site dataset, covering activities such as lying down, jumping, or sitting. After noise reduction using a Median filter, a BLSTM model classified activities. Like many other approaches in the literature, no post-processing was introduced. Performance comparisons invoked references to computer vision [Mobasheri, 2022 ; Sultana, 2021] and wearable sensor-based systems [Chai, 2021 ; Lee, 2022].

2023: Yang, Z. et al. [Yang, 2023b] (*FallDar*): *FallDar* focuses exclusively on fall detection for independent elderly living at home. Preprocessing includes speed estimation (*Doppler Frequency Shift* (DFS)), *Autocorrelation Function* (ACF) Analysis, and a *Variational Autoencoder* (VAE) to generate synthetic data for unseen fall types. An adversarial learning-based DNN (feature extractor, fall detector, user discriminator) processes speed sequences for classification. The authors compared *FallDar* with NB, SVM, *FallDeFi* [Palipana, 2018], and *DeFall* [Hu, 2022], highlighting strong adaptability to user and motion diversity.

3.3 Summary and Next Steps

Across both focus categories (multi-activity classification and fall detection), these studies showcase a variety of signal processing techniques, ML architectures, and attempts to mitigate noise or sparse data challenges. Many approaches demonstrate promising results for recognizing spontaneous activities

or detecting falls, however, they differ in how they handle data quality, model complexity, and system adaptability.

By reviewing each paper's dataset, preprocessing pipeline, and classification methods, we gain a foundational understanding of how Wi-Fi CSI-based HAR methodologies have evolved. This overview also underlines the recurrent challenges surrounding data collection, noise reduction, hyper-parameter sensitivity, and the presence (or absence) of post-processing mechanisms. In the next subsections, we further identify and categorize these challenges (Section 4) and detail how performance is typically assessed (Section 5). Finally, we consolidate these findings into a classification scheme and interpret trends and gaps in Section 7, paving the way for our subsequent contributions.

4 Key Challenges and Limitations in CSI-Based HAR

Based on the systematic review of the literature in the domain of "CSI-based HAR," we identified potential key challenges and limitations. These can be grouped into two main categories: **Data-Related Concerns** and **Model-Related Concerns**. These categories provide a structured overview of the constraints within the domain of CSI-based HAR for healthcare using ML, helping to understand where studies might have gaps or areas for improvement. Acknowledging these limitations is crucial, as it guides future research in addressing these gaps and enhances the outcomes, validity, and generalizability of methodologies in this field.

4.1 Data-Related Concerns

The "Data-Related Concerns" category encompasses potential limitations with the data used for model training and evaluation, which are crucial for the performance and reliability of CSI-based HAR methodologies [Chen, 2023a].

- **Data Unavailability:** Limited access to or sharing of relevant datasets impacts the reproducibility and comparability of results.
- **Suboptimal Data Quality:**
  - *Noisy Training Data:* The presence of unwanted signals or distortions reduces model accuracy and reliability.
  - *Missing Training Data:* Incomplete datasets or training frames lead to biased models and affect performance.
  - *Absence of Empty Data:* Lack of empty data signals required for environment-specific calibration.
  - *Inadequate Data Filtering:* Insufficient filtering may retain noise, while excessive filtering might remove critical information.
  - *Synthetic Data Usage:* Reliance on synthetic data instead of real-world data, which may not accurately reflect actual scenarios.
- **Limited Data Diversity:** A lack of representativeness across various demographics, environments, scenarios, and hardware configurations affects the model's generalizability in diverse real-world settings.
- **Small-sized Data:** Insufficient data quantities hinder robust model training and validation [Alzubaidi, 2021].
- **Imbalanced Data:** Overrepresentation of certain classes leads to models biased towards majority classes [Fernández, 2018].

4.2 Model-Related Concerns

The "Model-Related Concerns" category pertains to the design, development, and evaluation of models and algorithms capable of effectively processing CSI data for HAR tasks. These concerns affect the effectiveness, efficiency, and versatility of CSI-based HAR methodologies in real-world applications. Due to their specificities and the distinct challenges they present, the model-related concerns are treated separately under the following subcategories: Processing Technique Concerns, Post-processing Mechanism Concerns, and Evaluation Process Concerns.

4.2.1 Processing Technique Concerns

The "Processing Technique Concerns" category includes potential limitations related to the models or algorithms employed for processing the CSI data.

— **Hyper-parameter Sensitivity:** Heavy dependence on the choice of parameters can lead to suboptimal performance and reliability issues (relevant for SVM, DL models) [Goodfellow, 2016].

— **Poor Scalability with Large Datasets:** Difficulties in handling large datasets reduce efficiency and increase computational costs (relevant for SVM, KNN) [Thanh Noi, 2018].

— **Overfitting Training Data:** Tendency to fit noise in the training data leads to poor generalization on new data (relevant for Random Forest, BLSTM, DL models) [Hastie, 2009].

— **Long Training Periods:** Significant time and resources required for model training limit rapid prototyping and experimentation (relevant for LSTM, BLSTM, DL models) [Brownlee, 2018].

— **Challenges with Noisy Data:** Difficulty managing datasets with high noise levels or irrelevant features affects accuracy and reliability in real-world applications (relevant for Random Forest, Naive Bayes) [Zhang, 2004].

— **High Model Complexity:** Requirement for considerable computational time or resources affects practical deployment, particularly in real-time applications and resource-constrained environments (relevant for Random Forest, BLSTM, DL models) [Goodfellow, 2016].

— **Difficulty Handling Long-term Dependencies:** Struggles with capturing long-range dependencies in data limit effectiveness in capturing temporal patterns (relevant for LSTM, 1D-CNN) [Bai, 2018 ; Goodfellow, 2016].

— **Data Hungry Nature:** Requirement for vast amounts of labeled data poses challenges in evolving fields with limited data (relevant for DL models) [LeCun, 2015].

— **Vanishing/Exploding Gradients:** Challenges in deep network training hinder model training and convergence (relevant for DL models, especially those with many layers) [Goodfellow, 2016].

— **Feature Engineering Dependence:** Reliance on manually extracted features potentially overlooks complex data patterns (relevant for traditional ML models like SVM, Naive Bayes) [LeCun, 2015].

— **Sub-optimal Processing Algorithm:** Inherent constraints of the technique impact its performance or suitability for the intended application.

4.2.2 Post-processing Mechanism Concerns

The "Post-processing Mechanism Concerns" category encompasses potential limitations related to the mechanisms developed for post-processing the outputs from the processing techniques [Ganaie, 2022 ; Asghari Oskoei, 2007 ; Wahid, 2020].

— **Absence of Post-processing Mechanism:** Lack of effective advanced decision mechanisms to enhance primary processing results.

— **Static Post-processing Mechanism:** Limited sensitivity of the mechanism to various real-world settings.

— **Sub-optimal Post-processing Algorithm:** Existing constraints within the mechanism affect its performance or suitability for the intended application.

4.2.3 Evaluation Process Concerns

The "Evaluation Process Concerns" category includes potential limitations with how the methodology's performance is validated and measured.

— **Absence of Comparison Against Others:** Lack of methodology comparison with established work from the research domain.

— **Unfair Benchmarking Against Others:** Failure to use standardized datasets and evaluation metrics compromises the validity of the comparison [Yang, 2023a ; Varga, 2024 ; Weber, 2019].

— **Inadequate Testing Procedures:** Use of inadequate procedures for testing the methodology impacts the reliability of the evaluation metrics [Varga, 2024].

By addressing these limitations individually, we can better tailor solutions to mitigate each concern. Collectively, these concerns impact the outcomes, validity, and generalizability of CSI-based HAR methodologies, affecting their effectiveness, efficiency, and versatility in real-world applications. Acknowledging these limitations is crucial for the development of the research, as it guides future studies in addressing these gaps and improving the overall performance of CSI-based HAR methodologies.

5 Performance Aspects for CSI-Based HAR Methodologies

After selecting and reviewing the relevant literature in the domain of CSI-based HAR, we identified key aspects of methodology performance crucial for a comprehensive evaluation. Given the unique challenges and objectives inherent to this domain, our evaluation focuses on three key aspects: Effectiveness, Efficiency, and Versatility. These aspects provide a comprehensive assessment of a methodology's capabilities and applicability. This structured approach enables us to determine the most promising methodologies for real-world applications, particularly in healthcare settings.

To evaluate these three aspects, we employ qualitative assessment values, ensuring a clear and understandable categorization of a methodology's capabilities. These assessment values provide categorical evaluations rather than numerical scores, helping to describe the performance or capability of the methodology in specific aspects. Using qualitative assessment values helps to simplify the evaluation process, making it more accessible and easier to understand, especially when communicating results to a broader audience. However, these qualitative assessments are grounded in quantitative data and analyses derived from the selected articles reviewed in Section 3.

Specifically, we integrate evaluation metrics that capture both computational performance (e.g., complexity, memory, latency) and predictive performance (e.g., accuracy, precision, recall), ensuring each aspect (Effectiveness, Efficiency, and Versatility) is evaluated on a quantitative foundation. To ensure transparency and consistency, we developed scales for each of these metrics based on the quantitative values extracted or estimated from the reviewed methodologies. For instance:

— Metrics such as **Accuracy**, **Precision**, **Recall**, and **F1-Score** were directly extracted from the studies, as they are commonly reported in the literature.

— For metrics like **Computational Complexity** and **Memory Requirements**, where values are not always explicitly provided by the authors, estimates were derived based on the descriptions of the methodologies (e.g., model architecture, data size).

— Similarly, **Latency** values were extracted when reported by the authors and estimated when sufficient methodological details were available to make reasonable approximations.

By analyzing the spread of the quantitative values for these metrics across the selected articles, we defined thresholds to categorize each metric into qualitative levels. This procedure ensures that the resulting scales appropriately reflect the range of performance reported in the literature, providing a meaningful basis for comparison.

By evaluating these aspects, we aim to ensure that the selected methodologies are not only effective and efficient but also versatile enough to be applied across various real-world scenarios, particularly in healthcare applications.

5.1 Effectiveness of the Methodology

In evaluating the effectiveness of CSI-based HAR methodologies, we focus on their ability to accurately recognize and classify human activities. Effectiveness is a critical aspect that directly impacts the reliability of HAR systems in real-world applications [Yang, 2023a ; Varga, 2024 ; Montavon, 2012]. To assess this, we use several key evaluation metrics:

— **Accuracy**: This metric measures the proportion of correctly identified activities out of the total activities observed. It provides a general sense of the methodology's performance but can be misleading if the data is imbalanced.

— **Precision**: Precision evaluates the number of true positive results divided by the number of all positive results, indicating how many of the identified activities are correctly classified.

— **Recall**: Also known as sensitivity, recall measures the number of true positive results divided by the number of all samples that should have been identified as positive, indicating how well the methodology detects actual activities.

— **F1-Score**: This is the harmonic mean of precision and recall, providing a single metric that balances the two. It is particularly useful when the data has an imbalanced class distribution.

Metrics Scales:

— **Accuracy, Precision, Recall, and F1-Score**:
  — High: $\geq 90\%$
  — Moderate: $\geq 80\%$ to $< 90\%$
  — Low: $< 80\%$

These metrics are qualitatively assessed as "Low", "Moderate", or "High", providing an overall view of the methodology's effectiveness in activity recognition.

5.2 Efficiency of the Methodology

Efficiency pertains to the computational resources required by the methodology, the speed at which it processes data, and its scalability. Efficient methodologies are crucial for real-time applications and large-scale deployments [Yang, 2023a ; Ng, 2017 ; Xu, 2020 ; Montavon, 2012]. We assess efficiency using the following metrics:

— **Computational Complexity**: This measures the amount of computational power needed to run the methodology, typically analyzed in terms of the number of Floating-point Operations (FLOPs) required to process a single input instance.

— **Memory Requirements**: This evaluates the amount of memory necessary for the methodology to function, which impacts the feasibility of deploying the system on devices with limited resources.

— **Response Time (Latency)**: This measures the time taken from data input to the output of results, which is crucial for applications requiring real-time analysis.

Metrics Scales:

— **Computational Complexity**:
  — Low (better): $< 17,835$ FLOPs
  — Moderate: $\geq 17,836$ to $< 11,781,401$ FLOPs
  — High: $\geq 11,781,401$ FLOPs

— **Memory Requirements**:
  — Low (better): < 1 MB
  — Moderate: ≥ 1 MB to < 10 MB
  — High: ≥ 10 MB

— **Latency**:
  — Low (better): < 0.1 seconds
  — Moderate: ≥ 0.1 seconds to < 1 second
  — High: ≥ 1 second

These efficiency metrics are evaluated qualitatively as "Low", "Moderate", or "High" to give a clear indication of the methodology's performance in practical scenarios.

5.3 Versatility of the Methodology

Versatility refers to the methodology's adaptability and reliability across diverse environments and user groups. It ensures that the methodology can maintain performance under various conditions, which is essential for the broad applicability of HAR systems [Yang, 2023a ; Chen, 2023a ; Montavon, 2012]. Versatility encompasses several key aspects:

— **Generalization**: This aspect assesses how well the methodology performs on unseen data, such as new activities, environments, or users. It can be measured using accuracy, precision, recall, f1-score, AUC-ROC, and leave-one-out cross-validation. The generalization score is qualitatively rated as "Low", "Medium", or "High".

— **Adaptability**: This measures the methodology's ability to remain effective as environments change and user behaviors evolve. It is assessed by observing changes in evaluation metrics under new conditions or after fine-tuning. The adaptability index is categorized as "Rigid", "Moderate", or "Flexible".

— **Robustness**: This evaluates the methodology's performance in the presence of noisy, incomplete, or corrupted data. Stability in evaluation metrics under adverse conditions indicates robustness. The robustness rating is described as "Fragile", "Resilient", or "Sturdy".

— **Transferability**: This aspect measures the methodology's ability to transfer knowledge from one domain to another, adapting effectively to new environments and user groups. It is assessed by evaluating evaluation metrics on the target domain after knowledge transfer. The transferability level is qualitatively rated as "Low", "Medium", or "High".

6 Potential Healthcare Services Benefiting from HAR

HAR solutions that exploit non-intrusive Wi-Fi CSI can enhance many aspects of digital healthcare. *Patient Monitoring* is the umbrella service they all enable, yet the specific *recognizable components* (the activities sensed) and their *recognition purposes—i.e. the concrete healthcare objective that each detected activity serves* dictate which downstream workflows—such as *Emergency Response*, *Telehealth*, or *Rehabilitative Care*—benefit most. For example, a HAR system devoted solely to the component "fall" naturally aligns with Emergency Response, whereas a system that captures a wider set of Activities of Daily Living (ADLs) supports broader monitoring and can, in turn, feed Diagnostic or Rehabilitative services. Hence the services listed below should be read as *sub-sets* or *extensions* of patient monitoring, assigned case-by-case according to each study's activity set and clinical intent.

— **Patient Monitoring**: HAR systems can continuously track spontaneous activities such as *walking*, *sitting*, and *lying down*, detecting deviations from an individual's normal routine or early signs of distress. For example, a progressive reduction in daily step count or prolonged bed-rest periods may herald post-operative complications or cognitive decline.

— **Emergency Response**: When a critical recognizable component such as a *fall* (or abrupt cessation of movement, which *may* indicate stroke or cardiac arrest) is detected, the system can trigger an immediate alert to caregivers or emergency teams, minimizing the time to intervention.

— **Diagnostic Support**: Fine-grained analysis of periodic or predefined movements (e.g. hand tremor amplitude, gait cadence) provides clinicians with quantitative biomarkers useful in diagnosing or staging diseases such as Parkinson's or multiple sclerosis [Chen, 2023b; Chen, 2022].

— **Rehabilitative Services**: During physiotherapy, recognizing range-of-motion exercises (e.g. knee bends, shoulder abduction) enables objective feedback on exercise quality, adherence, and recovery progress—without requiring wearable instrumentation [Touhiduzzaman, 2025].

— **Palliative Care**: For terminal or chronically ill patients, subtle restlessness, guarding, or rapid position changes can signal possible pain or discomfort. Continuous HAR therefore supports timely adjustments of analgesia and enhances patient comfort [Alzaabi, 2024].

— **Health Education**: By quantifying daily activity patterns, HAR can personalize lifestyle recommendations—e.g. encouraging additional light-intensity movement when sedentary time exceeds a threshold—and thus promote preventive health behaviors [Chung, 2024]

— **Mental Health Services**: Changes in recognizable components linked to circadian rhythm (e.g. fragmented sleep, nocturnal wandering) or social withdrawal (reduced locomotion within the home) may serve as early indicators of depression or anxiety, providing actionable data to mental-health professionals [Ware, 2018].

— **Nutritional Services**: Preliminary research on step-count and movement intensity points toward future CSI-based energy-expenditure estimation, which could one day help dietitians tailor caloric intake and weight-management plans [Guo, 2018; Liu, 2019].

— **Telehealth Services**: Gesture-based predefined actions (e.g. hand-wave to answer a video call, arm-raise to request assistance) facilitate remote interaction with clinicians, supplementing sensor readings with context-rich behavioral data for virtual consultations [Ahmed, 2020; Wang, 2024].

By integrating CSI data into these services, HAR systems can offer non-intrusive, continuous, and accurate monitoring, greatly enhancing patient care and healthcare service delivery. This alignment not only reflects current technological advancements but also meets the growing demand for sophisticated, non-intrusive healthcare monitoring systems, marking a significant step forward in the integration of technology into everyday health-management practices [Ge, 2023; Tan, 2018; Wang, 2018].

7 Synthesis and Interpretation of the Mapped Literature

This section consolidates all findings from the literature review (Section 3), the key challenges and limitations (Section 4), the performance evaluation aspects (Section 5), and the potential healthcare services (Section 6). We integrate these elements into a systematically selected classification scheme (TABLE 2.1) that compares 19 state-of-the-art "Wi-Fi CSI-based HAR" articles identified through a multi-stage screening process. Below, we first describe how each column in the table is defined and tied to Subsections 4, 5, and 6. We then illustrate how we filled the table via a concrete example, before analyzing overarching trends, limitations, and gaps. Finally, we highlight how these gaps motivate our three subsequent contributions (Chapters 3, 4, and 5).

7.1 Classification Scheme Description

As introduced in Section 2, our table dissects each article into the following columns:

1. **Reference:** This column lists each study's citation, illustrating the evolution of methodologies in CSI-based HAR over time.
2. **Limitations:** Drawing on the categories identified in Section 4, this column highlights which data- and model-related concerns each study exhibits (e.g., *Small-sized Data*, *Overfitting*, *Absence of Post-processing Mechanism*). Since not all limitations from Section 4 appear in a single study, this column indicates how frequently each limitation arises and whether a given study addresses (or overlooks) certain challenges.
3. **Data:**
   - *Healthcare Service(s)*: Refers to the potential application(s) in healthcare, referencing Section 6 (e.g., *Emergency Response*, *Patient Monitoring*).
   - *Dataset*: Summarizes the environment and type of activities (e.g., *indoor fall*, *lab gestures*).
   - *FD* and *AR* are binary indicators specifying whether the study focuses on *Fall Detection* (FD) or *general Activity Recognition* (AR). Some works address both.
4. **Recognition Methodology:** Organized in line with Section 3:
   - *Pre-processing Method(s)*: Includes denoising, segmentation, filtering, or feature extraction steps, as well as any signal processing approaches.
   - *Processing Method(s)*: Encompasses the main machine learning (e.g., CNN, SVM, BLSTM) or signal processing techniques (e.g., MUSIC algorithm, DTW-based pattern matching) used for classification or decision-making.
   - *Post-processing Method(s)*: Any advanced decision mechanism or output refinement stage (e.g., user-feedback relabeling).
5. **Methodology Performance:** This column assesses each approach's *Effectiveness*, *Efficiency*, and *Versatility*, as defined in Section 5.
   - *Effectiveness*: Although we provide a *qualitative* rating (Low, Moderate, High) in the table, we primarily rely on **recall** as a *quantitative* indicator of prediction performance. This choice is made because recall is consistently reported across many studies and is more indicative of detection capability in imbalanced scenarios (e.g., fall detection).
   - *Efficiency*: Typically derived from the *Computational Complexity* (FLOPs) or a coarse estimate if not explicitly reported. We use the qualitative scales (Low ($\downarrow$), Moderate ($\rightarrow$), or High ($\uparrow$)) from Subsection 5.2.

— *Versatility*: Rated using *Generalization Score*, *Adaptability Index*, *Robustness Rating*, and *Transferability Level*, as described in Subsection 5.3. To simplify notation in the table, we employ arrows (↓, →, ↑) for these sub-aspects whenever applicable. We mark *Not Provided (NP)* if the original study's reports do not allow any reasonable assignment.

Overall, TABLE 2.1 provides a single reference point for comparing how each study navigates the challenges of CSI-based HAR for healthcare applications. Note that our classification is not an exhaustive benchmarking of each approach against a standardized dataset and evaluation metrics; rather, we rely on the authors' own reported performances and experimental setups. Although a universal benchmark would facilitate direct comparisons, each paper's data and methods differ, limiting the feasibility of cross-publication comparisons.

7.2 Illustrative Example of Table Entry Filling

Before examining broader trends, we present an example from the study [Wang, 2019], one of our 19 reviewed articles, to show how each column is filled:

— **Limitations:** By cross-checking with Section 4, we identified issues such as *Absence of Empty Data* and *Small-Sized Data* (no baseline CSI frames, only 1,394 samples). We also noted *High Model Complexity* (a large ResNet1D architecture) and *Difficulty Handling Long-term Dependencies* (CNN-based methods excel at local patterns but may miss longer temporal correlations). Finally, no advanced post-processing was reported (*Absence of Post-processing*).

— **Data:**

  — *Healthcare Service(s):* Linking hand-gesture recognition to *Telehealth* (Section 6) for potential remote patient monitoring.

  — *Dataset*: Six gestures collected with Ettus N210 USRPs in a single indoor environment.

  — *FD/AR Indicators*: AR (✓), FD (×).

— **Recognition Methodology:**

  — *Pre-processing*: Manual annotation, temporal segmentation, linear interpolation.

  — *Processing*: ResNet1D for gesture classification.

  — *Post-processing*: Not employed.

— **Methodology Performance:**

  — *Effectiveness*: A macro-average recall of 88% computed from the reported per-class values falls in the *Moderate* range (80–90%) defined in Subsection 5.1.

  — *Efficiency*: An estimated 5.7×10^8 FLOPs (*High* complexity in Subsection 5.2).

  — *Versatility*: The authors did not test multiple environments or user groups, so *Generalization* and *Adaptability* are limited (Medium/Moderate). They showed resilience to minor variations but offered no noise stress tests (*Resilient*), and no cross-domain evaluation (*NP* for Transferability).

This walkthrough demonstrates how reported or inferred details map onto the classification scheme. Applying this procedure uniformly to all 19 articles enables fair comparisons and highlights gaps in the literature.

7.3 Observed Trends and Common Limitations

After completing the table entries for all 19 selected articles, several patterns emerge, both in **quantitative** and **qualitative** terms:

- **Quantitative Observations**:
  - *Post-processing* is rarely employed: Out of 19 works, only **3** studies introduce any advanced decision layer to refine outcomes. This aligns with the "*Absence of Post-processing Mechanism*" in Subsection 4.2.
  - *High Model Complexity* is widespread: More than half the articles adopt deep networks (CNN, BLSTM, ResNet variants), frequently incurring *High* computational overhead. Only a small fraction rely on simpler algorithms like SVM with modest input dimensions.
  - *Data Unavailability* or *Small-Sized Data* appears in over half of the studies, underscoring the challenge of collecting large, diverse CSI datasets.
  - *Additional Limitations:* Certain other concerns also stand out. For instance, *Absence of Empty Data* emerges in about 58% of the literature, while *Overfitting* and *Long Training Periods* are each identified by roughly 21%. Handling *Long-term Dependencies* (around 16%) and *Unfair Benchmarking* (about 16%) occur less frequently but remain critical for real-world deployment.
- **Qualitative Insights**:
  - **Preprocessing is considered essential**: Many authors invest in filtering (e.g., Hampel, wavelet, PCA) and advanced segmentation (e.g., Doppler spectrum), which correlate with higher recognition rates.
  - **Limited data diversity** inhibits generalization: Single-user or single-environment evaluations dominate, yielding moderate or unverified *Adaptability* and *Transferability* scores (Subsection 5.3).
  - **Lack of standardized benchmarking**: Although some articles compare to WiFall or FallDeFi, no universal dataset or agreed-upon evaluation protocol emerges, impeding fair cross-publication comparisons.
  - **Minimal focus on advanced post-processing**: Even studies with strong accuracy in controlled setups seldom introduce threshold tuning or user-feedback loops to manage real-world variability.

In sum, the largest gaps include *unfair or inconsistent benchmarking*, *limited use of post-processing*, and *insufficient data diversity*. Many studies also omit *baseline (empty) CSI frames* or advanced robustness tests, risking overfitting to a single environment or user group.

7.4 Bridging Identified Gaps to Our Contributions

While these limitations collectively define the problem space, most works address only a fraction of the recognized concerns, leaving ample room for further exploration. In this thesis, we propose three distinct contributions (Chapters 3, 4, and 5) that partially alleviate the **Data-Related** and **Model-Related** problems introduced in the General Introduction (Section 2) and detailed in Section 4:

- **Contribution 1 (Chapter 3) – Multi-Activity Recognition via a Lightweight CNN:** Addresses *High Model Complexity* by designing a streamlined CNN architecture (*HAR-LightCNN*), and tackles *Small-Sized* and *Imbalanced Data* via data augmentation. It also emphasizes proper testing splits to counteract *Inadequate Testing Procedures* and *Unfair Benchmarking Against Others*.
- **Contribution 2 (Chapter 4) – Comprehensive Fall Detection (HFD-Adv):** Expands on the *Data-Related Concerns* of *Missing Training Data*, *Absence of Empty Data*, and *Limited Data Diversity*, while introducing a new *Post-processing* stage (ADM) to handle *Absence of Post-processing Mechanisms*. Additionally, it addresses *Overfitting*, *Hyper-Parameter Sensitivity*, and *Long Training Periods* using genetic algorithms and an LRFinder approach.

— **Contribution 3 (Chapter 5) – CSI-based Human Identity Recognition (HIDR):** In parallel to broader HAR, user identity recognition faces *intra-class similarity* challenges, especially in fall scenarios. While HIDR shares *long-term dependency* issues with HAR, it also must distinguish different users exhibiting similar fall patterns. Our transformer-based model targets these subtle user-specific features.

  — *Note on HIDR Domain Scope:* Because CSI-based HIDR hinges on *behavioral* rather than static biometrics, it inherits many HAR-related constraints (e.g., high-dimensional data, overfitting risks). Yet it introduces unique complexities such as *intra-class similarity* between users. Hence, a dedicated methodology and literature review appear in Chapter 5, distinct from these 19 AR/FD-focused articles.

In conclusion, no existing single study resolves both *Data-Related* and *Model-Related* issues comprehensively. Our thesis thus concentrates on specific subsets of these widely recognized gaps to strengthen the applicability of CSI-based HAR in healthcare scenarios. The subsequent chapters demonstrate how each contribution systematically tackles selected limitations, ultimately furthering the field's progress.

8 Conclusion

In this chapter, we conducted a *systematic mapping* and *narrative-style review* of the **Wi-Fi CSI-based HAR literature**, isolating a final set of 19 articles that represent current state-of-the-art methodologies, particularly in healthcare-oriented scenarios. By applying structured search, screening, and classification (Section 2), we identified recurring technical innovations and gaps in the field (Sections 3–6), then synthesized our findings in a unified classification scheme (Section 7). This approach allowed us to pinpoint common **data-related concerns** (e.g., small-sized data, limited data diversity) and **model-related concerns** (e.g., high model complexity, absence of post-processing mechanisms, inadequate benchmarking) that frequently hinder the deployment of robust, real-world HAR solutions for healthcare scenarios.

The review revealed **three major areas** requiring more focused research:

— *Data Availability and Quality*: Most existing work struggles with insufficient or underrepresented datasets, making it difficult to train and validate models adequately for diverse healthcare applications.

— *Model Complexity vs. Practical Deployment*: Advanced deep learning architectures often offer strong accuracy but come at the cost of computational overhead, limiting their use in resource-constrained or real-time healthcare environments.

— *Post-processing and Evaluation Rigor*: Few studies implement advanced post-processing strategies or employ standardized benchmarking practices, leading to uncertainty in performance comparisons and limited adaptability to changing clinical needs.

Together, these limitations outline a significant research gap for CSI-based HAR in healthcare, underscoring the need for more efficient architectures, improved data collection practices, and robust evaluation frameworks. In light of these findings, the next chapters of this thesis (*Chapters 3, 4, and 5*) propose novel contributions aimed at selectively addressing these key concerns:

— Chapter 3 introduces a lightweight CNN solution (*HAR-LightCNN*) targeting small-sized and imbalanced data, while ensuring low computational complexity suitable for patient monitoring scenarios.

— Chapter 4 presents a comprehensive fall detection framework (*HFD-Adv*) that not only tackles data unavailability and diversity issues but also integrates a new post-processing mechanism to enhance real-world reliability.

— Chapter 5 expands the CSI-based HAR domain to human identity recognition (*HIDR*), addressing intra-class similarity and further pushing the boundaries of adaptability in healthcare environments.

By addressing these data- and model-related gaps, our aim is to advance CSI-based HAR towards more deployable, scalable, and clinically relevant methodologies. The forthcoming contributions highlight how strategic modifications in architecture design, data handling, and evaluation protocols can elevate HAR's utility for diverse healthcare services. In doing so, we hope to further reinforce Wi-Fi CSI's promise as a non-intrusive, privacy-preserving platform for activity recognition and patient care.

TABLE 2.1 – State-of-the-art of CSI-based HAR for healthcare: ML and signal processing methodologies

Reference	Limitations	Data				Recognition Methodology			Methodology Performance					
		Healthcare Service(s)	Dataset	FD	AR	Pre-proc. method(s)	Proc. method(s)	Post-proc. method(s)	Effectiveness	Efficiency	Versatility			
											Gen. Score	Adapt. Index	Robust. Rating	Transf. Level
[Wang, 2017b]	- Absence of Empty Data - Small-sized Data - Overfitting Training Data - Challenges with Noisy Data - Feature Engineering Dependence	Elderly Monitoring	Indoor Activities	✓	✓	WMA filter SVD LOF	SVM	False Alarms Feedback	94%	→	→	NP	→	↓
[Wang, 2017a]	- Data Unavailability - Absence of Empty Data - Overfitting Training Data - Feature Engineering Dependence	Emergency Response	Indoor Fall	✓	×	1-D Linear Interpolation BP Filter Features Extraction	SVM	User feedback relabeling	93%	NP	→	→	→	→
[Dong, 2018]	- Absence of Empty Data - Inadequate Data Filtering - Feature Engineering Dependence - Challenges with Noisy Data - Absence of Post-processing - Inadequate Testing Procedures	Patient Monitoring & Diagnostic Support	Indoor Motion	✓	✓	Hampel Filter LPF Filter PCA MVS Segmentation Doppler Spread Spectrum Features Extraction	RF	×	94%	↑	↑	NP	→	↓
[Xiao, 2019]	- Small-Sized Data - High Model Complexity - Long Training Periods	Emergency Response & Health Education	Lab Sign Language & Indoor Fall	✓	✓	Static Segment Removal Downsampling	GAN	Manifold Regularization	NP	↓	↑	↑	→	→
[Wang, 2019]	- Absence of Empty Data - Small-sized Data - High Model Complexity - Handling Long-term Dependencies - Absence of Post-processing	Telehealth	Indoor Hand Gestures	×	✓	Manual Annotation Temporal Segmentation Linear Interpolation	ResNet1D	×	88%	↓	→	→	→	NP
[Chen, 2019]	- Data Unavailability - Imbalanced Data - Long Training Periods - High Model Complexity - Absence of Post-processing - Inadequate Testing Procedures	Patient Monitoring & Emergency Response	Indoor Activities	✓	✓	Sliding Window Segmentation Amplitude Extraction	ABLSTM	×	96.3%	↓	→	→	→	↓
[Wang, 2021a]	- Data Unavailability - Absence of Empty Data - Limited Data Diversity - Overfitting Training Data - Challenges with Noisy Data - Feature Engineering Dependence - Absence of Post-processing	Emergency Response & Hospice Care	Indoor Fall	✓	×	Wavelet Thresholding Denoising SFO Calibration PCA Features Extraction EMD	SVM	×	97.6%	NP	→	→	→	→
[Schäfer, 2021b]	- Inadequate Data Filtering - High Model Complexity - Handling Long-term Dependencies - Vanishing/Exploding Gradients - Absence of Post-processing	Elderly Monitoring	Indoor Activities	✓	✓	Hampel Filter DWT PCA Normalization Scaling	SVM LSTM	×	NP	↓	→	→	NP	↓
[Fard Moshiri, 2021]	- Absence of Empty Data - Small-sized Data - High Model Complexity - Absence of Post-processing - Inadequate Testing Procedures	Patient Monitoring & Emergency Response	Indoor Activities	✓	✓	Pseudocolor Plotting Normalization	2D-CNN	×	95%	↓	↑	→	→	NP
[Hu, 2022]	- Data Unavailability - Limited Data Diversity - Feature Engineering Dependence - Absence of Post-processing	Emergency Response & Hospice Care	Dummies and Human Fall	✓	×	DTW SLN-DTW DBA Algorithm Fall Event Template	DTW-based pattern matching	×	95%	→	↑	↑	→	→
[Nakamura, 2022]	- Data Unavailability - Limited Data Diversity - High Model Complexity - Absence of Post-processing - Inadequate Testing Procedures	Emergency Response & Hospice Care	Room Fall	✓	×	1-D Linear Interpolation PCA STFT	ResNet34	×	95%	↓	→	↑	→	↑
[Shi, 2022]	- Data Unavailability - Small-sized Data - Long Training Periods - High Model Complexity - Absence of Post-processing	Patient Monitoring	Indoor Activities	×	✓	Linear Recursive Operation Correlation-Based Compression Timing Offset Compensation	CNN-ABLSTM	×	93.4%	↓	↑	↑	→	↑
[Wang, 2022]	- Data Unavailability - Absence of Empty Data - Limited Data Diversity - Feature Engineering Dependence - Absence of Post-processing	Emergency Response	Fall	✓	×	Normalization Amplitude Extraction	SVM XGBoost	×	NP	→	→	NP	→	↓
[Chen, 2023b]	- Limited Data Diversity - Overfitting Training Data - Absence of Empty Data - Absence of Post-processing - Unfair Benchmarking	Emergency Response & Hospice Care	Indoor Fall	✓	×	CSI Phase Unwrapping Signal Sanitation Hampel Filter MA Filter	MUSIC algorithm	×	84.9%	→	↑	→	↑	→
[Cardenas, 2023]	- Data Unavailability - Noisy Training Data - Limited Data Diversity - Handling Long-term Dependencies - Challenges with Noisy Data - Absence of Post-processing	Emergency Response	Large Indoor Activity	✓	✓	Gaussian Window Function Gaussian Filtering	LSTM CNN	×	82%	→	NP	→	NP	↓
[Xia, 2023]	- Data Unavailability - Absence of Empty Data - High Model Complexity - Absence of Post-processing - Unfair Benchmarking	Emergency Response & Hospice Care	Indoor Fall	✓	×	Subcarrier Pruning Cubic Spline Interpolation DBSCAN	DeepConvLSTM	×	91%	↓	→	NP	→	↓
[Chu, 2023]	- Absence of Empty Data - High Model Complexity - Absence of Post-processing	Emergency Response & Hospice Care	Fall	✓	×	Sliding Window Segmentation Wavelet Denoising Downsampling Reshaping Normalization	EfficientNet	×	100%	↓	↑	↑	↑	↓
[Guo, 2023]	- Data Unavailability - Absence of Empty Data - Long Training Periods - High Model Complexity - Absence of Post-processing - Unfair Benchmarking - Inadequate Testing Procedures	Construction-worker Monitoring	Construction Site Activities	✓	✓	Median Filter	BLSTM	×	95%	↓	NP	NP	NP	↓
[Yang, 2023b]	- Data Unavailability - Sub-optimal Data Quality - Absence of Post-processing	Emergency Response & Hospice Care	Household Fall	✓	×	Signal Segmentation Algorithms Speed Estimation Algorithm Kernel & Residual Speed Modeling VAE	Adversarial-based DNN	×	NP	→	↑	↑	↑	→

* NP = Not Provided.

Part II

CONTRIBUTIONS

Chapter 3

CSI-Based HAR Approach via a Lightweight CNN and Data Augmentation for Patient Monitoring Service

Objectives

This chapter presents *HAR-LightCNN*, a lightweight CNN-based solution designed for CSI-based multi-activity recognition in healthcare settings. By integrating data augmentation techniques, the methodology effectively addresses small-sized and imbalanced CSI datasets, ensuring efficient and accurate patient monitoring.

Chapter Outline

1 Introduction **65**
- 1.1 Background and Importance of HAR in Modern Healthcare 65
- 1.2 Defining the Problem Space and Its Challenges 65
- 1.3 Preliminaries 66
- 1.4 Contributions 70
- 1.5 Organization of the Chapter 71

2 HAR-LightCNN Methodology Overview **72**
- 2.1 Introduction to the Methodology 72
- 2.2 Objectives and Design Principles 73

3 Data Preparation and Preprocessing **74**
- 3.1 Datasets Overview and Initial Setup Considerations 74
- 3.2 Overview of the Preprocessing Steps 75

4 Lightweight CNN Activity Recognizer **83**
- 4.1 Selecting CNNs for Multidimensional Time-Series Analysis 83
- 4.2 Architecture of the Recognizer 84
- 4.3 Implementation Details 89

5 Results and Evaluation **89**
- 5.1 Benchmarking with Standardized Testing Subsets and Evaluation Metrics 89
- 5.2 Effectiveness and Efficiency of HAR-LightCNN 91

6 Conclusions **98**
- 6.1 Summary of Key Findings 98
- 6.2 Benefits for Healthcare Services 98
- 6.3 Future Directions 100

1 Introduction

1.1 Background and Importance of HAR in Modern Healthcare

In the context of healthcare, Human Activity Recognition (HAR) plays a pivotal role in monitoring and assessing patients' physical activities, which is critical for managing chronic diseases, supporting rehabilitation processes, and ensuring the well-being of elderly individuals.

HAR relies on data-driven techniques to automatically identify and classify human activities based on sensor data. Its integration in healthcare systems allows for continuous and unobtrusive patient monitoring. Traditional activity monitoring methods, based on wearable sensors and cameras, often face challenges like patient compliance and privacy concerns. Wi-Fi Channel State Information (CSI)-based HAR, on the other hand, leverages existing Wi-Fi infrastructure to monitor activities without the need for additional devices, offering a non-invasive and cost-effective solution. Implementing HAR in healthcare settings can significantly improve patient outcomes by enabling timely interventions and personalized care. For instance, real-time activity recognition can help in detecting falls, monitoring rehabilitation progress, and alerting caregivers to unusual behaviors that may indicate health issues. Consequently, HAR contributes to enhancing the quality of care and ensuring patient safety, making it an essential component of modern healthcare. Current research trends in HAR focus on improving accuracy and robustness, particularly through the use of advanced machine learning (ML) models like Convolutional Neural Networks (CNNs). These advancements highlight the potential of HAR to transform healthcare services, ensuring better patient care and efficient resource utilization [Tan, 2018; Wang, 2018; Bibbò, 2023].

HAR is widely applied in various application fields, including robotics and the Internet of Things (IoT), but its role in healthcare is particularly crucial for monitoring patients such as children, the elderly, or the disabled, enabling timely interventions in emergencies such as falls or abnormal respiration [Tan, 2018; Wang, 2018; Bibbò, 2023].

1.2 Defining the Problem Space and Its Challenges

There is a great deal of research on HAR using various activity measurement technologies, including cameras [Zhang, 2018a; Gupta, 2024], sensors [Demrozi, 2020], and wireless communications (self-powered sensors). Despite the advancements in HAR, there remain significant challenges and limitations in current methodologies, particularly those that rely on wearable sensors and camera-based systems. Wearable sensors require consistent user compliance and may cause discomfort, while cameras pose privacy concerns and are limited by line-of-sight constraints [Zhang, 2018a; Gupta, 2024]. These issues make it difficult to implement continuous and unobtrusive activity monitoring in healthcare settings. Moreover, using cameras in a residential environment requires good lighting, while sensor data can be inaccurate if not worn properly.

Wi-Fi CSI offers a promising alternative for HAR, leveraging existing wireless infrastructure to monitor activities without the need for additional devices. Wi-Fi routers are affordable, power-efficient, insensitive to lighting conditions, easy to implement, and have a long range of connectivity [Tan, 2018; Wang, 2018; Zou, 2017]. However, Wi-Fi CSI data for HAR comes with its own set of challenges. The quality of CSI data can be affected by environmental factors and multipath effects, making it complex to process and analyze. Additionally, developing models that can handle the high-dimensional and noisy nature of CSI data while maintaining real-time processing capabilities is a significant hurdle [Strohmayer, 2024; Fard Moshiri, 2021]. Technical challenges include the high computational requirements of existing models, which limit their deployment on devices with limited resources. Furthermore, the complexity of CSI data, influenced by environmental factors, poses difficulties in accurate interpretation and activity recognition.

Data-related challenges include the small and imbalanced nature of available datasets, which can lead to biased models and poor generalization. Effective data augmentation (DA) techniques are needed to address these issues and improve model performance. Designing a lightweight CNN that balances accuracy and efficiency is another significant challenge. To be practical for real-time applications, the model must achieve high accuracy while minimizing computational complexity. Finally, the evaluation and benchmarking of HAR models present challenges in ensuring fair and meaningful comparisons. Standardized testing subsets and metrics are essential to validate the performance improvements of new models. Addressing these challenges is crucial for advancing HAR research and ensuring its practical implementation in healthcare.

In light of these challenges, HAR using Wi-Fi CSI data holds considerable promise for enhancing continuous and unobtrusive patient monitoring, particularly in healthcare environments. The specific problem addressed in this study is accurately recognizing various indoor activities based on the subtle variations in CSI data. This research aims to overcome the aforementioned obstacles by developing a robust and efficient HAR methodology that leverages Wi-Fi CSI data, setting a new benchmark for future studies.

1.3 Preliminaries

1.3.1 Overview of Time-Series Data Augmentation Techniques

In the realm of deep learning (DL), one practical solution to the constraints of limited data is the generation of synthetic frames (i.e., single time-series instances) through DA. This method serves as a universal, model-independent approach aimed at enhancing the generalization capabilities of models. By augmenting data, we not only mitigate overfitting but also expand the decision boundary, thereby increasing the model's robustness to real-world variations. Such enhancements are crucial for addressing the inherent challenges in real datasets, such as small sizes and class imbalances, which can severely impede the effectiveness of DL techniques. To navigate these issues, this subsection introduces and explores two main families of time-series DA techniques: random-transformation-based and pattern-mixing-based [Iwana, 2021a ; Alzubaidi, 2021 ; Fernández, 2018].

Inspired by image DA, the random-transformation-based DA techniques generate new frames $\boldsymbol{x'}_{\text{method}}$ by applying transformations to existing ones $\boldsymbol{x} = (\boldsymbol{x_1}, \ldots, \boldsymbol{x_T})$. Various types of transformations are applied to time-series data frames, such as those related to magnitude, time, or frequency domains [Iwana, 2021a ; Ashfaq, 2024].

Magnitude domain transformations occur with respect to the amplitude or value axes of a time series frame. These transformations modify the values of each element while keeping the time steps constant. Examples include:

— **Jittering**: it adds noise, usually Gaussian noise, to the time series frames, as described by the equation:

$$x'_{\text{jittering}} = (x_1 + \epsilon_1, \ldots, x_t + \epsilon_t, \ldots, x_T + \epsilon_T)$$

where ϵ_t is the noise added to each sample (data point) x_t at time step t, such that $\epsilon \sim N(0, \sigma_\epsilon^2)$. Jittering is an effective transformation-based DA method as it increases the generalization of neural networks by creating new frames under the assumption that unseen test frames are only different from the training frames by a factor of noise. It is particularly useful in mitigating time series drift, which occurs when the data distribution changes due to the introduction of new data.

— **Scaling**: The Scaling technique is applied to change the global magnitude of a time series frame by multiplying it with a random scalar value called the scaling parameter α, as described by the equation:

$$x'_{\text{scaling}} = (\alpha x_1, \ldots, \alpha x_t, \ldots, \alpha x_T)$$

The scaling parameter α is a random variable with a Gaussian distribution, such that $\alpha \sim N(1, \sigma_\alpha^2)$. Scaling is different from the image domain scaling, as it modifies the magnitude of the elements rather than resizing the time series. This technique has been effectively used in various applications, including the classification of sensor data.

Time domain transformations affect the time steps of a time series frame, thereby modifying the temporal dimension. Examples include:

— **Slicing**: Slicing transforms time series frames by randomly cropping a window of size W from the samples' time steps, per equation:

$$x'_{\text{slicing}} = (x_\varphi, \ldots, x_t, \ldots, x_{W+\varphi})$$

where φ is a random integer value $\varphi \in \mathbb{N}$, such that $1 \leq \varphi \leq T - W$. This method, also known as window slicing, is analogous to cropping in image augmentation, and it helps in creating varied frames by trimming the ends of the time series frames. When required to maintain a consistent input size for downstream models, various approaches can be applied to restore the original sequence length, such as interpolation, padding, or other length restoration techniques.

— **Time Warping**: it disturbs the time series frames with respect to the time axis. Window warping, random warping, vowel stretching, time stretching, and time perturbation are all time warping methods. Window warping, for example, is a specific method that stretches or contracts a random window of the frame using a multiplier value, often 2 for stretching and 1/2 for contracting [Iwana, 2021a]. Time warping with a smooth warping path can be described by the equation:

$$x'_{\text{time warping}} = (x_{\tau(1)}, \ldots, x_{\tau(t)}, \ldots, x_{\tau(T)})$$

where $\tau(\cdot)$ is a warping function that warps the time steps based on a smooth curve. This smooth curve is often defined by a cubic spline $S(u)$ with knots $u = (u_1, \ldots, u_I)$, where the height of the knots u_i are taken from a distribution. This ensures a smooth transition between stretches and contractions.

The pattern-mixing-based DA techniques generate new patterns (where "patterns" in [Iwana, 2021a] are equivalent to what we call "frames") by combining two or more similar time series frames. These techniques operate in the magnitude, time, or frequency domains, ensuring that the resulting augmented data retains meaningful characteristics from the original frames without relying on the assumption that any transformation is typical for the dataset [Iwana, 2021a ; Iwana, 2021b].

In the time domain mixing, time series frames are combined based on the time axis. One such technique is *Guided Warping* (GW), which combines time series by warping a reference frame to a teacher frame using Dynamic Time Warping (DTW). DTW is a classic, yet effective, method for determining an optimized distance measure between time series. It measures the distance between two time series $\boldsymbol{r} = (r_1, \ldots, r_I)$ and $\boldsymbol{s} = (s_1, \ldots, s_J)$, where elements r_i and s_j can be univariate or multivariate, with dimensions $r_i = (\alpha_1, \ldots, \alpha_U)^T$ and $s_j = (\beta_1, \ldots, \beta_V)^T$, respectively [Iwana, 2021a ; Iwana, 2021b].

DTW finds the optimal alignment between two sequences by non-linearly matching elements along the time dimension, thereby removing temporal distortions. It achieves this by warping the sequences

to minimize the global cost under constraints. The alignment is determined by calculating a minimal cumulative sum matrix $D(i, j)$ using the recurrent function:

$$D(i,j) = C(r_i, s_j) + \min_{(i',j') \in \{(i,j-1),(i-1,j),(i-1,j-1)\}} D(i', j')$$

where $D(i, j)$ is the cumulative sum of the i-th and j-th elements and $C(r_i, s_j) = \|r_i - s_j\|$ is the local distance between r_i and s_j. The Euclidean distance is used as the cost function in this context. The global distance is then defined as the value at $D(I, J)$ [Iwana, 2021a ; Iwana, 2021b].

GW leverages the warping path obtained from DTW to align the elements of a reference frame to a teacher frame, resulting in a time-warped version of the reference frame. This method differs from simple averaging because the mixing occurs exclusively in the time domain [Iwana, 2021a ; Iwana, 2021b].

We use *Random Guided Warping* (RGW), a variant of GW that selects a random intra-class teacher frame. RGW warps the elements of one frame to the time steps of the other, preserving the temporal structure while introducing variations. Specifically, the RGW-D version employs standard DTW with a symmetric slope constraint and a warping path boundary constraint of 10% of the time series length. The process involves the following steps:

1. Calculate the minimal cumulative sum matrix $D(i, j)$ using the above recurrent function.
2. Trace back from $D(I, J)$ to the origin $D(0, 0)$ to find the minimal warping path.
3. Create the new time series $\boldsymbol{s'}$ by warping $\boldsymbol{s}$ according to the time steps of $\boldsymbol{r}$ under the warping path constraints.

Using this method, it is possible to synthesize a large number of augmented time series by randomly selecting pairs of frames within the same class and applying the guided warping process [Iwana, 2021a ; Iwana, 2021b].

1.3.2 Theoretical Background on Convolutional Neural Networks

In this subsection, we provide a general overview of how Convolutional Neural Networks (CNNs) work on a theoretical level. We use this type of deep *Artificial Neural Network* (ANN) to deal with our multidimensional time-series classification task. A CNN is composed of two main components: the Convolutional and Pooling layers. Understanding these components helps researchers to build the suitable model architecture required for their tasks. The architecture choice is a critical factor in improving the model's accuracy and efficiency. Mainly, CNNs were intended for processing images, as they are better at emulating the human visual system than other ANN architectures. They extract the basic features of an image, such as vertical and horizontal bars, in an end-to-end manner, relying on the mentioned layers. Note that a Convolutional layer has fewer parameters than a fully connected Dense layer [Yang, 2023a ; Alzubaidi, 2021 ; Goodfellow, 2016 ; Zhao, 2017 ; Yi, 2017 ; Sadouk, 2018].

The Convolutional layer is the fundamental building block of CNNs. It uses a set of fixed-size kernels (filters) to perform convolution operations on an input feature map. The kernel, also called a feature detector, is a grid of weights (matrix). Traditionally, these filters were used in image processing practices, like detecting edges and creating effects (blur). Recently, in the DL field, they are used by CNNs to extract different features of an image. The convolution operation works by applying (overlaying) a kernel on a submatrix of an input feature map, where the submatrix suits the kernel size. The result of this operation at each position (x, y) in the output feature map is calculated using the following equation:

$$O(x,y) = \sum_m \sum_n K(m,n) \cdot F(x+m, y+n)$$

where $K(m, n)$ represents the elements of the kernel matrix, $F(x + m, y + n)$ denotes the corresponding elements of the input feature map at the positions shifted by m and n relative to the location (x, y). This equation essentially sums the element-wise product of the kernel and the feature map region, resulting in a single output value at each position in the output feature map. This process is repeated across the entire feature map to produce a complete output feature map, which captures the filtered features such as edges or textures depending on the kernel's values. Each kernel outputs a single feature map. The kernels are grids of weights where these weights are the trainable parameters of a CNN model. To perform well in its classification task, the model learns these kernels' values (weights) during the training phase. Upon building a CNN, there are several architectural decisions (settings) for each Convolutional layer, including the size and number of kernels, the stride, the padding, and the depth [Alzubaidi, 2021 ; Goodfellow, 2016].

The kernel's size represents the number of weights for the height and width of a grid, usually odd-valued. The stride determines the step size for sliding the kernel on the input feature map. Padding indicates the number of extra columns and rows to add evenly outside an input feature map. When padding is set to "valid", no padding is done, whereas when it is "same", zeros are added evenly to the left/right and up/down of the input feature map. After applying a kernel on an input feature map, its size decreases. Consequently, after successive Convolutional layers, this size reduces significantly and it becomes too small. Therefore, these layers use padding to preserve the size of their input feature maps. The result of using "same" padding with stride equals 1, is an output feature map having the same size as the input one. The kernel's third dimension, known as depth, corresponds to the number of channels in the input feature map. For instance, the depth is one for a single-channeled (grayscale) image, and three for a colored image with RGB channels. This ensures that each filter fully encompasses the depth of the input data. In the context of time series data, the depth can refer to the number of variables (dimensions) being considered, allowing the model to learn complex interactions between them. The overall dimensions of the kernel, expressed as height $\times$ width $\times$ depth, are thus defined with the depth matching the number of input channels. This comprehensive representation highlights how each kernel interacts with the entire volume of the input feature map [Alzubaidi, 2021 ; Goodfellow, 2016 ; Zhao, 2017 ; Yi, 2017 ; Sadouk, 2018].

When applying CNNs to multidimensional time-series classification, these architectural choices become crucial. Time series data have inherent temporal and, depending on the domain, possibly frequency, spatial, or other multi-variable dependencies that need to be captured for accurate classification. The convolution operation, with its ability to apply filters across both temporal and (for instance) frequency or spatial dimensions, is particularly well-suited for this task. By designing kernels that can move along the time axis, CNNs can capture temporal dependencies effectively. Additionally, if the time series data have multiple variables (dimensions), the depth of the kernel can accommodate these dimensions, allowing the model to learn complex interactions between them [Goodfellow, 2016 ; Zhao, 2017 ; Yi, 2017 ; Sadouk, 2018].

The Pooling layer is the other main component for building CNNs. Using a pooling grid, it performs pooling operations on an input feature map to reduce its dimensions (downsampling) without significant information loss. This layer takes as input a series of feature maps and it outputs a new series with lower dimensions. The pooling grid, also called a window, is a grid of fixed size over which the layer applies the pooling operation. This window is shifted by strides (step size) along the dimensions (height and width) of an input feature map. There are two primary types of pooling operations commonly used:

— Max Pooling: Where the maximum value over the sliding window for each channel of the input feature map is selected. The operation can be mathematically represented as:

$$O(x, y) = \max\{F(i, j) \mid x \leq i < x + k,\ y \leq j < y + k\}$$

— Average Pooling: Where the average value over the sliding window is calculated. This can be represented as:

$$O(x, y) = \frac{1}{k^2} \sum_{i=x}^{x+k-1} \sum_{j=y}^{y+k-1} F(i, j)$$

Using Pooling layers is useful for reducing the variability in the hidden activations, which enables extracting the dominant features. These features are rotational and positional invariant. Note that Max Pooling performs better than Average Pooling in discarding the noisy activations [Alzubaidi, 2021 ; Goodfellow, 2016].

In the context of multidimensional time-series classification, pooling operations help in managing the temporal resolution of the data. By reducing the temporal dimensions, pooling layers allow the network to focus on the most salient features, mitigating the effect of noise and variability over time. This is particularly advantageous when dealing with long sequences, where pooling can help in maintaining computational efficiency while preserving critical temporal patterns [Goodfellow, 2016 ; Zhao, 2017 ; Yi, 2017 ; Sadouk, 2018].

Overall, the combination of convolutional and pooling layers in CNNs provides a powerful framework for capturing and leveraging the complex dependencies present in multidimensional time series data. This makes CNNs particularly effective for tasks involving temporal dynamics and inter-variable relationships, which are essential for accurate time-series classification [Goodfellow, 2016 ; Zhao, 2017 ; Yi, 2017 ; Sadouk, 2018].

1.4 Contributions

This chapter outlines the significant contributions of this research in the domain of Wi-Fi CSI-based HAR. These contributions highlight the advancements made in addressing key limitations and improving the state of the art in this field.

1. Lightweight CNN Architecture

— **Development of HAR-LightCNN**: A novel lightweight CNN model named HAR-LightCNN was developed, specifically optimized for real-time HAR using Wi-Fi CSI data. This model significantly reduces computational complexity and memory usage compared to existing lightweight CNN architectures, making it suitable for real-time applications and deployment in resource-constrained environments such as residential healthcare settings.

— **Efficiency Optimization**: The architecture of HAR-LightCNN balances the depth (number of layers) and width (number of units or filters per layer) to operate efficiently within the constraints of devices with limited computational resources. This optimization was substantiated through comparisons with models like ShuffleNet, SqueezeNet, and MobileNet, demonstrating substantial reductions in parameters, prediction complexity, and memory requirements.

2. Carefully Selected Data Augmentation Techniques

— **Time-Series Data Augmentation**: The implementation of carefully selected time-series DA techniques to generate additional training samples and balance the dataset. This approach addresses the limitations of small-sized and imbalanced data, enhancing the model's robustness and generalizability, particularly for minority classes.

— **Enhanced Training Procedures**: Rigorous validation procedures, including holdout and k-fold cross-validation, were applied to ensure reliable training. These procedures involved train-test and train-validation splitting at appropriate preprocessing stages to avoid shared data samples between training and testing subsets, preventing data leakage and ensuring robust and unbiased evaluation.

3. Improved Evaluation Framework

— **Benchmarking Against State-of-the-Art Models**: The HAR-LightCNN model was benchmarked against various state-of-the-art models using standardized testing subsets and evaluation metrics, ensuring fair and valid comparisons. This benchmarking highlighted HAR-LightCNN's superior performance, achieving higher accuracy and a significant reduction in computational demands, making it ideal for real-time applications.

— **Standardized Testing and Evaluation Metrics**: Preparation of unified testing subsets from two datasets used in this study provided a reliable and unbiased evaluation environment. This approach facilitated meaningful benchmarking and ensured that comparisons with literature approaches were fair and valid, addressing the limitation of unfair benchmarking practices.

4. Performance Improvements

— **Significant Accuracy and Efficiency Gains**: The HAR-LightCNN model demonstrated significant improvements in testing accuracy and F1-score, outperforming existing models. The model's lightweight design resulted in fewer parameters and less prediction complexity compared to a 2D-CNN benchmark, proving its efficiency and effectiveness.

— **Robust Real-Time Application**: The lightweight nature and optimized architecture of HAR-LightCNN ensure efficient real-time processing capabilities, making it suitable for practical deployment in healthcare settings for patient monitoring. The model's robustness in handling unseen data and its stability in response times during real-time applications validate its practicality and applicability.

These contributions collectively advance the research domain of Wi-Fi CSI-based HAR, providing a foundation for future research and practical implementations in healthcare and other application fields. This research not only addresses critical limitations in existing methodologies but also sets new benchmarks for efficiency and performance in real-time HAR.

1.5 Organization of the Chapter

The rest of this third chapter (first chapter of the contributions part II) is structured as follows:

— Section 2 sets the stage for our approach—*HAR-LightCNN* methodology—by summarizing its motivation, core phases (*data acquisition*, *data framing*, and *activity recognition*), and how it leverages Wi-Fi CSI and CNN-based architectures for multi-activity recognition.

— Section 3 details the steps required to transform raw CSI signals into robust training and testing subsets. This includes partitioning the datasets, ensuring fair evaluation, and applying time-series DA to mitigate small-sized and imbalanced data limitations.

— Section 4 introduces the architectural design and efficiency optimizations of the proposed CNN model. It explains why CNNs are well-suited for multidimensional time-series analysis and how our lightweight design addresses high model complexity while maintaining strong classification performance. Additionally, it provides details on the software and hardware environments used for constructing and training the model.

— Section 5 presents a comprehensive evaluation of HAR-LightCNN, comparing its effectiveness and efficiency across different datasets and validation procedures. It also benchmarks our model against state-of-the-art approaches, highlighting key performance gains.

— Section 6 concludes the chapter by summarizing the main findings and discussing the benefits of our methodology for healthcare services. It further outlines potential research directions that build upon the advantages of Wi-Fi CSI-based HAR.

By the end of this chapter, we establish a robust and efficient CSI-based HAR solution for patient monitoring service. In the following Chapter 4 (second chapter of the contributions part), we will expand our research focus to address additional limitations, particularly those related to *fall detection task* in healthcare scenarios, thereby advancing the scope of CSI-based HAR systems for emergency response service.

2 HAR-LightCNN Methodology Overview

2.1 Introduction to the Methodology

HAR is a crucial field for identifying and predicting single or multiple users' behaviors, such as periodic body part movements, predefined actions, and spontaneous activities. HAR is especially significant in healthcare applications and it is vital for patient monitoring, enabling timely interventions [Tan, 2018 ; Ge, 2023].

The HAR-LightCNN methodology leverages Wi-Fi wireless communication technology to address the limitations posed by traditional activity measurement technologies, such as cameras and sensors. Cameras, for instance, require good lighting and can only detect actions within their line-of-sight, while sensor data can be inaccurate if not worn properly. Wi-Fi signals, on the other hand, can penetrate walls and are less invasive, offering fewer privacy concerns. Wi-Fi routers are affordable, power-efficient, and insensitive to lighting conditions, making them an ideal choice for activity recognition in residential environments [Tan, 2018 ; Wang, 2018 ; Zou, 2017].

Wi-Fi signals interact with obstacles like furniture or moving humans, creating multipath effects that can be analyzed using CSI. CSI provides detailed information about the Wi-Fi signal's journey from the transmitter to the receiver, capturing the effects of reflection, scattering, fading, and power decay. Analyzing these interactions allows for the recognition of human activities [Chen, 2023a ; Wang, 2018 ; Zou, 2017].

DL methods, particularly CNNs, have shown superior performance in handling CSI data for HAR compared to classical ML techniques. Previous research has explored various DL architectures and data modalities to enhance accuracy and robustness in Wi-Fi CSI-based HAR. Approaches have included transforming CSI data into RGB images for 2D-CNN models, using Attention-based Bi-directional Long Short-Term Memory (ABLSTM) models [Fard Moshiri, 2021], and employing semi-supervised Generative Adversarial Networks (GANs) for generating synthetic training samples [Xiao, 2019].

Building on these advancements, the HAR-LightCNN methodology introduces a lightweight CNN designed for recognizing seven single-user human activities (bending, falling, lying down, running, sitting down, standing up, and walking) using Wi-Fi CSI data. This approach eliminates the need for wearables, making it particularly useful for monitoring patients, especially the elderly, in resource-constrained environments. The lightweight design of HAR-LightCNN reduces both computational complexity and memory usage while maintaining robust performance, making it suitable for real-time activity recognition applications.

The methodology comprises three main phases: data acquisition, data framing, and activity recognition. During data acquisition, Wi-Fi signals are transmitted, and the receiver collects real-time CSI amplitude data. In the data framing phase, the collected data are grouped into frames of fixed duration, processed to comply with the input requirements of the activity recognition phase. This final phase involves categorizing the activity frames into one of the seven considered daily activities using the designed

and trained HAR-LightCNN classifier.

To address data-related concerns, including the limitations of small-sized and imbalanced data, the methodology incorporates time-series DA techniques. These techniques generate additional training samples, enhancing the model's performance and generalizability across minority classes. The effectiveness of HAR-LightCNN is validated using two testing subsets, demonstrating superior performance compared to state-of-the-art models. HAR-LightCNN is smaller in terms of parameters, less complex in terms of *Floating-point Operations* (FLOPs), and more precise in terms of accuracy compared to the 2D-CNN model proposed in previous research [Fard Moshiri, 2021]. This underscores its improved efficiency and effectiveness for real-time activity recognition.

In summary, HAR-LightCNN methodology presents an innovative, lightweight CNN-based approach for HAR using Wi-Fi CSI data. By addressing the limitations of high model complexity and data-related concerns through efficient architectural design and effective DA techniques, HAR-LightCNN sets a new benchmark for future research in Wi-Fi CSI-based HAR.

2.2 Objectives and Design Principles

The primary objective of the HAR-LightCNN methodology is to develop a robust and efficient HAR system that can be deployed in resource-constrained environments, such as residential healthcare settings, without relying on wearable devices. This approach aims to address several key limitations—first catalogued in Section 4 and prioritised in Section 7—within three categories in the domain of Wi-Fi CSI-based HAR:

1. **Data-Related Concerns**:
   - **Small-Sized Data**: To address the limitation of small-sized data, HAR-LightCNN incorporates DA techniques. Specifically, we implemented a train data over-framing preprocessing step, utilizing time-series DA techniques to generate additional training frames for the seven activities. This approach enriches the training data with diverse samples, enhancing the robustness and generalizability of our multi-activity classifier. By applying these techniques during the training phase, the model improves its effectiveness.
   - **Imbalanced Data**: To address the limitation of imbalanced data, HAR-LightCNN also employs DA techniques to balance the training subset. We applied a train data over-framing preprocessing step to generate additional training frames for the minority classes, matching the number of frames of the majority class. This ensures a balanced training subset, enhancing the robustness and generalizability of our multi-activity classifier. By addressing this limitation, the model not only improves its overall effectiveness but also enhances the performance on minority classes, leading to better detection and classification of these activities.
2. **Processing Technique Concerns**:
   - **High Model Complexity**: To address the limitation of high model complexity, HAR-LightCNN is based on a lightweight CNN, designed to significantly reduce computational complexity and memory usage compared to existing lightweight CNN architectures. This is achieved by optimizing the architecture to balance depth (number of layers) and width (number of units or filters per layer), ensuring that the model operates efficiently within the constraints of devices with limited computational resources. This optimization occurs at the model-building phase, resulting in reduced complexity and improved efficiency.
3. **Evaluation Process Concerns**:
   - **Unfair Benchmarking Against Others**: To address the limitation of unfair benchmarking, the HAR-LightCNN methodology benchmarks its performance using standardized testing subsets and evaluation metrics, ensuring fair and valid comparisons with existing models. By

employing a unified benchmarking framework, this approach highlights genuine advancements and improvements, providing a robust validation of the model's effectiveness and efficiency. This ensures that performance comparisons are meaningful and technically sound.

— **Inadequate Testing Procedures**: The HAR-LightCNN methodology addresses the limitation of inadequate testing procedures by implementing rigorous validation procedures that prevent overlap between training, validation, and testing subsets. By ensuring that activity frames used for validation and testing are distinct and non-overlapping, the methodology maintains the integrity of the evaluation process. This prevents the issue of data leakage and provides a more accurate reflection of the model's performance in real-world scenarios.

The HAR-LightCNN methodology comprises three main phases: data acquisition, data framing, and activity recognition.

— **Data Acquisition**: During data acquisition, Wi-Fi signals are transmitted, and the receiver collects real-time CSI amplitude data. This phase leverages the properties of Wi-Fi signals, which can penetrate walls and provide detailed information about the environment, including the presence and movement of individuals.

— **Data Framing**: In the data framing phase, the collected data are grouped into frames of a fixed duration (1 second). These activity time series data frames are then processed to comply with the input requirements of the activity recognition phase. This may include downsampling or oversampling to ensure the data is suitable for classification.

— **Activity Recognition**: The final phase involves categorizing the activity frames using a lightweight CNN-based multi-activity classifier. The classifier's architecture includes three convolutional (Conv) blocks, a Fully Connected (FC) block, and an output block. Each Conv block contains Conv2D, LeakyReLU, MaxPooling2D, and BatchNormalization layers, which help extract and normalize features from the input data. The FC block consists of Flatten, Dense, and Dropout layers, which reduce overfitting and improve generalization. The output block uses a Dense layer with softmax activation to produce a probability distribution for each activity, with the highest score indicating the predicted activity.

The HAR-LightCNN methodology effectively addresses the limitations related to high model complexity and data-related concerns. FIGURE 3.1 provides a schematic representation of this methodology, illustrating the three phases, with the CNN being part of the activity recognition phase.

By integrating these design principles and objectives, the HAR-LightCNN methodology demonstrates superior performance and efficiency compared to state-of-the-art models. The lightweight nature of the model makes it an ideal candidate for deployment in real-time activity recognition applications, contributing to the advancement of Wi-Fi CSI-based HAR research.

3 Data Preparation and Preprocessing

3.1 Datasets Overview and Initial Setup Considerations

In our approach, we utilized two publicly available activity CSI datasets to evaluate the proposed HAR methodology, HAR-LightCNN. These datasets (First and Second CSI Datasets), introduced in Chapter 1 (Section 8), offer comprehensive CSI data for seven common daily activities performed in an indoor environment. This subsection provides an overview of these datasets and the initial setup considerations for using them in our research.

The 1st dataset we used is provided by [Fard Moshiri, 2021]. The data collection was conducted in an indoor environment using a transmitter and receiver equipped with a Nexmon CSI extractor [Gringoli,

2019]. This setup operated on a single channel with 52 subcarriers, capturing CSI amplitude data at a sampling rate of 200 samples per second. The dataset includes recordings of seven daily activities: bending, falling, lying down, running, sitting down, standing up, and walking. These activities were performed 20 times by three participants of different age groups, resulting in a total of 420 CSV files (60 files per activity). Each file comprises 52 columns, corresponding to the subcarriers, and contains between 600 to 1100 rows, depending on the activity's duration.

The 2nd dataset, collected by [Yousefi, 2017], also focuses on the same seven daily activities but was gathered using a different setup. This dataset employed an Intel 5300 Wi-Fi Network Interface Card (NIC) with three antennas at the receiver end and a single antenna at the transmitter end, producing data across three channels with 30 subcarriers each. The sampling rate for this dataset was 1000 samples per second. Six participants performed each activity multiple times, resulting in 557 signal data files (approximately 80 files per activity). Each CSV file contains 90 columns (representing the 30 subcarriers across the three channels) and around $20,000$ rows, reflecting the duration of each activity trial.

These datasets form the foundation for testing our HAR-LightCNN methodology, providing diverse and extensive data to develop and evaluate its performance in recognizing human activities based on Wi-Fi CSI data.

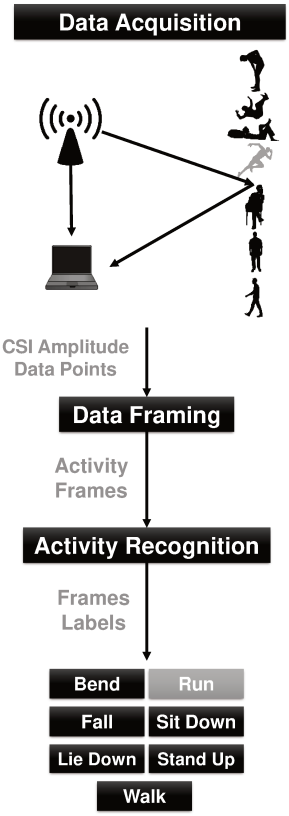


FIGURE 3.1 – HAR-LightCNN three-phase methodology schema: This schematic representation illustrates the three phases of HAR-LightCNN methodology, namely data acquisition, data framing, and activity recognition, with the CNN being part of the activity recognition phase.

3.2 Overview of the Preprocessing Steps

This subsection will delve into the preprocessing steps and procedures applied to prepare the datasets for robust training and unbiased evaluation of our methodology.

3.2.1 Testing Subset Preparation for Ensuring Fair Evaluation

To ensure an unbiased evaluation of our multi-activity classifier, we carefully prepared two testing subsets from both datasets used in this study. The 1st dataset from [Fard Moshiri, 2021] was split into 85% training and 15% testing (one of the commonly used split percentages in ML solutions), resulting in 357 and 63 files, respectively. This split was chosen to allocate a larger portion of the data for model training, considering its multidimensional nature, while still reserving enough data for evaluation. The seven activities in each of these two subsets have an equal number of files (signals).

To process this data, we applied overlapped framing with a frame size of 200, matching the sampling rate, and a slide size of 40, generating 1-second activity snapshots. Framing is the process of extracting samples from the signal and bundling them into frames of sufficient time steps to make sense from the classification task perspective. This approach produced $5,681$ training frames and 994 testing frames of size $(200, 52)$. In an activity labels array, we added the class label corresponding to each extracted frame. By splitting the dataset before framing, we ensured that the testing frames were unseen during model training, providing an unbiased evaluation [Varga, 2024 ; Brownlee, 2018 ; Goodfellow, 2016].

However, we noticed a slight class imbalance in the testing subset, with the minority class (sitting down) having 128 frames. To address this, we discarded extra frames to match the number of frames of the minority class, ensuring a fair and unbiased evaluation. The class distribution in the test data array before balancing is shown in FIGURE 3.2. To prevent any order effects on model evaluation metrics, we shuffled the testing data randomly. This comprehensive preprocessing ensured the testing subset was suitable for providing an objective evaluation of the trained model, reflecting its performance on real-world unseen data [Brownlee, 2018 ; Goodfellow, 2016 ; Fernández, 2018].

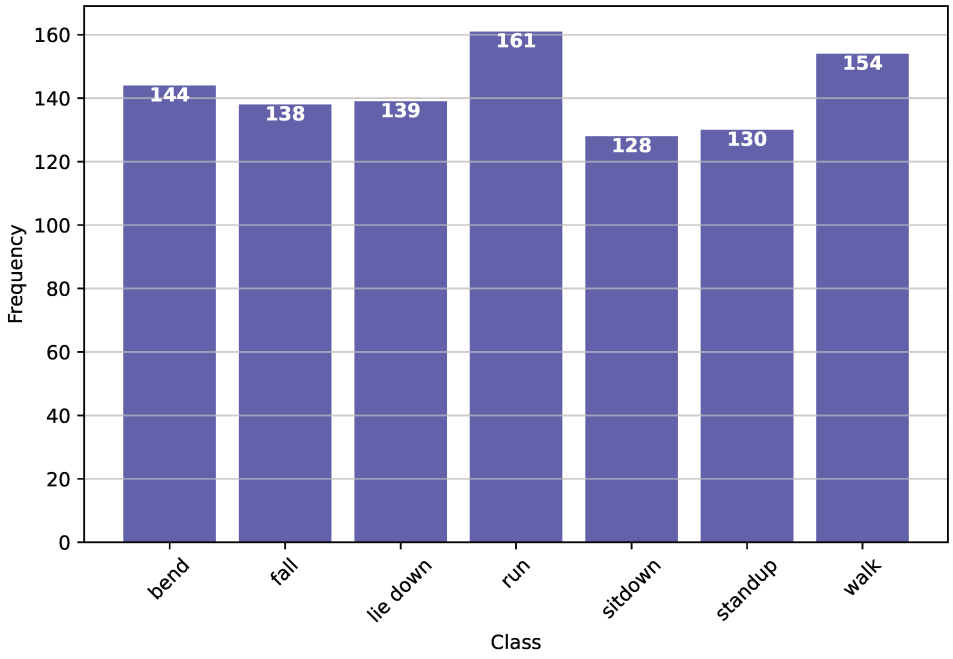


FIGURE 3.2 – Class Distribution in test data array of 1st dataset before balancing: This figure shows the initial distribution of classes in the test data array from the first dataset, highlighting the imbalance with the minority class (sitting down) having fewer frames.

For the 2nd dataset from [Yousefi, 2017], we applied similar preprocessing steps. This dataset was also split into 85% training and 15% testing, resulting in 473 and 84 files, respectively. Each activity trial signal was framed with a frame size of 1000 (equal to the sampling rate), representing a 1-second

snapshot. The frames overlapped with a slide size of 200 and were downsampled by a factor of 2 to avoid memory errors and reduce computational cost, resulting in frames of size $(500, 90)$. We obtained $8,497$ training frames and $1,451$ testing frames.

We observed a severe class imbalance in the test data array, visualized in FIGURE 3.3. To balance the classes, we discarded extra frames to match the minority class (standing up) with 97 frames. Additionally, we shuffled the testing data to ensure random class order, achieving an unbiased evaluation.

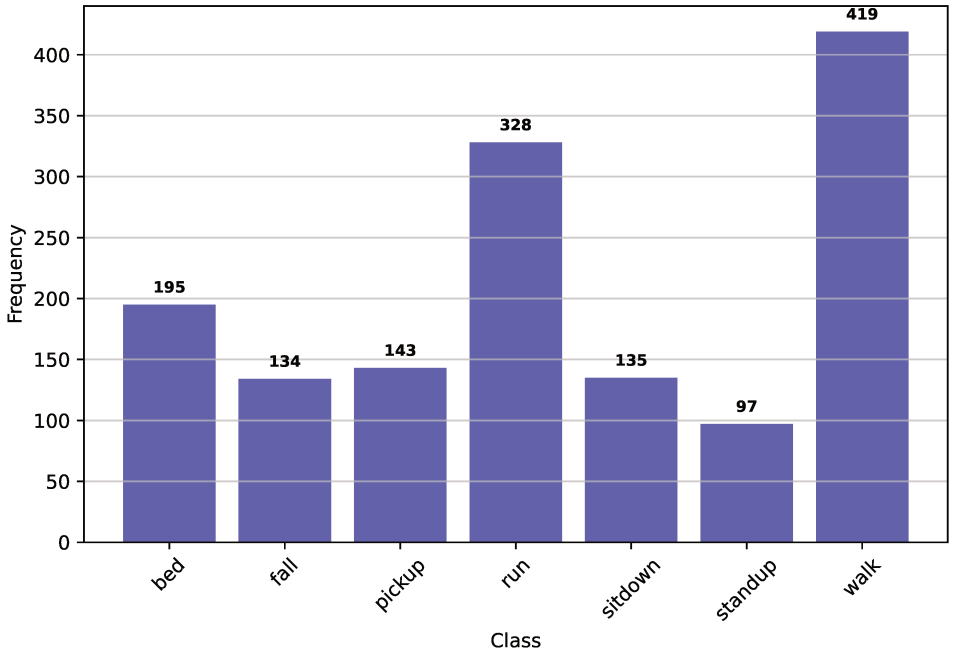


FIGURE 3.3 – Class Distribution in test data array of 2nd dataset before balancing: This figure illustrates the initial class distribution in the test data array from the second dataset, showing a severe imbalance with the minority class (standing up) having the fewest frames.

3.2.2 Validation Procedures for Ensuring Reliable Training

In the domain of Wi-Fi CSI-based HAR, a common limitation within the "Evaluation Process Concerns" category pertains to "Inadequate Testing Procedures," which impacts the reliability of evaluation metrics. Notable examples include using the same activity frames for both validating and testing models and performing framing before splitting the dataset into training and testing subsets. Both practices lead to biased results that do not accurately reflect the ML model's performance. Using the same activity frames for both validation and testing skews evaluation accuracy, as it causes overlap in the data the model is exposed to during training and evaluation. Similarly, framing the dataset before splitting results in shared data samples between the training and testing subsets, compromising the integrity of the evaluation. This is particularly problematic with overlapped framing, where data points are bundled into frames that may overlap, leading to common samples being present in both subsets.

In our supervised ML experiments, this limitation is addressed by employing both train-test and train-validation splittings at appropriate stages of preprocessing. Train-test splitting involves initially dividing the dataset into training and testing subsets, ensuring that the testing data remains unseen during the training process. To avoid shared data samples, the dataset is first split into training and testing subsets, followed by the framing process, ensuring that the training and testing frames do not share common data points. Train-validation splitting is then applied to the training subset, further dividing it into training

and validation subsets to tune the model and prevent overfitting. This sequence of preprocessing steps ensures that our model is trained and validated on distinct data frames, providing a robust and unbiased evaluation of its performance [Varga, 2024 ; Brownlee, 2018 ; Goodfellow, 2016].

Simultaneously with preprocessing the testing subsets from the two datasets, the training subsets for model training were also preprocessed. To ensure robust training of our multi-activity classifier, we experimented with two validation procedures: the conventional holdout method and k-fold *Cross-Validation* (CV). These validation procedures involve train-validation splitting, where the training data is further divided into training and validation subsets.

The holdout method involves randomly dividing the training subset frames between training and validation subsets in a stratified manner, preserving the class distribution percentage of the training frames. This stratified splitting ensures that the model properly learns the class distribution in the training data array. Additionally, this method includes random shuffling of the class order before splitting to enhance robustness. At each epoch, we trained the model parameters using the training subset frames. At the end of each epoch, the model was validated on the validation subset frames, using the validation results to optimize the model parameters. This process ensures that while the model does not learn from the validation frames, it occasionally sees them. This distinction underscores why using the same activity frames for both validating and testing is unreliable. Finally, upon completion of the experiment, the model's performance was evaluated on the yet unseen testing subset frames, providing an accurate assessment of its performance [Varga, 2024 ; Brownlee, 2018 ; Goodfellow, 2016].

However, with a small dataset like the 1st dataset from [Fard Moshiri, 2021], dividing it among three subsets (training, validation, testing) can significantly reduce the number of training frames, potentially excluding some activity frames with useful information from the training process. To mitigate the impact of random split choices in the holdout method and address the small-sized data limitation present in the 1st dataset, k-fold CV can be used. This method splits the training subset into k smaller subsets (folds), training on $k-1$ of these folds and validating on the remaining fold. While k-fold CV is more computationally expensive than the holdout method, it is more reliable and avoids many of the holdout method's shortcomings. It does not waste much data and provides a reliable alternative for validation, ensuring robust and generalizable models [Brownlee, 2018 ; Goodfellow, 2016].

Considering these facts, we conducted two experiments with different validation procedures. Each experiment was executed on an independent duplicate of the 1st dataset, with each duplicate following the preprocessing steps. After training on the resulting preprocessed training subsets, we evaluated the trained classifiers on the testing subset frames to determine the best-trained version (instance) of our activity classifier. The first experiment employed the conventional holdout method, repeated N times with a (train: 85%, validation: 15%) split percentage and $N = 4$. We trained four distinct instances of our model on each resultant split to illustrate the impact of split choices on the model's experimental results. In the second experiment, we employed the k-fold CV procedure with k set to 5 (5-fold CV).

Given that k-fold CV is computationally expensive and does not propose a radical approach to solving the small-sized data limitation, we introduced an additional preprocessing step to the training subset before employing the validation procedure. Specifically, we implemented a train data over-framing preprocessing step to mitigate this limitation. In this step, we utilized time-series DA techniques, as introduced in Subsection 1.3.1, to generate additional new training frames for the seven activities (7-classes). This approach aimed to enhance the robustness and generalizability of our multi-activity classifier by enriching the training data with diverse samples. In the subsequent Subsection 3.2.3, we will discuss the costs associated with the small-sized data limitation and describe our proposed data preprocessing step in detail, highlighting its efficacy in addressing this challenge using DA.

For the training subset derived from the 2nd dataset, we visualized the class distribution in the train data

array to check for balance. As illustrated in FIGURE 3.4, there is a severe imbalance of classes in this array.

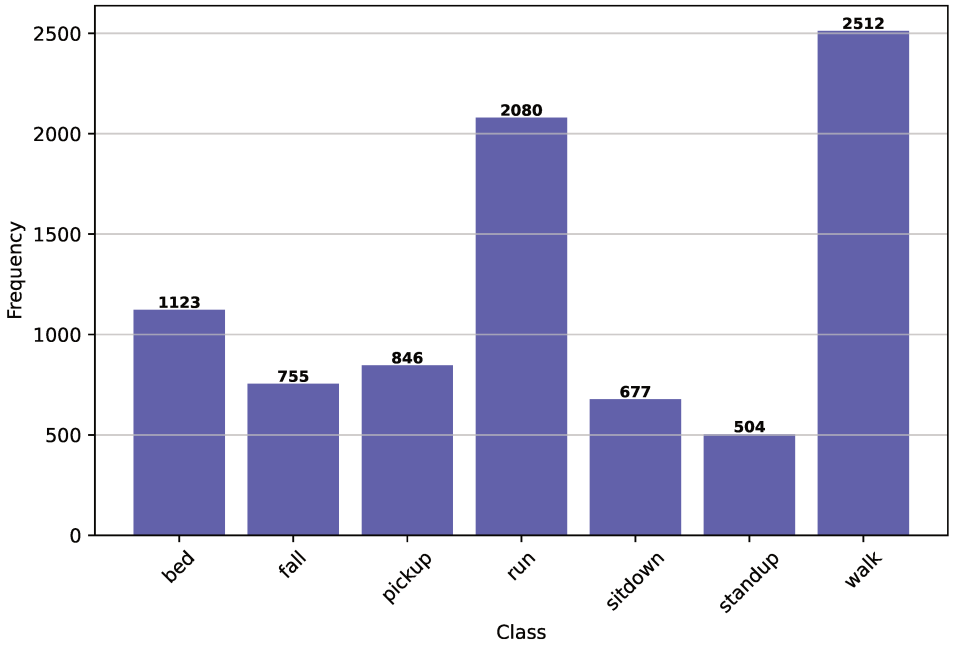


FIGURE 3.4 – Class Distribution in train data array of 2nd dataset before balancing: This figure illustrates the initial class distribution in the train data array from the second dataset, highlighting the imbalance with the majority class (walking) having significantly more frames.

This imbalance indicates that the 2nd dataset of [Yousefi, 2017] suffers from the imbalanced data limitation. We applied the train data over-framing preprocessing step to the training subset before the validation procedure to address this limitation. In this step, we utilized time-series DA techniques to generate additional training frames, matching the number of frames of the majority class (walking, class 6) with 2512 frames. This step ensured a balanced training subset, enhancing the robustness and generalizability of our multi-activity classifier. In the subsequent Subsection 3.2.3, we will explain the costs associated with the imbalanced data limitation and describe our proposed data preprocessing step in detail, highlighting its efficacy in addressing this challenge using DA.

We conducted experiments similar to those performed on the training subset derived from the 1st dataset. The first experiment used the conventional holdout method with the same split percentage (train: 85%, validation: 15%) repeated $N = 4$ times, while the second experiment employed a 5-fold CV. Each experiment was performed on an independent duplicate of the 2nd dataset.

The results of all experiments, including those for both datasets, will be displayed and discussed in Section 5, aiming to determine the best final trained version (instance) of our activity classifier.

The preprocessing steps for both datasets are summarized in FIGURE 3.5, which provides a comprehensive overview of the procedures applied to prepare the data for robust training and unbiased evaluation of our multi-activity classifier.

3.2.3 Implementing Data Augmentation for Enhancing Generalization

Small-sized data, specifically insufficient data quantities, hinders robust model training and validation in the domain of CSI-based HAR. This limitation falls under the "Data-Related Concerns" category. In complex tasks like our multi-activity classification task, DL techniques generally outperform traditional

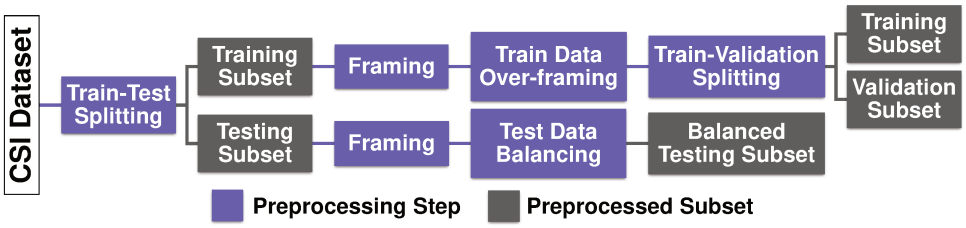


FIGURE 3.5 – Dataset preprocessing steps: This figure illustrates the comprehensive preprocessing steps applied to the CSI datasets, including train-test splitting, framing, over-framing for training data (using DA to generate additional frames), balancing for testing data (by discarding excess frames), and train-validation splitting. Each step is crucial for preparing robust and unbiased datasets for training and evaluating the multi-activity classifier.

ML techniques. However, DL techniques require a large training dataset and considerable computing power. The small size of the training subset impacts the DL model's performance on unseen data. Fitting a model on a training subset with insufficient frames causes overfitting to the training data, resulting in reduced performance. Specifically, insufficient data prevents learning the mapping function approximated by deep neural networks [Alzubaidi, 2021 ; Goodfellow, 2016]. These issues highlight the costs of working with a small-sized dataset, such as reduced generalization from training to unseen data, potentially leading to inaccurate and unreliable predictions.

The 1st dataset from [Fard Moshiri, 2021] faces the small-sized data limitation. To address this limitation, we introduced a train data over-framing preprocessing step. In this step, we utilized time-series DA techniques, as introduced in Subsection 1.3.1, to increase the quantity of the training subset frames. Specifically, on the training frames of each class (activity), we applied four random-transformation-based DA techniques to generate additional frames. These techniques are jittering, scaling, slicing, and window warping. Jittering added Gaussian noise with a mean $(\mu) = 0$ and standard deviation $(\sigma) = 0.03$ to the frames. Scaling multiplied the frames by a random scaling parameter from a Gaussian distribution with mean $(\mu) = 1$ and standard deviation $(\sigma) = 0.02$. Slicing, specifically window slicing, cropped the frames using a window of size set to $0.9 \times$ frame_size, and then interpolated the generated frames back to the original frame length of 200 samples. Window warping selected a random window of size set to $0.1 \times$ frame_size from the frames and warped this window by a multiplier value. Furthermore, on the training frames of the 7 classes (activities), we applied the pattern-mixing-based DA technique, RGW, to generate additional synthetic frames [Iwana, 2021a] [1]. As a result of applying these DA techniques, the total number of training subset frames increased from $5,681$ to $52,360$, with each activity class now containing $7,480$ frames, ensuring a balanced class distribution.

By applying these DA techniques, we aimed to expand the decision boundary of our model and alleviate the overfitting problem. Consequently, we anticipated that these techniques would enhance the model's generalization ability on unseen data, thereby improving its overall performance. Our main goal is to obtain an effective activity classifier. Therefore, to examine the influence of our proposed approach on model performance, we conducted two experiments with different validation procedures as described in the previous subsection. In Section 5, we will display and discuss the results of these experiments.

FIGURE 3.6 illustrates the application of four random-transformation-based DA techniques on two frames, each derived from real CSI data signals representing bending and lying down activities from the 1st dataset. For clarity, only subcarriers 1, 26, and 52 are visualized in each subplot. Each subplot shows the original frame followed by its augmented versions. These visualizations highlight how each

1. Implementation of the augmentation techniques was based on the open-source repository provided by [Iwana, 2021a], available at: https://github.com/uchidalab/time_series_augmentation.

technique modifies the original frame, introducing variability that helps our model generalize better to unseen data. Notably, the transformed frames retain key characteristics of the original activity, ensuring that the augmented data remains relevant for accurate activity classification.

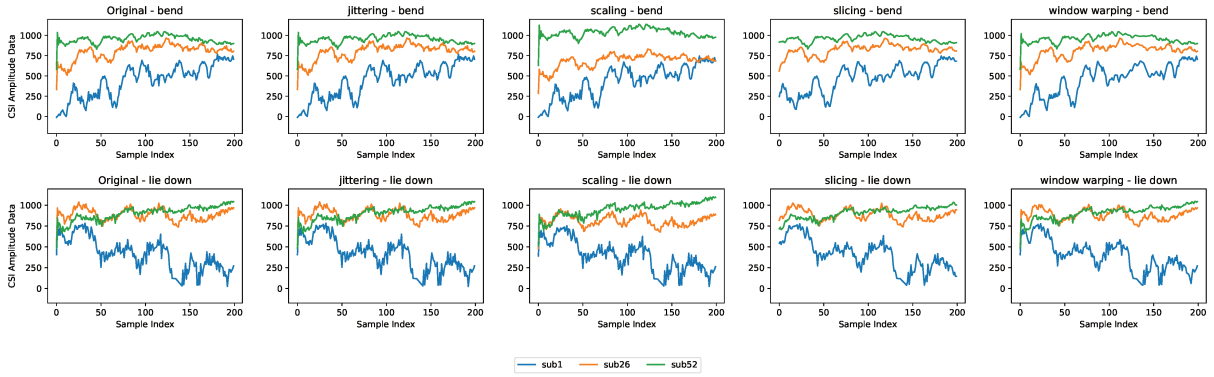


FIGURE 3.6 – Augmentation effect of random-transformation techniques: This figure illustrates the original and augmented frames of bending and lying down activities using jittering, scaling, slicing, and window warping techniques. The augmentation introduces variability while retaining the key characteristics of the original frames, thereby improving the model's generalization ability.

FIGURE 3.7 presents a comparative analysis of original and augmented frames of bending and lying down classes from the 1st dataset by applying the RGW pattern-mixing-based DA technique. Each subplot showcases the original frame, its augmented version, and their absolute differences, with the samples for subcarriers 1, 26, and 52. The Mean Squared Error (MSE) values associated with these frames (2295.1749 and 1392.3665) highlight the effectiveness of the RGW technique in introducing subtle yet significant modifications. These MSE values, substantially lower than's variance the dataset$\sigma^2 = 65,987.3344$, confirm that the alterations to the CSI signal frames are minor and do not drastically distort the underlying data patterns. This level of augmentation ensures the training data remains representative of real-world variations, fostering the development of robust DL models. The subtle changes introduced are sufficient to add the necessary diversity for the models to classify the exact signals encountered during training and their slight variations, thereby enhancing the models' ability to generalize across real-world scenarios.

Imbalanced data, specifically the overrepresentation of certain classes, hinders robust model training and validation in the domain of CSI-based HAR. This limitation falls under the "Data-Related Concerns" category. In multi-class classification tasks, most ML techniques assume that the number of instances in the training dataset for each class label is balanced. Training our model on an imbalanced dataset results in a classifier with reduced classification performance. This classifier tends to misclassify the activity frames of classes with insufficient training frames compared to others [Fernández, 2018 ; Goodfellow, 2016]. Moreover, we may encounter severe analytical errors based on the classifier's experimental results, as these results could obscure the classification errors for the minority classes since they are less represented compared to the majority ones.

The 2nd dataset from [Yousefi, 2017] suffers from the imbalanced data limitation. To address this issue, we introduced a train data over-framing preprocessing step to the training subset before the validation procedure. We proposed utilizing time-series DA techniques to balance the class distribution in our training data array. Initially, we identified the majority class in this array, which is class 6 (walking), having 2,512 frames. Then, we determined the number of frames needed by each other class to match this majority amount. For the training frames of each underrepresented class, we applied several random-transformation-based DA techniques to generate additional frames. These techniques include jittering,



FIGURE 3.7 – Augmentation effect of RGW pattern-mixing technique: This figure compares original and augmented frames of bending and lying down activities. The augmented frames and their absolute differences, along with the MSE values, highlight the subtle yet effective modifications introduced by the RGW technique, ensuring the training data remains representative of real-world variations.

scaling, slicing, and window warping. Specifically:

— For class 0 (bed) frames, we applied slicing and window warping.

— For class 1, 2, and 4 (falling, picking Up, and sitting down) frames, we applied scaling, slicing, and window warping.

— For class 3 (running) frames, we applied window warping.

— For class 5 (standing up) frames, we applied jittering, scaling, slicing, and window warping.

Each minority class needed a distinct number of new frames to reach the $2,512$-frame target, so transformations were chosen partly based on how many synthetic samples each class required, along with memory constraints and the desire to avoid excessive distortion. For example, class 0 (bed) required relatively few additional frames and thus only slicing and window warping were sufficient. By contrast, severely underrepresented classes like standing up (class 5) combined jittering, scaling, slicing, and window warping to generate enough diverse examples. Applying different transformations per class also mitigated the risk of repeatedly creating near-duplicate patterns. Ultimately, any excess frames that exceeded $2,512$ per class were discarded, preserving all original data while balancing each class count.

Additionally, for the training frames of all seven classes, we applied the pattern-mixing-based DA technique, RGW, to generate further synthetic frames. As a result of applying these DA techniques, the total number of training subset frames increased from $8,497$ to $34,884$, ensuring that each activity class now contains around $4,980$ frames, achieving a balanced distribution.

Moreover, a brief empirical assessment indicated that this tailored approach outperformed other tested technique combinations, which failed to provide sufficient improvement to justify the additional memory usage and risk of over-distortion. Hence, the final class-specific distribution of transformations (plus RGW) offered a practical balance between augmentation diversity and computational feasibility.

By applying these DA techniques, we aimed to improve the classification ability of our model on the minority classes' unseen data. Consequently, we anticipated that these techniques would enhance

the reliability of our model's experimental results, thereby improving our analysis. Our main goal is to obtain an effective activity classifier. Therefore, to examine the influence of our proposed approach on model performance, we conducted two experiments with different validation procedures as described in the previous subsection. In Section 5, we will present and discuss the results of these experiments.

To justify the choice of DA techniques implemented in our research, we conducted empirical evaluations to identify methods that best suit the unique characteristics of our datasets. By carefully selecting jittering, scaling, slicing, window warping, and RGW, we aimed to enhance our model's ability to generalize to unseen data and improve recognition accuracy. These techniques were chosen not only because they were readily available but also due to their proven effectiveness in previous studies, ensuring seamless integration into our methodology [Iwana, 2021a ; Alzubaidi, 2021 ; Fernández, 2018].

These DA techniques were selected because they introduced sufficient variability in the training data while retaining the core characteristics of the original activities. This ensures the augmented data remains relevant for accurate activity classification. Visualizations confirmed that the alterations to the CSI data signal frames were minor, avoiding drastic distortions to the underlying data patterns. Such subtle changes are crucial as they reflect real-world variations without compromising the integrity of the data, thus fostering the development of robust DL models.

Moreover, each technique was chosen for its specific technical advantages in handling time-series data. Jittering effectively introduces noise to mitigate time series drift. Scaling adjusts the global magnitude of signals, enhancing variability without altering the sequence's temporal structure. Slicing creates varied frames by cropping parts of the data, while window warping modifies the temporal dimension to introduce additional variability. RGW combines frames within the same class to generate realistic, augmented variations [Iwana, 2021a ; Alzubaidi, 2021 ; Fernández, 2018].

Ultimately, these techniques expanded the decision boundary and mitigated overfitting, allowing the model to classify both the exact signals encountered during training and their slight variations. This careful balance of augmentation ensures our DL models are better equipped to handle real-world scenarios, enhancing their overall performance and reliability.

4 Lightweight CNN Activity Recognizer

4.1 Selecting CNNs for Multidimensional Time-Series Analysis

CNNs, a type of deep ANNs, are particularly well-suited for feature extraction from raw data. Originally designed for image processing, CNNs have been adapted to work with time series data, including multidimensional time series. A multidimensional time series consists of data points organized along multiple dimensions, such as receivers, transmitters, subcarriers, and time steps. Utilizing CNNs for multidimensional time-series classification leverages their powerful feature extraction capabilities without the need to convert time series frames into images. This approach effectively captures the temporal dependencies and patterns within the time series data, which is crucial for accurate classification. Additionally, CNNs can handle high-dimensional data, which is often the case in multidimensional time series, and can learn temporal and spectral features, making them suitable for tasks where the relationships across different dimensions are significant [Yang, 2023a ; Alzubaidi, 2021 ; Goodfellow, 2016 ; Zhao, 2017 ; Yi, 2017 ; Sadouk, 2018 ; Brownlee, 2018].

In Wi-Fi CSI-based HAR approaches, the raw CSI data for human activities are represented as multidimensional time series, with the CSI for each OFDM subcarrier serving as data points across multiple dimensions. The input to the CNN layers consists of the raw CSI data tensor $\mathrm{CSI}^{(\mathrm{act})}$, where act denotes the performed activity, with dimensions $\mathcal{T} \times N_r \times N_t \times N_s$, where $\mathcal{T}$ represents the total number of time steps recorded over the duration of the activity, N_r and N_t denote the number of receivers and trans-

mitters (spatial domain), and N_s refers to the number of subcarriers (frequency domain). This structure represents the CSI variations across time, space, and frequency domains, capturing how human activity influences the signal over multiple dimensions (see FIGURE 1.5 in Chapter 1). This multidimensional tensor is directly fed into the model, which performs deep temporal and spectral analysis to extract meaningful features for accurate activity recognition [Ma, 2019].

The presence of multiple subcarriers adds complexity to the data, as each subcarrier captures different frequency components influenced by human activities. CNNs are particularly advantageous in this context because they can simultaneously process and extract features from these complex, high-dimensional datasets, identifying intricate patterns and dependencies that might be missed by other models. Given that each activity's CSI data signal exhibits a unique pattern, employing CNNs in our research enhances the performance of multi-activity classification tasks [Ma, 2019 ; Goodfellow, 2016 ; Zhao, 2017 ; Yi, 2017 ; Sadouk, 2018 ; Brownlee, 2018].

Additionally, CNNs are capable of leveraging local spatial dependencies through convolutional operations, which is particularly beneficial for multidimensional data. This ability allows CNNs to capture the local interactions between subcarriers, receivers, and transmitters, thereby providing a more comprehensive understanding of the activity-induced variations in the Wi-Fi signals. The hierarchical feature extraction process of CNNs further aids in capturing both low-level and high-level features, making them highly effective for recognizing complex patterns in the CSI data [Ma, 2019 ; Goodfellow, 2016 ; Yi, 2017 ; Brownlee, 2018].

By utilizing CNNs for multidimensional time-series analysis, our approach not only benefits from their powerful feature extraction capabilities but also ensures robust and accurate classification of human activities based on the intricate CSI data patterns.

4.2 Architecture of the Recognizer

4.2.1 Layers and Configurations

The proposed HAR-LightCNN methodology utilizes a CNN to process the CSI activity frames, classifying them into one of seven possible activities. This CNN is specifically designed to be lightweight, minimizing model complexity while preserving classification accuracy, as further detailed in Section 4.2.2. The CNN classifier comprises three convolutional (Conv) blocks, a Fully Connected (FC) block, and an output block. In the following, we describe each component of the architecture:

— **Input Layer:** Batches of CSI amplitude frames constitute the input to this network. Each activity frame represents a contiguous segment of the amplitude component of $\text{CSI}^{(\text{act})}$ recorded for a specific activity act. Each frame is of size (frame_size, num_subcarriers), where:

  — frame_size is the number of consecutive time steps (satisfying frame_size $\leq \mathcal{T}$, the total samples recorded for the activity).

  — num_subcarriers is the count of data subcarriers that we feed into the model. For example, in the 1st dataset there are 52 data subcarriers, whereas in the 2nd dataset each of the three receiver antennas provides 30 data subcarriers, yielding 90 columns of CSI amplitude data when concatenated.

A depth dimension of 1 is added to these 2D frames for convolutional compatibility, yielding frames of shape (frame_size, num_subcarriers, 1). During training, multiple frames are batched into a 4D tensor:

$$(\text{batch_size}, \text{frame_size}, \text{num_subcarriers}, 1),$$

where batch_size remains unspecified at graph-construction time (and thus may appear as `None` in layer summaries). Accordingly, the input shape for the 1st dataset is (batch_size, 200, 52, 1), while

for the 2nd dataset it is (batch_size, 500, 90, 1). By representing both the temporal (time-step) and frequency (subcarrier) dimensions, this input layer enables subsequent convolution operations to learn relevant patterns, often referred to as temporal and spectral features, that are essential for accurate activity recognition.

— **Conv blocks:** The three Conv blocks form the core of the feature extraction process. In the first two blocks, the combination of Conv2D, LeakyReLU, and MaxPooling2D layers identifies and reduces key features, transforming the raw input into a set of informative and manageable representations. In the third block, smaller filters and pooling windows focus on finer-grained details, further refining the feature set. The result is a hierarchical extraction of temporal and spectral patterns that help distinguish between activities with similar CSI signatures (e.g., sitting down vs. standing up or falling vs. lying down).

  — **Conv2D Layer:** This layer performs convolution operations using multiple filters to extract local patterns such as edges, textures, and more complex features from the input data. In our architecture, the first two Conv blocks contain 32 filters of 3×3 in the Conv2D layer, which effectively captures temporal and spectral patterns while maintaining a sufficiently large receptive field. The third Conv block uses 2×2 filters to refine the learned representations, enhancing activity-specific patterns. This choice balances feature extraction and computational efficiency, aligning with the lightweight design of the proposed model.

  — **LeakyReLU Activation:** LeakyReLU, with a slope of 0.3, introduces non-linearity to the model. This helps in capturing more complex patterns in the data and addresses the dying ReLU problem, where neurons might otherwise become inactive and output zero for any input.

  — **MaxPooling2D Layer:** This layer performs downsampling by selecting the maximum value within a 3×3 pooling window for the first two Conv blocks and a 2×2 window for the third block. Pooling reduces the spatial dimensions of the data, lowering computational cost and controlling overfitting.

  — **BatchNormalization Layer:** It normalizes the activations, stabilizing and accelerating the training process by maintaining the mean output close to 0 and the output standard deviation close to 1.

— **FC Block:**

  — **Flatten Layer:** The Flatten layer reshapes the output of the third Conv block into a single column (1D array). For the 1st dataset, for instance, the feature maps after three Conv+Pooling operations measure $(24 \times 6 \times 32)$, resulting in $(\text{None}, 24, 6, 32)$ being converted to $(\text{None}, 4608)$. This shape reflects the progressive downsampling from an original input of $(200, 52)$. By contrast, the 2nd dataset's larger input yields a different final shape, $(\text{None}, 62, 10, 32)$, which flattens to $19,840$ features. In both cases, this step prepares the 2D feature maps for the subsequent Dense layer, which requires 1D input.

  — **Dense Layer:** This layer, with 64 neurons and ReLU activation, is chosen to achieve a compact representation while avoiding excessive parameters. Such a design balances accuracy and computational cost, aligning with the lightweight goals of the proposed CNN model. By reducing thousands of flattened features to 64 neurons, this Dense layer consolidates the most relevant high-level features extracted by the Conv blocks, facilitating effective classification while controlling model complexity.

  — **Dropout Layer:** To avoid overfitting, the Dropout layer randomly sets 30% of the input units to zero during training, encouraging the network to learn more robust features that generalize well to unseen data.

— **Output Block:**

— **Dense Layer with Softmax Activation:** This final Dense layer has seven neurons, corresponding to the seven different indoor daily human activities we aim to classify. The softmax activation function transforms the outputs into a probability distribution. Mathematically, for a given input vector $\mathbf{z} = [z_1, z_2, \ldots, z_7]$ from the Dense layer, the softmax function σ is defined as:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{7} e^{z_j}} \quad \text{for } i = 1, 2, \ldots, 7.$$

Here, z_i represents the input to the i-th neuron in the output layer. The exponential function e^{z_i} ensures that all transformed values are positive, and the normalization by the sum $\sum_{j=1}^{7} e^{z_j}$ ensures that the output values sum to 1, effectively creating a probability distribution across the seven activities. This probability distribution allows the network to output the likelihood of each activity being the correct one. In the context of classification, the predicted activity $\hat{y}$ is determined by identifying the neuron with the highest probability:

$$\hat{y} = \arg\max_i \sigma(z_i)$$

This approach ensures that the activity corresponding to the neuron with the highest softmax output is selected as the network's prediction. The theoretical foundation of softmax can be viewed as a generalization of the logistic regression model to multiple classes, often referred to as multinomial logistic regression. By applying the softmax function in the output layer, our classifier can handle multi-class classification problems, such as distinguishing between multiple human activities, with improved accuracy and reliability.

The overall architecture of our multi-activity classifier is illustrated in FIGURE 3.8 [Alzubaidi, 2021 ; Goodfellow, 2016 ; Brownlee, 2018 ; Sadouk, 2018].

Our model's initial configuration employed the Adam optimizer with a learning rate of 10^{-4}. The Adam optimizer is chosen for its adaptive learning rate capabilities, computational efficiency, and low memory requirements, which can improve convergence speed and performance. The learning rate of 10^{-4} controls the step size at each iteration while moving toward a minimum of the loss function. The loss function was set to Sparse Categorical Cross-Entropy, which is appropriate for multi-class classification problems where the labels are provided as integers. This loss function measures the performance of the model by comparing the predicted probabilities to the actual class labels. We trained the model over 150 epochs with a batch size of 32. The number of epochs denotes how many times the entire training subset passes through the model, while the batch size specifies the number of training samples utilized in one forward/backward pass. Training over 150 epochs ensures that the model has ample opportunity to learn from the data, providing sufficient learning iterations. A batch size of 32 balances computational efficiency with gradient estimation stability and model convergence [Yang, 2023a ; Alzubaidi, 2021 ; Goodfellow, 2016 ; Brownlee, 2018 ; Sadouk, 2018 ; Montavon, 2012]. Various training callbacks were utilized to enhance the training process:

— **EarlyStopping:** This callback terminates training if no improvement in validation loss is observed for 10 epochs. It helps in preventing overfitting by stopping the training early once the model stops learning.

— **ReduceLROnPlateau:** This callback adjusts the learning rate by a factor of 0.5 if no improvement in validation loss is seen for 5 epochs. By reducing the learning rate, the model can make finer adjustments, potentially leading to better performance.

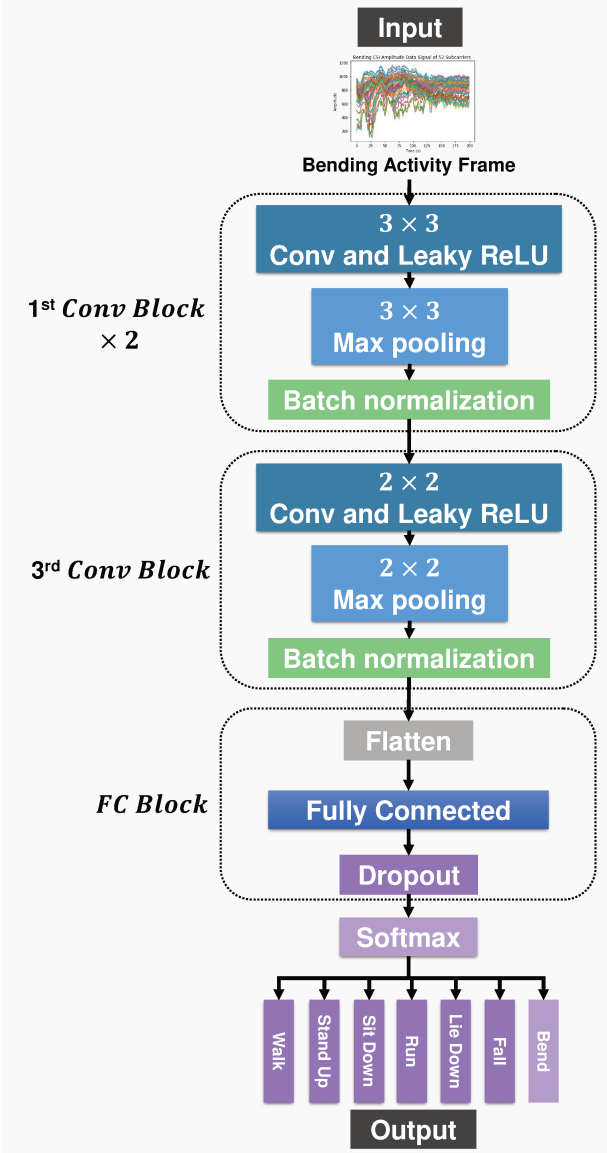


FIGURE 3.8 – Architecture of the lightweight CNN multi-activity classifier: This figure presents the design of our proposed lightweight CNN multi-activity classifier, specifically tailored for multidimensional time-series classification of human activities based on Wi-Fi CSI data. The architecture includes three convolutional blocks, a fully connected block, and an output block, each with specific configurations to optimize the classification performance.

— **ModelCheckpoint:** This callback saves the best-performing model version based on validation accuracy. It ensures that the best model is available for evaluation and deployment, even if the training process undergoes fluctuations in performance.

These hyper-parameters and training strategies are summarized in TABLE 3.1.

4.2.2 Lightweight Design and Efficiency Optimizations

Building on the architecture outlined in the previous section, the proposed CNN model is designed to be lightweight, emphasizing efficiency while maintaining high accuracy. This design specifically addresses the "High Model Complexity" limitation, a common challenge within the "Processing Technique Concerns" category in the domain of Wi-Fi CSI-based HAR. The high model complexity limitation refers

TABLE 3.1 – Summary of Model Hyper-Parameters

Hyper-parameter	**Chosen Value**
Optimizer	Adam
Learning Rate	10^{-4}
Loss Function	Sparse Categorical Cross-Entropy
Metrics	accuracy
Batch Size	32
Number of Epochs	150
EarlyStopping Patience	10 epochs
ReduceLROnPlateau Factor	0.5
ReduceLROnPlateau Patience	5 epochs
ReduceLROnPlateau Min LR	10^{-5}

to the requirement for considerable computational time or resources, which affects practical deployment, particularly in real-time applications and resource-constrained environments.

The lightweight designation of our model stems from its efficient architecture, which substantially reduces the number of parameters compared to existing lightweight CNN architectures, thereby making it computationally less expensive. The architecture balances depth (number of layers) and width (number of units or filters per layer) to operate within the constraints of devices with limited computational resources, ensuring real-time processing capabilities.

To substantiate our claim of lightweight design, we compared the efficiency of our model against established benchmarks in the field: ShuffleNet [Iandola, 2016], SqueezeNet [Zhang, 2018c] (versions 1.0 and 1.1), and MobileNet (versions V2 [Sandler, 2018] and V3 [Howard, 2019]). The comparison focused on several efficiency metrics, including prediction complexity, which quantifies the maximum operations needed to process a single instance by the model as a function of input size, calculated in million floating-point operations (MFLOPs) using the keras_flops library designed to assess the computational demands of Keras-based models; the number of parameters, indicating the total quantity of model parameters; and memory requirements, representing the amount of storage needed to hold the model's parameters, quantified in megabytes (MB).

Each model (with default configurations) was adapted to process inputs of the same dimensions $(200, 52, 1)$ (tailored to the 1st dataset) to ensure a fair comparison. TABLE 3.2 compares the efficiency of the proposed CNN model against other lightweight CNN models cited in the literature.

TABLE 3.2 – Efficiency Benchmarking of the Proposed CNN Against Established Lightweight CNN Architectures

Model	**Prediction Complexity** (MFLOPs)	**N° Parameters**	**Memory Requirements** (MB)
ShuffleNet[†]	74.2	1,035,559	3.9
SqueezeNet v1.0[‡]	224.8	729,607	2.7
SqueezeNet v1.1[‡]	80.8	724,935	2.7
MobileNetV2[§]	149.6	2,266,375	8.6
MobileNetV3[§]	**28.6**	942,871	3.6
Proposed CNN	**53.2**	**309,511**	**1.1**

[†]Implemented by Haikoitoh (GitHub)
[‡]Implemented by cmasch (GitHub)
[§]Implemented by TensorFlow's Keras API

The results demonstrate the lightweight nature of the proposed model, showcasing its potential for deployment in environments with stringent memory and computational constraints. By substantially reducing computational complexity and memory usage, our model effectively addresses the high model complexity limitation, standing as a viable solution for efficient real-time activity recognition.

4.3 Implementation Details

The practical aspects of our methodology, including the software and hardware environments used for constructing and training the CNN model, are crucial for understanding the implementation and efficiency of our approach.

Our classifier is constructed using the Keras layers API [2], a high-level neural networks API written in Python. The software environment, which includes the tools and frameworks used to build and train the model, is critical for the effective implementation of our methodology. Keras is designed to enable rapid experimentation with deep neural networks, offering a variety of built-in layers and simplifying the process of building and training ML models through its user-friendly, modular, and extensible interface.

For the training phase, we conducted numerous experiments that required fitting multiple distinct instances of our model on the preprocessed datasets (1st or 2nd datasets presented in Section 3). The hardware environment, encompassing the physical machines and specifications on which the model is trained, played a key role in allowing multiple rounds of experimentation with minimal overhead. These experiments were performed on Google Cloud servers equipped with NVIDIA GA100 GPUs running at 1095 MHz. Each server had 83.48 GB of RAM and 166.77 GB of storage capacity.

The combination of Keras and high-performance Google Cloud servers with NVIDIA GA100 GPUs was pivotal for the development and training of our lightweight CNN model. Keras' user-friendly API facilitated rapid prototyping and model adjustments, while the powerful GPU hardware significantly reduced training times, allowing us to conduct extensive experiments and fine-tune our model for optimal performance. The high RAM and storage capacities ensured that large datasets could be processed efficiently, contributing to the robustness and reliability of our results.

5 Results and Evaluation

5.1 Benchmarking with Standardized Testing Subsets and Evaluation Metrics

In ML research, particularly for classification tasks, it is essential to employ standardized datasets and evaluation metrics. This practice ensures that research outcomes are aligned and allows for direct and valid performance comparisons. Preparing a specific testing subset for benchmarking is a common and valid approach, providing a controlled and consistent environment for evaluating different models. This method facilitates fair comparisons of performance outcomes. Benchmarking against established work is crucial to highlight advancements or differences in aspects like accuracy and efficiency. Using common benchmarks, which involve standardized datasets and evaluation metrics, ensures that comparisons are valid and results are directly comparable. By curating dedicated testing sets, researchers can maintain a standardized environment for benchmarking various models, enabling fair comparisons under consistent conditions. These practices and considerations are fundamental in ML research for ensuring fair, meaningful, and technically sound comparisons of new models or approaches with existing literature [Weber, 2019 ; Beyer, 2019].

Another limitation within the "Evaluation Process Concerns" category pertains to "Unfair Benchmarking Against Others." This limitation arises when authors compare their approaches with others in the literature without employing a common benchmark, instead relying solely on reported results. This approach compromises the validity of such comparisons due to the absence of standardized datasets and evaluation metrics. Specifically, the authors often conduct comparative analyses by referencing the reported evaluation metrics from other studies, without utilizing a unified benchmarking framework. This

2. Keras: The Python DL library. Available at: https://keras.io

practice aims to highlight potential improvements in performance but inherently acknowledges the limitations posed by the lack of standardized benchmarks. The comparisons made in this manner fail to provide a robust validation of the approach's effectiveness and efficiency, as variations in datasets, experimental conditions, and evaluation criteria can lead to misleading conclusions. In contrast, proper benchmarking involves using standardized datasets and evaluation metrics, which ensures a fair and valid comparison across different methodologies. Hence, the failure to adopt such standardized practices remains a critical concern in the evaluation process of Wi-Fi CSI-based HAR research [Yang, 2023a ; Varga, 2024].

Unified benchmarks exist in various research fields to ensure standardized and comparable evaluations of different methodologies. Here are some examples of unified benchmarks and their respective research fields:

- **Machine Learning and Computer Vision:**
  - **CIFAR-10 and CIFAR-100**[3]: These datasets are commonly used for evaluating image classification algorithms. They consist of labeled images that span 10 and 100 categories, respectively.
  - **MNIST**[4]: A dataset of handwritten digits used for image recognition tasks.
  - **ImageNet**[5]: A large dataset used for image classification and object detection, consisting of millions of images labeled with thousands of categories.
- **Bioinformatics:**
  - **ENCODE**[6] : A project that provides a comprehensive encyclopedia of functional elements in the human genome, used for various genomic research tasks.
  - **TCGA (The Cancer Genome Atlas)**[7]: A dataset used for studying cancer genomics, including mutations, copy number variations, and gene expression.

Unified benchmarks are critical in these fields as they provide a standardized way to measure and compare the performance of different approaches, ensuring that improvements and innovations are assessed fairly and consistently across studies. These benchmarks often come with specific guidelines and protocols to ensure the reliability and validity of the evaluations [Weber, 2019 ; Beyer, 2019].

To address the lack of a unified benchmark in Wi-Fi CSI-based HAR, we have prepared two testing subsets from both datasets used in this study, specifically the 1st dataset from [Fard Moshiri, 2021] and the 2nd dataset from [Yousefi, 2017]. These testing subsets, described in Section 3, serve as unified benchmarks for comparing our proposed approach against state-of-the-art methods. By curating these dedicated testing sets, we aim to provide a reliable and unbiased evaluation environment. This approach ensures that our comparisons with literature approaches are fair and valid, enabling meaningful benchmarking of performance within the domain of Wi-Fi CSI-based HAR. The establishment of these standardized benchmarks is critical for achieving research alignment and facilitating direct and accurate performance comparisons across different DL models.

In this context, we utilize four primary evaluation metrics to assess the effectiveness and efficiency aspects of the ML methodology performance: accuracy, F1-score, training time, and response time. These metrics are integral to the benchmarking process as they provide a comprehensive view of model performance. Accuracy and F1-score measure the correctness of the model's predictions, while training time and response time evaluate the model's efficiency in terms of learning and operational speed. By employing these standardized evaluation metrics alongside common benchmarks, we ensure that our performance comparisons are valid and directly comparable, thereby maintaining the integrity and relevance of our

3. https://www.cs.toronto.edu/~kriz/cifar.html
4. http://yann.lecun.com/exdb/mnist/
5. http://www.image-net.org/
6. https://www.encodeproject.org/
7. https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga

research within the domain of Wi-Fi CSI-based HAR [Yang, 2023a ; Varga, 2024 ; Ng, 2017 ; Montavon, 2012 ; Goodfellow, 2016]. To further elaborate on these metrics:

— **Accuracy** (Effectiveness): This metric measures the percentage of correctly classified activity frames out of the total frames. It is crucial for providing an overall measure of how well the model performs across all categories of activities (bending, falling, lying down, running, sitting down, standing up, and walking). A high accuracy indicates that the model is generally performing well in recognizing these different activities from the Wi-Fi CSI data. It is calculated as:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

— **F1-score** (Effectiveness): The F1-score is the harmonic mean of precision and recall, providing a balance between the two. In the context of HAR, precision measures the number of true positive activity frames (e.g., frames correctly identified as "running") divided by the total number of frames classified as that activity, including those incorrectly classified. Recall measures the number of true positive activity frames divided by the total number of frames that should have been classified as that activity. The F1-score ensures that both false positives (e.g., frames incorrectly classified as "running") and false negatives (e.g., frames that were "running" but classified as another activity) are considered, offering a more nuanced view of the model's performance. This is particularly important in ensuring that critical activities, such as "falling" which might indicate an emergency, are accurately detected without significant misses or false alarms. It is calculated as:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

— **Training Time** (Efficiency): This metric records the average duration required to train the model per epoch, measured in seconds (s). Monitoring training time is crucial for optimizing model performance, making efficient use of computational resources, and ensuring the training process aligns with the operational constraints and objectives of deployment environments.

— **Response Time** (Efficiency): Response time quantifies the latency of the model's predictions for each instance within the testing subset, measured in milliseconds (ms). This is a critical metric for real-time applications, where quick decision-making is essential.

Based on the prepared testing subsets and the evaluation metrics detailed above, we evaluated the performance of our proposed methodology, HAR-LightCNN, with and without DA. This rigorous evaluation process involved comparing our methodology against baseline models from existing literature. By using standardized benchmarks and evaluation metrics, we ensured the credibility and reliability of our evaluation results. We detailed the performance of the classifier in terms of both effectiveness and efficiency, providing insights into the trade-offs between accuracy and computational costs. By examining these metrics, we aim to demonstrate the robustness and practicality of our HAR-LightCNN methodology in real-world healthcare scenarios. This comprehensive evaluation highlights the strengths and potential areas for improvement, showcasing the practical applicability of our approach in addressing the challenges in the Wi-Fi CSI-based HAR research field.

5.2 Effectiveness and Efficiency of HAR-LightCNN

Building upon the standardized benchmarks and evaluation metrics outlined in the previous subsection, this subsection delves into the detailed assessment of the HAR-LightCNN methodology's performance. Our evaluation focuses on both the effectiveness and efficiency of the methodology, providing a

comprehensive analysis through various experimental setups. Initially, we establish a baseline by examining the methodology's performance without DA, followed by an evaluation incorporating DA to enhance training data diversity and volume. These experiments are conducted on two distinct datasets using different training subset validation procedures to assess the stability and robustness of our methodology. The experimental setup for these evaluations is detailed below:

1. **Experiments without DA:**
   - **1st and 2nd Datasets:**
     - **1st Experiment (Holdout Method):**
       - Split the training data into 85% for training and 15% for validation.
       - Repeat this split 4 times, resulting in four different training-validation splits: S_0, S_1, S_2, and S_3.
     - **2nd Experiment (k-fold CV):**
       - Perform 5-fold CV by dividing the training data into 5 subsets.
       - Train on 4 folds and validate on the remaining fold, repeating this process for each fold.
2. **Experiments with DA:**
   - **1st Dataset (Small Dataset):**
     - Apply time-series DA to increase the training dataset's size and diversity.
     - Repeat the two experiments as described above.
   - **2nd Dataset (Imbalanced Dataset):**
     - Balance the training data by generating additional frames for the minority classes, ensuring equal representation of all activities.
     - Repeat the two experiments as described above.

In Subsection 5.2.1, we discuss the evaluation of HAR-LightCNN without DA. By analyzing the experiments, we assessed the impact of different training subset validation procedures on the model's performance, focusing on key metrics such as accuracy, F1-score, training time, and response time. This baseline evaluation provides insights into the classifier's effectiveness and efficiency in handling the datasets without additional data enhancements.

Subsection 5.2.2 explores the evaluation of HAR-LightCNN with DA. We applied various augmentation techniques to increase the training dataset's size and diversity for the 1st dataset and to balance the classes for the 2nd dataset. The results highlight the improvements in classification performance and generalization ability due to DA. Detailed evaluations of the most efficient classifier instances, particularly focusing on critical activities like falling, are discussed to demonstrate the practical benefits of DA in enhancing model robustness.

Finally, Subsection 5.2.3 benchmarks our HAR-LightCNN against state-of-the-art models, including 2D-CNN, LSTM, and ABLSTM. By comparing our proposed model's performance against these established models, we contextualize HAR-LightCNN's effectiveness within the broader research landscape. This benchmarking highlights the improvements brought by our methodology, emphasizing its efficiency and potential for real-time applications, particularly in resource-limited environments.

Our approach ensures a balanced evaluation of both classification performance and computational efficiency. The findings from these experiments underscore the strengths of HAR-LightCNN, offering insights into its efficiency and generalization capabilities, particularly in real-world healthcare scenarios requiring real-time processing. By situating our methodology within the broader landscape of Wi-Fi CSI-based HAR, we demonstrate its practical applicability and potential for enhancing real-world healthcare

services. Given the dependency of response time on factors like processor speed, parallelism, and system architecture, the testing phase for the DL models was conducted on a Google Cloud server equipped with an NVIDIA GA100 @ 1095 MHz GPU.

5.2.1 Evaluation of HAR-LightCNN Without DA

We conducted two experiments on the 1st dataset of [Fard Moshiri, 2021], using different validation procedures for the training subset. Specifically, we used the holdout method and k-fold CV. TABLE 3.3 shows the testing results on the preprocessed unseen testing subset, with the holdout method repeated 4 times, resulting in four different train-validation splits (S_0, S_1, S_2, S_3) for the first experiment and 5-fold CV for the second. The testing accuracies and F1-scores indicate stable and good classification performance, demonstrating that the variation in train-validation splits and validation procedures had minimal impact on the results. This stability highlights the effectiveness of our CNN models, even with limited training data. Additionally, the consistent training and response times across different splits underscore the efficiency of our CNN models.

TABLE 3.3 – Testing Results on [Fard Moshiri, 2021] without DA

Experiment / Metric	1st Experiment				2nd Experiment					Average
	S_0	S_1	S_2	S_3	F_1	F_2	F_3	F_4	F_5	
Accuracy (%)	89.6	91.8	88.7	88.9	90.4	88.0	89.1	89.7	89.3	89.5
F1-score (%)	89.4	91.7	88.6	88.8	90.2	87.8	88.9	89.6	89.3	89.3
Training Time (s)	1.090	1.166	1.038	1.074	1.361	1.080	1.090	1.109	1.107	1.124
Response Time (ms)	0.263	0.246	0.252	0.245	0.261	0.258	0.257	0.257	0.255	0.255

The same experiments were conducted on the 2nd dataset of [Yousefi, 2017], with results shown in TABLE 3.4. As with the 1st dataset, the testing accuracies and F1-scores indicate stable and good performance regardless of the validation method used. The minimal impact of different train-validation splits and validation procedures on the results further emphasizes the effectiveness of our CNN models, even with limited training data. Training and response times remained stable, further demonstrating the efficiency of our CNN models.

TABLE 3.4 – Testing Results on [Yousefi, 2017] without DA

Experiment / Metric	1st Experiment				2nd Experiment					Average
	S_0	S_1	S_2	S_3	F_1	F_2	F_3	F_4	F_5	
Accuracy (%)	83.6	81.0	83.9	75.6	77.7	81.5	81.1	83.2	82.3	81.1
F1-score (%)	83.5	81.2	83.8	76.1	77.8	81.6	81.3	83.1	82.1	81.1
Training Time (s)	8.523	8.342	8.461	8.352	5.086	5.040	5.037	5.030	5.041	6.546
Response Time (ms)	0.774	0.754	0.756	0.758	0.749	0.751	0.743	0.749	0.745	0.753

By systematically evaluating HAR-LightCNN without DA across both datasets and using different validation procedures, we were able to effectively assess the model's baseline performance. This evaluation serves as a foundation for understanding the potential improvements brought by DA, as discussed in the subsequent subsections.

5.2.2 Evaluation of HAR-LightCNN With DA

Building on the baseline performance evaluation without DA, we proceeded to investigate the impact of DA on the HAR-LightCNN methodology. The experiments with DA aimed to increase the training subset size and balance the classes for both the 1st and 2nd datasets.

For the 1st dataset, we applied time-series DA to increase the size and diversity of the training subset. TABLE 3.5 shows the testing results on the unseen testing subset, using the same two validation procedures: the holdout method and k-fold CV. The testing accuracies and F1-scores demonstrate improved classification performance, with an average improvement of 3.4% and 3.6%, respectively. Applying DA enhanced the classifiers' performance and generalization ability. Our proposed approach demonstrated its effectiveness in improving the models' performance.

TABLE 3.5 – Testing Results on [Fard Moshiri, 2021] with DA

Experiment / Metric	1st Experiment				2nd Experiment					Average
	S_0	S_1	S_2	S_3	F_1	F_2	F_3	F_4	F_5	
Accuracy (%)	92.7	91.8	93.6	92.8	92.1	93.7	93.7	92.5	93.7	92.9
F1-score (%)	92.7	91.8	93.6	92.8	92.1	93.7	93.7	92.5	93.7	92.9
Training Time (s)	9.311	10.666	10.052	10.105	9.684	10.000	9.954	10.041	10.000	9.979
Response Time (ms)	0.255	0.254	0.252	0.254	0.250	0.246	0.243	0.244	0.245	0.249

Next, we evaluated the most effective multi-activity classifier (3rd fold classifier with DA) focusing on the seven activities in the 1st dataset, prioritizing the detection of the falling activity due to its critical importance. We used a confusion matrix for evaluation, as shown in FIGURE 3.9, and computed precision, recall, and F1-score for each activity class. TABLE 3.6, the classification report, compares the values of these metrics. For the falling activity, the classifier achieved perfect precision of 100% in identifying only the correct instances of falling and a recall of 97%, indicating a strong ability to identify nearly all correct instances. For the other activities, the precision and recall values were also high. The F1-score, which is the harmonic mean of precision and recall, was our chosen metric for assessing overall classification performance. The results indicate excellent performance in achieving a balance between these two aspects across all activities.

TABLE 3.6 – Classification Report of Most Efficient Classifier on [Fard Moshiri, 2021]

Metric / Class	Precision (%)	Recall (%)	F1-score (%)
Bending	100	98	99
Falling	100	97	98
Lying Down	84	86	85
Running	96	98	97
Sitting Down	87	96	91
Standing Up	93	86	89
Walking	98	96	97

For the 2nd dataset, we balanced the training subset by generating additional frames for the minority classes, ensuring equal representation of all activities. TABLE 3.7 illustrates the testing results on the unseen testing subset, using the same two validation procedures. The testing accuracies and F1-scores indicate improved and stable classification performance, with an average improvement of 3.91% and 3.93%, respectively. The application of DA techniques improved the classification performance and generalization ability of the classifiers. Our proposed approach demonstrated its effectiveness in enhancing the models' performance.

Next, we evaluated the performance of the most effective multi-activity classifier (S_0 split classifier with DA) on the 2nd dataset. Through the classifier's confusion matrix, displayed in FIGURE 3.10, we computed its precision, recall, and F1-score for all activities, detailed in TABLE 3.8. For the falling activity, the classifier achieved a precision of 94% and a strong recall of 91%. For most of the other activities, the precision and recall values were good, and the F1-scores showed a good balance between precision and

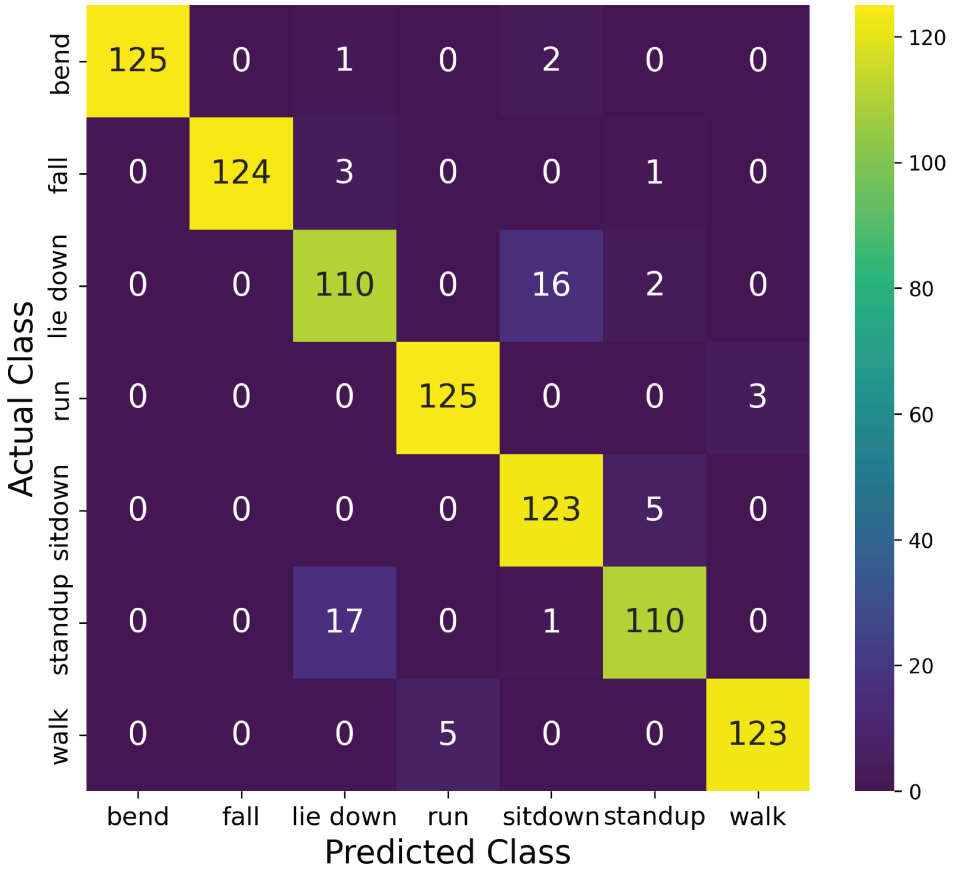


FIGURE 3.9 – Confusion Matrix for Most Efficient Classifier (with DA) on 1st dataset: This figure presents the confusion matrix for the most efficient classifier on the 1st dataset, illustrating the classifier's performance across different activity classes with DA.

TABLE 3.7 – Testing Results on [Yousefi, 2017] with DA

Experiment / Metric	1st Experiment				2nd Experiment					Average
	S_0	S_1	S_2	S_3	F_1	F_2	F_3	F_4	F_5	
Accuracy (%)	87.3	86.4	84.9	86.5	83.0	84.6	83.3	83.3	85.8	85.01
F1-score (%)	87.1	86.4	85.0	86.5	83.0	84.6	83.3	83.5	85.9	85.03
Training Time (s)	21.285	20.245	34.606	34.352	24.374	24.154	24.156	24.7737	24.7767	25.858
Response Time (ms)	0.801	0.773	0.746	0.757	0.746	0.748	0.742	0.749	0.746	0.756

recall. For comparison, TABLE 3.9 presents the F1-scores of the previous best classifier (without DA) for the minority classes (sitting down and standing up) and the walking majority class. Compared to this previous classifier, the F1-scores for minority classes improved by 8% and 2%, respectively, highlighting enhanced classification of the minority classes' unseen data. Likewise, the F1-score for the majority class increased by 11%, indicating an overall improvement in the classifier's generalization ability.

In the conducted experiments, DA significantly enhanced the classifier's performance by increasing the diversity and volume of the training subset. However, this augmentation naturally led to an increase in the average training time. Notably, despite the longer training periods, the response times during the testing phase remained consistently unaffected. This stability is crucial for real-time applications like activity recognition, where prompt decision-making is essential. Therefore, while the augmented training phase requires additional time, which is an acceptable trade-off in scenarios where training occurs offline, the

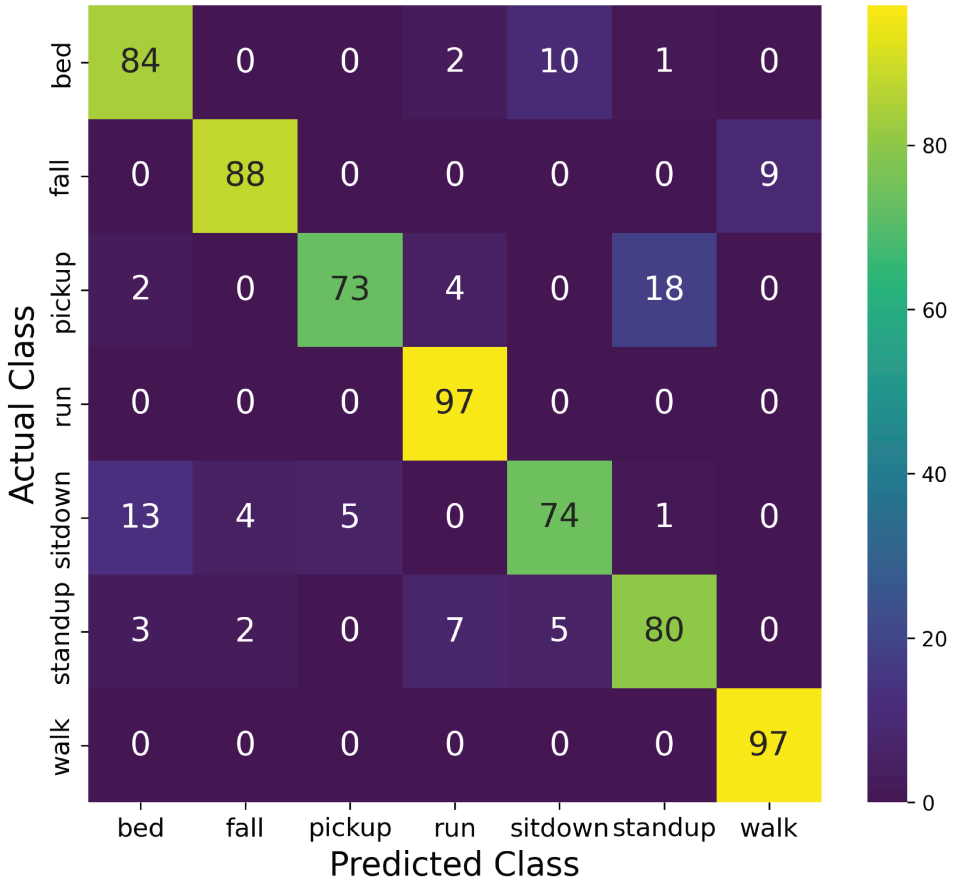


FIGURE 3.10 – Confusion Matrix for Most Efficient Classifier (with DA) on 2nd dataset: This figure shows the confusion matrix for the most efficient classifier on the 2nd dataset, highlighting the classifier's effectiveness in distinguishing between various activity classes with DA.

TABLE 3.8 – Classification Report of Most Efficient Classifier on [Yousefi, 2017]

Metric / Class	Precision (%)	Recall (%)	F1-score (%)
Bed	82	87	84
Falling	94	91	92
Picking Up	94	75	83
Running	88	100	94
Sitting Down	83	76	80
Standing Up	80	82	81
Walking	92	100	96

TABLE 3.9 – F1-score Values of Best Classifier on [Yousefi, 2017] without DA

Class	F1-score (%)	Class Type
Sitting Down	72	Minority
Standing Up	79	Minority
Walking	85	Majority

consistent response time ensures that the operational efficiency during real-time application is maintained. This validates the effectiveness of DA in improving classification accuracy without compromising the

model's responsiveness in practical settings.

5.2.3 Benchmarking Against State-of-the-Art Models

Our study compared different DL models, specifically the 1D-CNN, 2D-CNN, LSTM, and BLSTM, as implemented according to the network configurations and hyper-parameter settings detailed in [Fard Moshiri, 2021], against our proposed CNN model on unseen testing data from their dataset (see TABLE 3.10). Among these models was the 2D-CNN, which served as a benchmark model and was directly compared to the proposed CNN. Our CNN, streamlined to 309,511 parameters, is approximately 21.32 times smaller than the 2D-CNN, which has 6,598,023 parameters. Additionally, it reduces prediction complexity by 2.77 times, requiring about 53.25 MFLOPs compared to 147.75 MFLOPs for the 2D-CNN. It performed significantly better in testing accuracy and F1-score, achieving an 8.7% higher accuracy and F1-score at 93.7%, compared to 85% for the 2D-CNN. This proves that the proposed model is more efficient and generalized, positioning it as a superior choice for embedding in resource-limited devices. Our CNN ensures efficient real-time processing by substantially reducing computational complexity and memory usage, achieving state-of-the-art performance with reduced computational demands. This balance of efficiency and effectiveness establishes the proposed CNN as an ideal candidate for applications requiring minimal computational load and real-time processing capabilities, outperforming not only the 2D-CNN but also the other authors' models, thereby presenting an effective solution for embedding within real-time activity recognition systems.

TABLE 3.10 – Results on [Fard Moshiri, 2021] for HAR-LightCNN vs. State-of-the-art

Activity Recognizer / Metric	1D-CNN	2D-CNN	LSTM	BLSTM	HAR-LightCNN
Accuracy (%)	51.6	85.0	74.5	81.7	**93.7**
F1-score (%)	50.6	84.8	74.3	80.5	**93.7**
Training Time (s)	0.540	1.146	1.046	1.053	9.954
Response Time (ms)	0.337	0.368	0.468	1.04	**0.243**
N° Parameters	630,631	6,598,023	106,279	568,407	**309,511**

We also compared the proposed model to the DL models, specifically the LSTM as implemented according to the network configurations and hyper-parameter settings detailed in [Yousefi, 2017] and the ABLSTM as detailed in [Chen, 2019], on unseen testing data from their dataset [Yousefi, 2017]. The results, shown in TABLE 3.11, indicate that our classifier significantly outperformed the state-of-the-art approaches.

TABLE 3.11 – Results on [Yousefi, 2017] for HAR-LightCNN vs. State-of-the-art

Activity Recognizer / Metric	LSTM [Yousefi, 2017]	ABLSTM [Chen, 2019]	HAR-LightCNN
Accuracy (%)	68.1	75.6	**87.3**
F1-score (%)	67.9	75.7	**87.1**
Training Time (s)	3.140	3.126	21.285
Response Time (ms)	0.916	1.294	**0.801**
N° Parameters	234,207	629,207	1,284,359

6 Conclusions

6.1 Summary of Key Findings

This Subsection summarizes the key findings of this research, highlighting the significant advancements made in the domain of Wi-Fi CSI-based HAR. These findings demonstrate the effectiveness of the proposed methodologies and their implications for real-world applications.

1. **Innovative Lightweight Design**: The research developed HAR-LightCNN, a lightweight CNN model, specifically optimized for real-time activity recognition using Wi-Fi CSI data. The model's design significantly reduces computational complexity and memory usage, making it feasible for deployment in resource-constrained environments such as residential healthcare settings.
2. **Enhanced Generalization and Performance**: By implementing carefully selected DA techniques, the research significantly improved the model's ability to generalize from small and imbalanced datasets. The application of these techniques resulted in substantial improvements in accuracy and F1-scores, demonstrating the model's robustness and effectiveness across different datasets.
3. **Rigorous and Fair Evaluation**: A comprehensive evaluation framework was established, utilizing standardized testing subsets and evaluation metrics. This framework ensured unbiased and fair comparisons with existing models, addressing previous limitations in testing procedures and benchmarking practices. The results from this rigorous evaluation demonstrated the model's superior performance and reliability.
4. **Superior Efficiency and Effectiveness**: The HAR-LightCNN model achieved significant gains in accuracy and efficiency, with notable improvements over existing models. The lightweight design reduced the number of parameters and prediction complexity, proving the model's suitability for real-time applications. These efficiency gains make HAR-LightCNN an ideal solution for practical deployment in environments with limited computational resources.

These key findings collectively advance the field of Wi-Fi CSI-based HAR, providing a robust foundation for future research and practical implementations in healthcare and other domains. This research not only addresses critical limitations in existing methodologies but also sets new standards for efficiency and performance in real-time human activity recognition.

6.2 Benefits for Healthcare Services

The integration of Wi-Fi CSI-based HAR systems, particularly using methodologies like HAR-LightCNN, offers significant benefits to healthcare services. These systems enable continuous, unobtrusive monitoring of patients, enhancing the quality of care and improving patient outcomes. By accurately recognizing daily activities without the need for wearables or cameras, healthcare providers can monitor patient well-being more effectively, leading to timely interventions and personalized care strategies.

6.2.1 Enhancing Patient Monitoring and Care

HAR systems play a crucial role in modern healthcare by providing real-time insights into patients' physical activities. This capability is essential for managing chronic diseases, supporting rehabilitation processes, and ensuring the safety of vulnerable populations such as the elderly [Tan, 2018 ; Ge, 2023]. Traditional monitoring methods, such as wearable sensors and cameras, often face challenges like patient compliance, discomfort, and privacy concerns [Zhang, 2018a ; Gupta, 2024]. The HAR-LightCNN

methodology addresses these issues by leveraging existing Wi-Fi infrastructure to monitor activities non-invasively.

By utilizing Wi-Fi CSI data, HAR-LightCNN can recognize seven common daily activities—bending, falling, lying down, running, sitting down, standing up, and walking—without requiring patients to wear additional devices. This non-intrusive approach respects patient autonomy and privacy, essential considerations in healthcare delivery [Duan, 2023]. Continuous monitoring allows healthcare providers to detect anomalies or changes in activity patterns promptly, facilitating early interventions that can prevent complications or deterioration of health conditions.

For example, sudden decreases in activity levels may indicate the onset of depression or chronic pain [Tan, 2018]. By identifying such changes through HAR-LightCNN, healthcare services can adjust care plans accordingly, providing more effective and personalized support to patients.

6.2.2 Advantages Over Traditional Monitoring Methods

The HAR-LightCNN methodology offers several advantages over traditional monitoring systems:

— **Non-Intrusive Monitoring**: Unlike wearable sensors, HAR-LightCNN does not require patients to wear any devices, eliminating issues related to discomfort and compliance [Ge, 2023]. This feature is particularly beneficial for elderly patients or those with disabilities who may find wearables cumbersome.

— **Privacy Preservation**: Camera-based systems raise privacy concerns and are limited by lighting conditions and line-of-sight constraints [Zhang, 2018a ; Gupta, 2024]. HAR-LightCNN utilizes Wi-Fi signals, which are less invasive and can operate effectively without direct visual contact, preserving patient privacy.

— **Cost-Effectiveness and Accessibility**: By leveraging existing Wi-Fi infrastructure, HAR-LightCNN reduces the need for additional hardware investments. Wi-Fi routers are affordable and widely available, making the system accessible for widespread adoption in various healthcare settings [Tan, 2018 ; Wang, 2018].

— **Real-Time Activity Recognition**: The lightweight design of HAR-LightCNN enables real-time processing, essential for timely interventions in critical situations such as falls or sudden health deteriorations. Its efficiency makes it suitable for deployment in resource-constrained environments like residential homes or small clinics.

— **Scalability and Adaptability**: HAR-LightCNN's architecture allows for easy adaptation to different environments and patient populations. Its ability to generalize from small and imbalanced datasets enhances its applicability across diverse healthcare scenarios.

6.2.3 Scenario: Application in Elderly Care Facilities

Consider an elderly care facility equipped with the HAR-LightCNN system for monitoring residents' daily activities. The system continuously analyzes Wi-Fi CSI data to recognize activities such as walking, sitting down, or lying down.

Scenario: An elderly resident, Mary, has a history of mobility issues and is at risk of falls. The HAR-LightCNN system detects that Mary has been lying down for an unusually extended period during daytime hours, deviating from her normal activity pattern. The system alerts caregivers to check on her well-being.

Upon inspection, caregivers find that Mary is experiencing dizziness and weakness. Early detection allows for prompt medical attention, preventing potential complications. The system logs this event,

contributing to Mary's health records and enabling healthcare providers to adjust her care plan accordingly.

This scenario illustrates how HAR-LightCNN enhances patient monitoring, enabling proactive care and improving patient outcomes. By providing continuous, non-intrusive monitoring, healthcare services can better support the elderly, reducing risks associated with inactivity or unnoticed health issues.

6.2.4 Advancing Healthcare Delivery

The implementation of HAR-LightCNN in healthcare services represents a significant advancement in patient monitoring technology. By addressing the limitations of traditional methods, this methodology contributes to more efficient and effective healthcare delivery.

The ability to monitor patients unobtrusively aligns with modern healthcare goals of patient-centered care and autonomy [Duan, 2023]. HAR-LightCNN's efficiency and accuracy support healthcare providers in delivering personalized care plans, tailored to individual patient needs and activity patterns.

Moreover, the system's adaptability ensures that it can be integrated into various healthcare environments, from residential homes to hospitals. Its scalability allows for widespread adoption, potentially transforming standard practices in patient monitoring and activity recognition.

6.2.5 Summary of Benefits

By leveraging Wi-Fi CSI data and the advanced capabilities of HAR-LightCNN, healthcare providers can enhance patient monitoring without imposing additional burdens on patients. The methodology's non-intrusive nature, coupled with its efficiency and adaptability, makes it a valuable tool for improving patient outcomes and optimizing healthcare resources.

The contributions outlined in Chapter 3 demonstrate the potential of integrating advanced HAR systems into healthcare services. HAR-LightCNN offers a practical solution to longstanding challenges in patient monitoring, paving the way for more responsive and personalized healthcare delivery.

6.3 Future Directions

Building upon the findings and contributions of this chapter, several avenues for future research emerge that can enhance the effectiveness and applicability of Wi-Fi CSI-based HAR systems, particularly the HAR-LightCNN methodology developed herein.

Firstly, one limitation of the current study is the reliance on two specific datasets, which, while sufficient for initial validation, may not capture the full variability present in real-world scenarios. Future work could involve collecting and incorporating larger and more diverse datasets, encompassing a broader range of activities, environmental settings, and participant demographics. Expanding the dataset diversity and scale would enable the model to generalize better across different contexts and improve its robustness.

Secondly, the HAR-LightCNN model is designed for single-user activity recognition. An important extension would be to adapt the methodology to handle multi-user scenarios, where multiple individuals perform activities simultaneously within the same environment. Developing models capable of distinguishing and recognizing activities from multiple users would significantly enhance the applicability of HAR systems in more complex real-world settings, such as hospitals or assisted living facilities.

Moreover, while the current approach utilizes several time-series DA techniques to address data scarcity and imbalance, future research could explore more sophisticated augmentation methods. Techniques such as Generative Adversarial Networks (GANs) could be employed to generate synthetic CSI data that closely mimic real-world signals. This would further enhance the diversity of training data and potentially improve model performance.

In addition, the HAR-LightCNN model primarily focuses on temporal and spectral feature extraction at the CSI frame level. However, incorporating deeper sequence modeling techniques, such as Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks, could improve the recognition of activities with stronger long-range temporal characteristics. Hybrid models that combine CNNs with sequence modeling would capture both subcarrier-level and extended temporal features, potentially leading to improved accuracy.

Another area for future exploration is environmental adaptability and robustness. Future studies could investigate the impact of environmental changes on model performance, such as variations in room layouts, furniture placement, or the presence of obstacles. Developing methods to adapt the model to new environments without extensive retraining would enhance its practicality. Techniques like domain adaptation or transfer learning could be explored to achieve environmental adaptability.

Furthermore, while HAR-LightCNN is designed to be lightweight, further optimization could enable deployment on edge devices with even more limited computational resources, such as microcontrollers or smartphones. Investigating model compression techniques, quantization, or efficient neural network architectures could facilitate real-time implementation in resource-constrained settings, expanding the system's accessibility and scalability.

Recognizing that individuals may have unique activity patterns, future research could focus on developing personalized HAR models. Incorporating user-specific data during training or employing adaptive algorithms that learn from individual users over time could improve recognition accuracy and user satisfaction. This personalization would be particularly beneficial in healthcare settings, where patient-specific monitoring is crucial.

Finally, investigating alternative deep learning architectures, such as transformer-based models or graph neural networks, could reveal new opportunities for improving activity recognition. These architectures may offer advantages in capturing complex patterns within CSI data and could be tailored to address specific challenges identified in this study.

By pursuing these future directions, the research community can continue to advance the field of Wi-Fi CSI-based HAR, enhancing the capabilities of methodologies like HAR-LightCNN. These efforts will contribute to more effective, personalized, and accessible healthcare services, ultimately improving patient care and outcomes.

Chapter 4

CSI-Based Human Fall Detection through a Comprehensive Four-Stage DL Approach for Emergency Response Service

Objectives

This chapter presents *HFD-Adv*, a comprehensive four-stage DL methodology designed to address key data-, processing-, and post-processing-related limitations in CSI-based fall detection. By tackling challenges such as the absence of empty data, hyper-parameter sensitivity, and the lack of post-processing mechanisms, the methodology delivers a robust, real-time solution for emergency response in healthcare scenarios.

Chapter Outline

1 Introduction . . . **104**
- 1.1 Public Health Impact of Falling Accidents . . . 104
- 1.2 Significance of Rapid Intervention in Healthcare . . . 105
- 1.3 Defining the Problem Space and Its Challenges . . . 107
- 1.4 Contributions . . . 107
- 1.5 Organization of the Chapter . . . 109

2 HFD-Adv Methodology Overview . . . **109**
- 2.1 Introduction to the Methodology . . . 109
- 2.2 Objectives and Design Principles . . . 110

3 Data Preparation and Signal Preprocessing . . . **114**
- 3.1 Collection and Structuring of Activity CSI Data . . . 115
- 3.2 System Calibration with Empty Activity Data . . . 116
- 3.3 User-Specific Train-Test Split . . . 116
- 3.4 Dimensionality Reduction with PCA . . . 116
- 3.5 Handling Incomplete Frames with Threshold-Based Padding Strategy . . . 117

4 ActiNet Flexible CNN Classifier . . . **120**
- 4.1 Architecture for Fall Detection . . . 121
- 4.2 Weighted Loss Function Optimization using Genetic Algorithm . . . 121
- 4.3 Learning Rate Optimization with LRFinder . . . 124
- 4.4 Hyperparameter Optimization and Training Strategies . . . 126

5 Transfer Learning for Model Adaptability . . . **127**
- 5.1 Role of Transfer Learning in Enhancing Model Adaptability . . . 127

5.2 Transfer Learning Strategy and Implementation for Model Fine-Tuning . . . 127
6 Advanced Decision Mechanism for Classifier Outputs Reassessment . . . **128**
6.1 Introduction to the Advanced Decision Mechanism . . . 128
6.2 Age-Sensitive Threshold Adjustment . . . 130
7 Results and Evaluation . . . **131**
7.1 Evaluation Metrics and Benchmarking . . . 131
7.2 Effectiveness, Efficiency, and Versatility of HFD-Adv . . . 132
7.3 Benchmarking Against State-of-the-Art Models . . . 133
8 Conclusions . . . **134**
8.1 Summary of Key Findings . . . 134
8.2 Benefits for Healthcare Services . . . 135
8.3 Future Directions . . . 137
9 Appendix: Additional Validation on a Different Left-out User (user_2) . . . **138**

1 Introduction

1.1 Public Health Impact of Falling Accidents

Falling accidents are a significant public health concern recognized globally for their severe consequences. The World Health Organization (WHO) considers falls the second leading cause of unintentional injury deaths worldwide, with an estimated 684,000 individuals dying from falls annually [World Health Organization, 2021]. This statistic underscores the urgent need to address falling accidents as a critical public health issue.

Demographically, older adults and individuals with certain medical conditions are the most affected. Adults over 60 years suffer the greatest number of fatal falls [World Health Organization, 2021]. In the United States alone, about 3 million older adults are treated in Emergency Departments (ED) for fall injuries each year [Centers for Disease Control and Prevention, 2023b]. These statistics highlight the vulnerability of older adults to falling accidents and their resulting health implications.

The prevalence and incidence of falls, particularly among vulnerable populations, are alarming. Approximately 37.3 million falls globally are severe enough to require medical attention annually [World Health Organization, 2021]. In Australia, there has been an average annual increase in deaths due to falls between 2011–12 and 2020–21 [Australian Institute of Health and Welfare AIHW, 2023]. In the United States, falls are the leading cause of injury-related death among adults aged 65 and older, with the fall death rate increasing by 41% from 55.3 per 100,000 older adults in 2012 to 78.0 per 100,000 older adults in 2021 [Centers for Disease Control and Prevention, 2023c]. Furthermore, every 11 seconds, an older person is treated for a fall in the emergency room [Centers for Disease Control and Prevention, 2021].

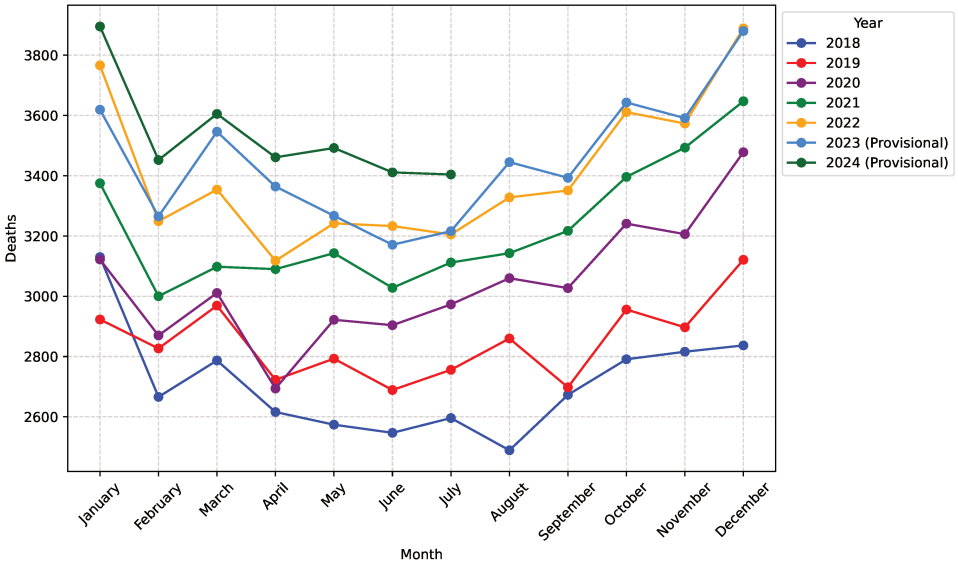


FIGURE 4.1 – Monthly Trends in Unintentional Older Adults Falls (Age 65 and Older) in the U.S.: This figure presents the monthly trends in deaths due to unintentional falls among older adults (age 65 and older) from 2018 to 2024. Data for 2023 and 2024 are provisional. Source: Data from CDC's Web-based Injury Statistics Query and Reporting System (WISQARS™) [Centers for Disease Control and Prevention, 2023c].

Recent trends indicate that the rates of falls are increasing due to aging populations. This is evident from the aforementioned statistics in Australia and the United States. FIGURE 4.1 illustrates the monthly trends in deaths due to unintentional falls among older adults from 2018 to 2024. Common risk factors include lower body weakness, vitamin D deficiency, balance difficulties, certain medications, vision problems,

poor footwear, and home hazards like uneven steps or clutter [Centers for Disease Control and Prevention, 2023a]. Additional factors such as age, gender differences in fall outcomes, occupational hazards, substance use, socioeconomic factors, and underlying medical conditions also play significant roles [World Health Organization, 2021]. For instance, conditions like diabetes, heart disease, and vision or hearing loss can affect balance, increasing the risk of falls [National Institute on Aging, 2022]. Fear of falling can lead to decreased physical activity, further raising the risk of falls, creating a vicious cycle [National Institute on Aging, 2022].

Various environments contribute to falling accidents, including homes, workplaces, and public spaces [World Health Organization, 2021]. Certain occupations, such as healthcare support, building cleaning and maintenance, transportation and material moving, and construction and extraction, are at higher risk for fall-related injuries [Occupational Safety, 2022].

Falls can result in a range of injuries, from minor bruises to severe fractures, head injuries, and even fatalities. One out of every five falls causes a serious injury, such as broken bones or a head injury [Centers for Disease Control and Prevention, 2023b]. Women tend to fall more often than men and account for three-quarters of all hip fractures, which are often caused by falls [Centers for Disease Control and Prevention, 2023b].

The health impacts of falls are both short-term and long-term, encompassing physical, psychological, and functional consequences. Falls among adults aged 65 and older are particularly burdensome, not only due to the immediate health impacts but also due to the significant financial strain on the healthcare system [Centers for Disease Control and Prevention, 2023c]. Complications from fall-related injuries, such as decreased mobility, further increase the risk of additional falls [National Institute on Aging, 2022].

The economic burden of falling accidents on healthcare systems is substantial. In the United States, around $50 billion is spent annually on medical costs related to older adult falls, with fatal injuries totaling $754 million and the remainder attributed to non-fatal fall injuries [Centers for Disease Control and Prevention, 2023c]. This includes costs associated with emergency response, hospitalization, rehabilitation, and long-term care. Occupational fall incidents alone cost approximately $70 billion annually in the United States [Occupational Safety, 2022]. Indirect costs, such as lost productivity and the financial strain on families, further exacerbate the economic impact of falls [Occupational Safety, 2022].

Falls significantly affect the quality of life for individuals and their families. Issues such as loss of independence, fear of falling, social isolation, and reduced participation in daily activities are common among fall victims. Falls are the leading cause of injury-related death among adults aged 65 and older, highlighting their severe impact on quality of life [Centers for Disease Control and Prevention, 2023c].

In summary, falling accidents represent a critical public health issue with significant health, economic, and quality-of-life implications. The rising prevalence of falls, particularly among older adults, calls for urgent intervention. The importance of prompt intervention strategies cannot be overstated, as they play a crucial role in mitigating the short and long-term consequences of falling incidents. Rapid intervention is critical in reducing the severity and complications associated with falls, thereby improving patient outcomes. Institutions such as the National Health Service (NHS) England highlight the significant impact of these strategies in enhancing the patient experience and streamlining emergency response processes [NHS England, 2023b].

1.2 Significance of Rapid Intervention in Healthcare

Prompt intervention following a fall incident is crucial in mitigating both immediate and long-term health consequences, particularly among vulnerable populations such as older adults. Rapid response can significantly reduce the severity of injuries, prevent complications, and improve overall patient outcomes [National Institute on Aging, 2022; NHS England, 2023a]. The National Institute on Aging em-

phasizes that immediate medical attention after a fall can be the difference between a minor injury and a life-threatening condition [National Institute on Aging, 2022].

Falls are the leading cause of injury-related deaths among adults aged 65 and older [Centers for Disease Control and Prevention, 2023c]. Without rapid intervention, individuals who fall may experience prolonged periods of immobility, leading to complications such as pressure ulcers, dehydration, hypothermia, and increased mortality risk [National Institute on Aging, 2022]. Moreover, delays in assistance can exacerbate psychological impacts, including fear of falling again, which can result in decreased physical activity and social isolation [National Institute on Aging, 2022].

From a healthcare system perspective, swift intervention reduces the burden on emergency departments and minimizes unnecessary hospital admissions [NHS England, 2023a]. According to NHS England, enhancing community-based rapid response services is essential for improving patient experiences and outcomes [NHS England, 2023a]. Implementing effective rapid intervention strategies can lead to:

— **Reduced Severity of Injuries**: Immediate medical attention can prevent minor injuries from escalating into severe conditions requiring extensive treatment or hospitalization [National Institute on Aging, 2022].

— **Improved Recovery Times**: Quick response facilitates faster rehabilitation, enabling individuals to return to their daily activities sooner [National Institute on Aging, 2022].

— **Decreased Healthcare Costs**: By preventing complications and hospital admissions, rapid intervention contributes to significant cost savings for healthcare systems [Centers for Disease Control and Prevention, 2023c].

The integration of rapid intervention mechanisms, such as emergency response healthcare service dedicated to fall incidents, plays a vital role in addressing this need [NHS England, 2023a]. These services are designed to provide immediate assistance and urgent care, ensuring that individuals receive timely medical attention from healthcare professionals or caregivers. This is especially crucial for those who may not be able to seek help independently due to age or pre-existing health conditions [National Institute on Aging, 2022].

Technological advancements have paved the way for innovative solutions to facilitate rapid intervention. Fall detection systems are at the forefront of this technological revolution. By accurately and efficiently detecting human falls, fall detection systems enhance the quality of emergency response service. They significantly reduce response times by automatically alerting medical teams or caregivers when a fall is detected, decreasing the potential severity of injuries through prompt medical intervention [NHS England, 2023a].

These technologies are particularly advantageous for vulnerable populations in high-risk settings, providing an additional layer of protection and reassurance that assistance is promptly available in case of a fall [National Institute on Aging, 2022]. The inclusion of fall detection systems in emergency response strategies enables healthcare providers to improve both the efficiency and effectiveness of their care, leading to better patient outcomes and experiences [NHS England, 2023a].

Recognizing the critical need for comprehensive fall detection systems, there has been increased research into various activity measurement technologies, including cameras [Alam, 2022], wearable sensors [Singh, 2020], and Wi-Fi self-powered sensors [He, 2020]. In our research, we focus on leveraging Wi-Fi Channel State Information (CSI) to develop non-intrusive, privacy-preserving fall detection systems. CSI provides a detailed measure of human movements without the need for wearable devices, making it an ideal choice for continuous monitoring in residential settings [Ma, 2019].

In summary, rapid intervention is a significant component in the healthcare response to falling accidents. It not only mitigates the immediate and long-term health impacts but also enhances the overall quality of life for individuals at risk of falls. By investing in rapid intervention strategies and technolo-

gies, healthcare systems can address the urgent need to reduce the prevalence and severity of fall-related injuries, ultimately contributing to improved public health outcomes.

1.3 Defining the Problem Space and Its Challenges

The critical need for rapid intervention in fall incidents, underscores the importance of reliable and timely fall detection. While existing technologies such as camera-based systems and wearable sensors have made strides in *Human Fall Detection* (HFD), they face inherent drawbacks in healthcare environments [Alam, 2022; Singh, 2020]. These limitations impede the goal of providing non-intrusive, round-the-clock monitoring for individuals at high risk of falls.

Wi-Fi CSI has emerged as a privacy-preserving and cost-effective alternative, leveraging ordinary wireless infrastructure to capture subtle human movements. Yet, deploying CSI-based HFD for emergency response presents its own set of challenges. First, accurate fall detection demands near-real-time processing of noisy, high-dimensional CSI data, which can be influenced by environmental factors such as multipath reflections and interference. Unlike general activity recognition, the severity of misclassifications in fall detection is amplified; false negatives (missed falls) can lead to serious health consequences if intervention is delayed.

Second, data quality and diversity remain problematic. In many real-world scenarios, CSI datasets for fall detection are small, imbalanced, or missing critical frames. Discarding incomplete data further reduces training samples and undermines model generalization, while limited user or environmental diversity hinders performance across varied populations. Third, robust model training is non-trivial, as overfitting, hyper-parameter sensitivity, and prolonged training times can diminish the system's reliability and usability. These problems are compounded by the fact that fall detection often requires heightened sensitivity, increasing the risk of false alarms if the model is not properly tuned.

Finally, post-processing and personalization are rarely addressed in existing CSI-based solutions. Falls can manifest differently depending on a person's age, physical condition, or environment, suggesting that a static threshold or a one-size-fits-all decision mechanism is inadequate. Without an adaptive post-processing strategy, even a strong classifier may struggle to maintain low false negatives across diverse user groups.

In summary, the problem space centers on developing a non-intrusive, accurate, and adaptable CSI-based HFD methodology that supports rapid emergency intervention. Key challenges include noisy and incomplete CSI data, limited data diversity, overfitting and hyper-parameter sensitivity in model training, and the absence of advanced post-processing mechanisms tailored to individuals.

1.4 Contributions

This chapter outlines the significant contributions of this research in the domain of Wi-Fi CSI-based Human Activity Recognition (HAR). In this study, we present a novel methodology, the Advanced Human Fall Detection (HFD-Adv), which significantly enhances the state-of-the-art in this domain. Our contributions are categorized into four primary areas:

1. **Overcoming Existing Limitations**:
   - **Data-Related Concerns**:
     - **Absence of Empty Data**: We address the challenge of the absence of empty data signals for environmental calibration by incorporating 'empty' activity data, allowing the model to better distinguish between human activities and environmental noise.
     - **Missing Training Data**: Our threshold-based padding strategy for handling incomplete frames ensures data integrity, preserving valuable information that would otherwise be

discarded.

- **Limited Data Diversity**: By employing dimensionality reduction techniques like PCA and implementing a refined transfer learning strategy, we enhance the model's generalizability across various demographics and environments.

- **Processing Technique Concerns**:
  - **Overfitting Training Data**: We mitigate overfitting through a weighted loss function optimized with Genetic Algorithms (GAs) and by employing transfer learning, ensuring robust performance on unseen data.
  - **Hyper-parameter Sensitivity**: Our use of GAs for loss function optimization and the LRFinder mechanism for learning rate tuning reduces dependency on manual hyper-parameter selection, enhancing model reliability.
  - **Long Training Period**: The LRFinder mechanism accelerates model convergence, reducing the training period and facilitating quicker iterations.
- **Post-processing Mechanism Concerns**:
  - **Absence of Post-processing Mechanism and Static Post-processing Mechanism**: The introduction of an Advanced Decision Mechanism (ADM) reassesses classifier outputs and includes demographic-sensitive adjustments, improving detection accuracy and system reliability.

2. **Introducing Novel Methodologies or Approaches**:
   - **Four-Stage Methodology**: We propose a comprehensive four-stage approach (preprocessing, processing, enhanced-processing, and post-processing) that systematically addresses the specific challenges in Wi-Fi CSI-based HAR.
   - **ActiNet Flexible CNN Classifier**: The development of ActiNet, a versatile Convolutional Neural Network (CNN) tailored for varied HAR tasks, marks a significant advancement. ActiNet's architecture is designed for adaptability and robustness, accommodating both binary and multi-class classification tasks.
   - **Innovative Optimization Techniques**: Implementation of weighted loss function optimization using GAs and learning rate optimization with the LRFinder mechanism demonstrates innovative approaches for enhancing model training and performance.
3. **Enhancing Performance**:
   - **High Fall Detection Rate (FDR)**: Our extensive evaluations demonstrate that the HFD-Adv methodology consistently achieves a high FDR across various experimental conditions, significantly outperforming existing methods and ensuring reliable fall detection in real-world scenarios.
4. **Application Innovations**:
   - **User-Specific Adaptability**: The refined transfer learning approach enables the model to adapt to new users and environments, even with limited initial data, ensuring precise fall detection across diverse conditions.
   - **Advanced Decision Mechanism**: The ADM not only enhances detection accuracy by reassessing classifier outputs but also includes age-sensitive threshold adjustments, tailoring the system's sensitivity based on demographic-specific fall likelihood.

These contributions collectively elevate the capabilities of non-intrusive, privacy-preserving fall detection systems, providing a robust, efficient, and adaptable solution for emergency response service. The HFD-Adv methodology represents a significant leap forward in the deployment of advanced HAR technologies, with extensive evaluations confirming its superiority over state-of-the-art methodologies.

1.5 Organization of the Chapter

The remainder of this fourth chapter (second chapter of the contributions part II) is organized as follows:

— Section 2 sets the stage for our approach—*HFD-Adv* methodology—by summarizing its motivation and four-stage structure (*Preprocessing*, *Processing*, *Enhanced-processing*, and *Post-processing*), which underpin the design and objectives of our fall detection solution.

— Section 3 delves into the initial preprocessing stage. It covers dataset collection and structuring, environment calibration using 'empty' activity data, dimensionality reduction with PCA, and the handling of incomplete frames. These steps ensure robust, well-organized input data for the fall detection classifier.

— Section 4 focuses on the processing stage, detailing the core classifier architecture (*ActiNet*). It explains how overfitting is mitigated, why a weighted loss function is used, and how hyperparameter optimization (including a Genetic Algorithm and LRFinder) fosters efficient and accurate fall detection.

— Section 5 highlights the enhanced-processing stage. It describes how the system adapts to new users via refined transfer learning, ensuring robustness and generalization even when data from unfamiliar users or environments is introduced.

— Section 6 details the post-processing stage. It presents the ADM, an approach to reassessing classifier outputs and introducing demographic-sensitive adjustments, further enhancing detection accuracy for vulnerable populations like the elderly.

— Section 7 provides a comprehensive performance analysis of the proposed *HFD-Adv* methodology. It examines both effectiveness and efficiency metrics, compares results against state-of-the-art models, and demonstrates the impact of each methodological stage.

— Section 8 concludes the chapter by summarizing key findings, outlining benefits for healthcare services, and proposing future research directions.

— Appendix 9 presents additional validation on a different unseen user, reinforcing *HFD-Adv*'s scalability across different users.

By the end of this chapter, we present a specialized and robust Wi-Fi CSI-based HFD solution for emergency response service. In the subsequent Chapter 5 (third chapter of the contributions part), we shift our focus to *user identification task*, thereby broadening the scope of CSI-based HAR to include personalized healthcare services.

2 HFD-Adv Methodology Overview

2.1 Introduction to the Methodology

The integration of fall detection systems into emergency response service frameworks is pivotal for enhancing the quality of care provided to individuals at risk of falls. Accurate identification of fall incidents allows for rapid intervention, significantly reducing the risk of severe injury through prompt medical response. This is especially crucial for high-risk groups, such as the elderly and individuals with chronic conditions, providing them with an added layer of security and assurance [National Institute on Aging, 2022 ; Ge, 2023].

The Advanced Human Fall Detection (HFD-Adv) methodology leverages Wi-Fi CSI, a sophisticated metric derived from Wi-Fi signals, to precisely capture and analyze human movements. CSI's sensitivity to subtle changes in the environment, caused by human activity, makes it a superior choice compared to

other metrics like the Received Signal Strength Indicator (RSSI), which is less precise due to environmental noise and its coarse-grained nature [Ge, 2023; Ma, 2019].

By integrating machine learning (ML) and deep learning (DL) techniques with CSI data, HFD-Adv is capable of distinguishing falls from other human activities with high accuracy. The use of ML and DL allows the system to learn from intricate patterns within CSI data, providing a robust framework for binary classification, specifically identifying fall incidents among various activities. This approach not only enhances the precision of fall detection but also ensures the system's adaptability and continuous improvement over time [Ma, 2019].

The HFD-Adv methodology introduces a comprehensive four-stage approach to fall detection systems, significantly enhancing their effectiveness, efficiency, and versatility. This holistic methodology includes:

1. **Preprocessing Stage**: Preparation of the activity CSI dataset to ensure system reliability and versatility across various settings.
2. **Processing Stage**: Development of an advanced flexible DL model for precise and efficient fall detection.
3. **Enhanced-processing Stage**: Extension of the initial processing stage to improve the model's transferability and adaptability, particularly for new and unseen users.
4. **Post-processing Stage**: Refinement of outputs to ensure precise, user-specific results, tailored to specific needs, such as those of the elderly.

Notable contributions of the HFD-Adv methodology include the development of ActiNet, a flexible CNN designed for varied HAR tasks. Additionally, the methodology incorporates innovative strategies such as weighted loss function optimization using Genetic Algorithms (GAs), refined transfer learning for enhanced model adaptability, and an Advanced Decision Mechanism (ADM) that adjusts sensitivity based on demographic factors like age.

The holistic design of HFD-Adv addresses several critical limitations identified in existing literature in the domain of Wi-Fi CSI-based HAR, such as the absence of empty data for environment-specific calibration, the issue of missing and incomplete training data, and the limited diversity in CSI data affecting model generalizability. Additionally, it tackles overfitting and hyper-parameter sensitivity in training models and mitigates long training periods. The methodology also incorporates an advanced post-processing mechanism to reassess classifier outputs and tailor sensitivity based on demographic factors. By addressing these challenges, HFD-Adv sets a new standard for the accuracy, adaptability, and applicability of HAR systems in diverse real-world conditions. This marks a significant advancement in the deployment of non-intrusive, privacy-preserving fall detection systems within emergency response service frameworks, thereby improving patient outcomes and experiences through technological innovation.

2.2 Objectives and Design Principles

The HFD-Adv methodology is designed to develop an effective, efficient, and versatile DL solution for Wi-Fi CSI-based HAR, with a specific focus on fall detection. This integrated approach sets a robust foundation for future advancements in the field. The methodology aims to address several key limitations—identified in Section 4 and prioritised in Section 7—within three categories in the domain of Wi-Fi CSI-based HAR:

1. **Data-Related Concerns**:
   - **Missing Training Data**: A significant limitation in CSI-based HAR is the issue of missing training data, specifically incomplete training frames. Incomplete frames lack the full set of data points necessary for consistent frame size, crucial for effective DL model training. A common solution in the literature is to discard these incomplete frames, leading to the loss

of valuable data and exacerbating the problem. The HFD-Adv methodology addresses this by implementing a threshold-based padding strategy in the Preprocessing Stage, which retains and completes these incomplete frames rather than discarding them. This approach ensures that valuable data is preserved and utilized, enhancing overall data integrity and consistency. By addressing the missing training data issue, HFD-Adv improves the model's training process, resulting in better performance and reliability in fall detection.

— **Absence of Empty Data**: A significant limitation in Wi-Fi CSI-based HAR is the absence of empty data signals, crucial for environment-specific calibration. This limitation hampers the model's ability to accurately distinguish between human activities and environmental noise. The HFD-Adv methodology addresses this by incorporating 'empty' activity data from an additional dataset, training the model to recognize periods of inactivity. This step, part of the Data Purity Optimization in the Preprocessing Stage, improves the model's robustness and reliability, ensuring effective differentiation between actual falls and random noise. Addressing this limitation enhances the system's accuracy and applicability in diverse environments, contributing to the overall effectiveness of the HFD-Adv methodology.

— **Limited Data Diversity**: In Wi-Fi CSI-based HAR, the sensitivity of CSI data to slight user or environmental changes poses a significant limitation. This affects data consistency and model generalizability across various demographics, environments, and scenarios. The HFD-Adv methodology addresses this by employing dimensionality reduction techniques like PCA to integrate diverse data sources, creating a more generalized model. Additionally, a refined transfer learning approach is utilized, which allows the model to adapt to new users and environments, even if it was not initially trained on a highly diverse dataset. This strategy involves fine-tuning the model with specific instances from new users to enhance adaptability and generalization. Implemented in both the Preprocessing and Enhanced-processing Stages, these strategies improve the model's robustness and accuracy, mitigating the impact of limited data diversity. As a result, the HFD-Adv system remains effective across various conditions, enhancing its performance and reliability in real-world applications.

2. **Processing Technique Concerns**:

— **Overfitting Training Data**: A significant limitation in Wi-Fi CSI-based HAR is overfitting, where the model fits noise in the training data, leading to poor generalization on new data. This compromises the model's reliability on unseen data. The HFD-Adv methodology addresses this by employing a weighted loss function optimized with GAs, assigning different weights to fall and non-fall classes to minimize *false negatives* (FNs) and enhance generalization. Additionally, a refined transfer learning approach is utilized, fine-tuning the model with data from new users to expose it to diverse patterns and further mitigate overfitting. Implemented in both the Processing and Enhanced-processing Stages, these strategies improve the model's robustness and adaptability, ensuring effective performance across various demographics and environments. By addressing overfitting, the HFD-Adv methodology enhances the system's accuracy and reliability in real-world applications.

— **Hyper-parameter Sensitivity**: Another significant limitation in Wi-Fi CSI-based HAR is hyper-parameter sensitivity, where the model's performance heavily depends on the choice of parameters, leading to suboptimal outcomes. This can cause reliability issues. The HFD-Adv methodology addresses this by employing optimization techniques, particularly GAs, to automate the tuning process of the weighted loss function. This allows for optimal weight identification, minimizing FNs and enhancing model reliability. Additionally, the LRFinder mechanism is used to optimize the learning rate, further reducing dependence on manual tu-

ning. Implemented in the Processing Stage, these strategies enhance the model's precision and adaptability, ensuring efficient and effective training. By addressing hyper-parameter sensitivity, the HFD-Adv methodology improves the system's performance and reliability, making it more robust in real-world applications.

— **Long Training Period**: Another limitation in Wi-Fi CSI-based HAR is the long training period required for model convergence. The significant time and resources needed for model training impede rapid prototyping and experimentation. The HFD-Adv methodology addresses this by employing the LRFinder mechanism, which quickly identifies the optimal learning rate, accelerating model convergence and potentially reducing overall training time. Implemented in the Processing Stage, this strategy fosters efficient and effective training by optimizing the learning rate, thereby mitigating the impact of prolonged training periods. By addressing this limitation, the HFD-Adv methodology enhances the system's efficiency, allowing for quicker iterations and improved adaptability in real-world applications.

3. **Post-processing Mechanism Concerns**:

— **Absence of Post-processing Mechanism and Static Post-processing Mechanism**: Significant limitations in Wi-Fi CSI-based HAR include the absence of an effective post-processing mechanism and the static nature of existing ones. These limitations hinder the enhancement of primary processing results, leading to inadequate detection accuracy and reduced system reliability in varying real-world conditions. The HFD-Adv methodology addresses these issues by introducing the ADM. The ADM reassesses classifier outputs, enhancing detection accuracy by decreasing FNs. Additionally, it incorporates demographic-sensitive adjustments, tailoring sensitivity based on age-specific fall likelihood. Implemented in the Post-processing Stage, these refinements ensure the system adapts to diverse real-world conditions and user demographics, improving overall effectiveness, personalization, and reliability of fall detection. By addressing these limitations, the HFD-Adv methodology significantly enhances the system's accuracy and applicability in real-world applications.

The HFD-Adv methodology is structured into four stages: preprocessing, processing, enhanced-processing, and post-processing. Each stage is designed to address specific limitations and contribute to the overall effectiveness, efficiency, and versatility of the methodology. The output of each stage serves as the input for the subsequent stage, ensuring a cohesive and integrated process.

1. **Preprocessing Stage**: This stage prepares the activity CSI dataset for training the classification model, ensuring system reliability and versatility across various conditions. The key steps include:

— **Data Collection**: Gathering diverse activity CSI data from multiple sources is essential for a comprehensive training set, ensuring the system's generalizability.

— **Data Purity Optimization**: Incorporating 'empty' activity data enhances the model's discernment between human activities and inactivity or noise, improving robustness and real-life applicability.

— **Data Harmonization**: Dimensionality reduction techniques like PCA are used to integrate data from various sources, leading to a more generalized model capable of adapting to different scenarios.

— **Data Split**: A user-specific train-test split of the dataset allows for a robust evaluation of the model's performance on new, unseen data.

— **Data Structuring**: Overlapped framing is implemented to handle the complexities of multi-dimensional CSI data signals, enhancing temporal resolution and ensuring comprehensive data representation.

— **Data Integrity Optimization**: Addressing incomplete frames through a threshold-based padding strategy maintains data integrity, balancing data preservation and representativeness.

These steps lay a solid foundation for DL model training in CSI-based HAR by producing preprocessed data, which serves as the input for the next stage, where the fall detector model is developed.

2. **Processing Stage**: This stage focuses on developing an advanced flexible DL model, ensuring precision and efficiency in fall detection. It encompasses several key strategies:
   — **Adaptive Classifier Development**: Introducing ActiNet, a versatile CNN designed for varied HAR tasks, emphasizing adaptability in model architecture to suit diverse requirements.
   — **Overfitting Prevention**: L2 regularization and dropout layers are employed in the classifier to ensure robust generalization to new data, a critical feature for reliable HAR systems.
   — **Loss Function Optimization**: A weighted binary cross-entropy loss function, refined with GAs, enhances precision and adaptability in model training. This approach effectively minimizes FNs, essential in fall detection contexts.
   — **Learning Rate Optimization**: A learning rate optimization mechanism is utilized for optimal learning rate selection, fostering efficient and effective training by avoiding arbitrary hyperparameter choices.

   These strategies develop an adaptive classifier, fine-tuned for efficiency and accuracy, tailored for the complexities of CSI-based HAR. The output of this stage is a trained fall detector model, which serves as the input for the enhanced-processing stage.

3. **Enhanced-processing Stage**: This stage represents a strategic extension of the initial processing stage, focusing on enhancing the transferability and adaptability of the DL model, particularly for new and unseen users. Key enhancements include:
   — **Augmenting New User Novel Insights**: Gathering CSI data from new users via interaction simulations provides a baseline of their unique activity patterns, despite limited initial data. Data Augmentation (DA) is then applied to enrich the model's understanding with a broader spectrum of user-specific behaviors.
   — **New User Adaptation**: Refined transfer learning strategy is employed to address model performance variability in different conditions. This is essential for calibrating the model to specific user profiles and environmental variables, ensuring precise interpretation of distinct user characteristics and environmental influences on CSI data.

   These enhancements improve not only the model's adaptability and transferability but also its precision and efficiency in diverse user situations. The output of this stage is an enhanced fall detector model, which serves as the input for the post-processing stage.

4. **Post-processing Stage**: This stage refines the outputs from the processing and enhanced-processing stages to ensure final results are precise and tailored to specific user needs, particularly for groups like the elderly. Key refinements include:
   — **Classifier Outputs Reassessment**: An ADM is executed to reassess the DL model classifications, decreasing FNs and enhancing fall identification accuracy.
   — **Demographic-sensitive Adjustment**: Customizing fall detection sensitivity in ADM based on age-specific fall likelihood, enhancing effectiveness and personalization.

   These refinements ensure precise, user-specific results by producing refined detector predictions, which are the final output of the HFD-Adv methodology.

FIGURE 4.2 provides a schematic representation of this methodology, illustrating the interconnectedness and flow between the stages.

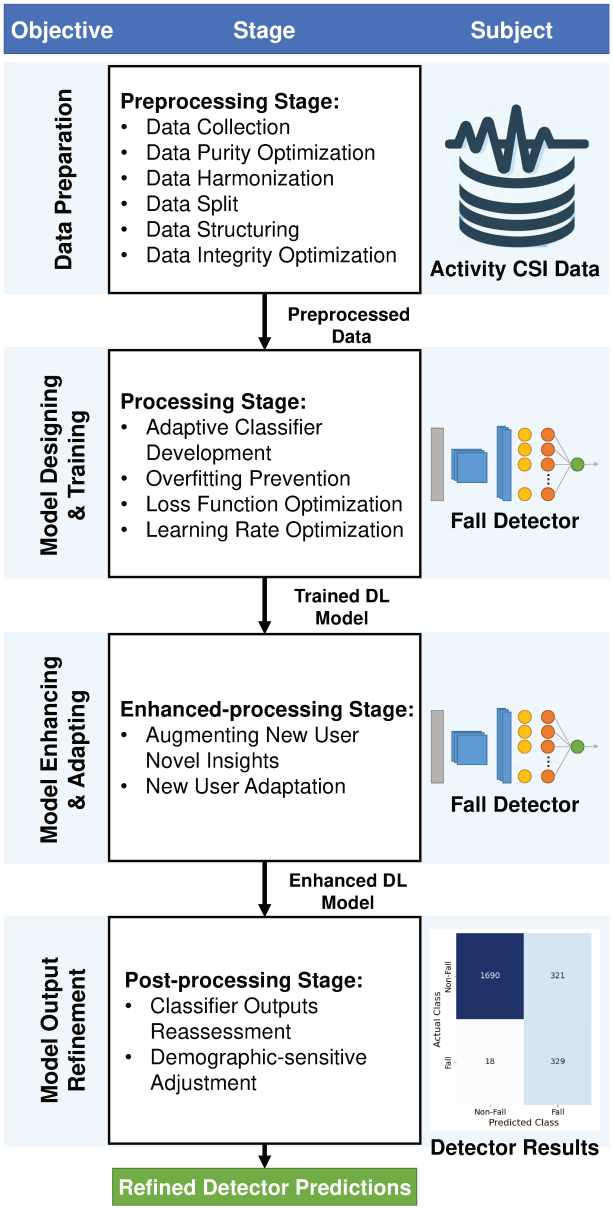


FIGURE 4.2 – HFD-Adv four-stage methodology schema: An illustrative overview of the HFD-Adv four-stage methodology, highlighting the objective, tasks, and subject of each stage. The objectives represent the goals of each stage, the tasks detail the specific activities performed, and the subject indicates the main component being worked on or produced in each stage. The arrows between stages represent the flow of processed data and models, showing how the output of each stage serves as the input for the next, culminating in refined detector predictions as the final output.

3 Data Preparation and Signal Preprocessing

The preprocessing stage is a critical initial step in the HFD-Adv four-stage approach, aimed at preparing the activity CSI dataset for effective fall detection. This stage is essential for ensuring the reliability and versatility of the system by facilitating the training of DL models. Key steps in this stage include data collection, structuring, calibration with empty activity data, dimensionality reduction using PCA, and handling of incomplete frames. These steps are outlined in FIGURE 4.3. The following subsections detail the specific steps involved in this stage.

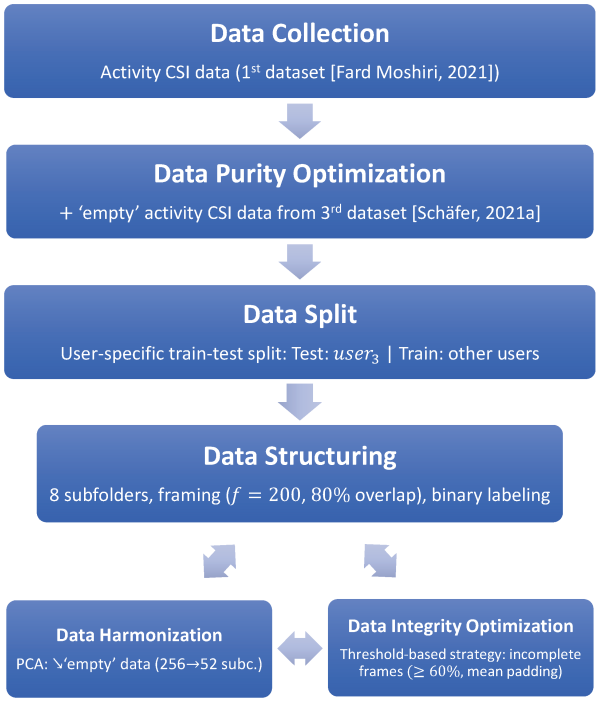


FIGURE 4.3 – Steps of the preprocessing stage: This figure outlines the key steps involved in the preprocessing stage of the HFD-Adv methodology. It begins with data collection, followed by data purity optimization by incorporating 'empty' activity data. The data is then split using a user-specific train-test split strategy. Next, data structuring is performed through framing and binary labeling. Finally, data harmonization using PCA and data integrity optimization via threshold-based handling of incomplete frames are conducted to ensure a robust and well-structured dataset for training DL models.

3.1 Collection and Structuring of Activity CSI Data

In the HFD-Adv four-stage approach, the preprocessing stage involves preparing the activity CSI dataset for fall detection task, utilizing the "First CSI Dataset" introduced in Chapter 1 (Section 8). This dataset, provided by [Fard Moshiri, 2021], includes CSI amplitude data signals from seven daily human activities: bending, falling, lying down, running, sitting down, standing up, and walking, all conducted indoors and collected using the Nexmon CSI Extractor. Three participants, referred to as $user_1$, $user_2$, and $user_3$, performed each activity 20 times, with each trial's data recorded in a separate CSV file.

We opted to use this dataset because it clearly labels each activity trial with a corresponding user identity, enabling blind-user testing (see Subsection 3.3). In contrast, the other datasets used in this thesis lack such organization, which limits their suitability for the strict user-specific train-test split strategy required for robust evaluation of our methodology.

In our study, we implemented overlapped framing to preprocess the continuous CSI amplitude data signals, segmenting them into overlapping frames to improve temporal resolution. By experimenting with frame sizes ranging from 500 to 100 samples and overlap percentages of 20%, 50%, and 80%, we determined the optimal configuration to be a 200-sample `frame_size` (equivalent to the sampling rate), with a 40-sample hop length, achieving 80% overlap. This setup effectively captures detailed signal representations and quick transitions within signals, essential for detecting nuances in human activities. For our (fall vs. non-fall) binary classification, frames indicating falls were labeled '1', while all others were labeled '0'.

3.2 System Calibration with Empty Activity Data

A significant limitation identified within the domain of Wi-Fi CSI-based HAR pertains to the absence of empty data signals, which are essential for environment-specific calibration. This limitation falls under the category of Data-Related Concerns. It is critical because the lack of empty activity data hampers the model's ability to distinguish between human activities and environmental noise accurately. Addressing this concern is vital for enhancing the reliability and robustness of fall detection systems, as it ensures the model can effectively differentiate between actual falls and non-activity periods.

To address the absence of empty data signals in Wi-Fi CSI-based HAR, we incorporated 60 CSV files from the "Third CSI Dataset," provided by [Schäfer, 2021a], representing 'empty' activity (non-activity) CSI amplitude data. Although the 3rd dataset employs a distinct CSI data-collection setup compared to the 1st dataset, non-activity segments are unavailable in the latter. Consequently, importing these external "empty" activity signals was the most practical approach to train the model to distinguish human motion from a purely environmental baseline. We note that this introduces potential domain mismatches, discussed further in Subsection 3.4.

To address the absence of empty data signals in Wi-Fi CSI-based HAR, we incorporated 60 CSV files from the "Third CSI Dataset" (introduced in Chapter 1) representing 'empty' activity (non-activity) CSI amplitude data. Although this dataset was collected with different hardware and in a different environment, importing these external non-activity trials was the most feasible and technically sound way to inject an environmental baseline signals. Signal baselines (mean amplitude and variance) inevitably differ across collection set-ups, so all activity frames are standardized to zero mean and unit variance. For the 256-subcarrier empty activity frames we apply PCA (see Subsection 3.4) to match the 52-dimension feature space of the 1st dataset. This pipeline introduces controlled domain variability while keeping the input dimensionality consistent, enabling the model to learn that non-activity is a distinct scenario.

This step trains the model to recognize periods of no activity, thus improving real-world utility by differentiating between human activities and random Wi-Fi signal noise. By calibrating the fall detection model on 'empty' activity, we enhance its performance and accuracy in real-world applications. Both training and testing sets were organized into eight activity subfolders, including one for 'empty' activity. These subfolders follow the binary labeling strategy described earlier, ensuring uniform preparation across all data.

3.3 User-Specific Train-Test Split

A strategic user-specific train-test split S_{user_3} was implemented to segregate $user_3$'s data for testing. This method ensured that our binary classifier was trained on separate data, providing an unbiased and precise assessment of its efficacy on unseen user data. Although this method centered on $user_3$, it offers a scalable strategy applicable to any user, enabling consistent model performance evaluation.

3.4 Dimensionality Reduction with PCA

To seamlessly integrate the 'empty' subfolder from the 3rd dataset, containing 256 subcarriers, with our primary dataset, which encompasses CSI amplitude signals from seven distinct human activities captured with 52 subcarriers, we employed *Principal Component Analysis* (PCA) for dimensionality reduction. This precise application of PCA aligns the higher-dimensional 'empty' data with the dimensional space of our primary activity CSI data, ensuring consistency in our analysis methodology. The PCA transformation is defined as:

$$\text{Transformed_Data} = \text{PCA}(X, 52)$$

where X denotes the matrix of CSI amplitude signals from the 'empty' subfolder, and the number 52 represents the target number of principal components (subcarriers) we aim to match with our primary dataset. This targeted dimensionality reduction to 52 principal components is pivotal for merging the 'empty' data with the activity CSI signals, allowing for a unified feature space critical for our fall detection model's learning process.

By focusing on the principal components, we distill the most meaningful variance from the 'empty' data, aligning it with the original dataset's structure. This ensures that our model is exposed to a coherent feature space, enhancing its ability to distinguish between different types of activities and periods of inactivity.

Employing PCA in this manner is a technical requirement and a strategic enhancement to the model's training regime. It ensures that the feature set derived from the 'empty' and activity-based data is optimized for relevance and discriminative power, improving the model's efficacy in real-world applications [Ma, 2019].

By integrating the 'empty' activity data into our training and testing sets, we enhance the fall detection model's ability to discern between actual human activities and random Wi-Fi signal noise. This step is essential for real-world utility, providing the model with a broader understanding of its operational environment. The unified feature space created through PCA plays a crucial role in maintaining the integrity and performance of our fall detection model, ultimately leading to more reliable and accurate fall detection capabilities.

Combined with the prior standardization step, this technical bridge broadens the training data without any additional data collection sessions and directly mitigates the *limited data diversity* limitation highlighted in Subsection 2.2. Consequently, the model is exposed to a wider range of environmental baseline patterns—and, by extension, could ingest other activity types collected on alternate setups in future domain-adaptation scenarios.

3.5 Handling Incomplete Frames with Threshold-Based Padding Strategy

In the domain of CSI-based HAR, particularly when preparing CSI amplitude data for DL model training, handling incomplete frames is crucial. These frames, often found towards the end of the signal, can affect data integrity and representativeness. To handle such incomplete signal frames, we implemented a threshold-based padding strategy. While this step has only a marginal effect when a very high overlap of 80% is used (as in our experiments), it becomes essential for lower overlaps (e.g., 20–50%) or shorter frame sizes, where otherwise a non-trivial portion of the signal would be silently lost.

A significant limitation within the domain of CSI-based HAR is the issue of missing training data, specifically incomplete training frames. This limitation falls under the "Data-Related Concerns" category. Incomplete frames lack the full set of data points required for consistent frame size, which is critical for training DL models effectively. Moreover, a common solution in the literature is to simply discard incomplete frames. However, this approach can result in the loss of valuable data as it leads to some samples not being considered. Specifically, when incomplete frames are discarded, the samples they contain, which were not part of any previous complete frame, are lost. This further exacerbates the limitation of missing training data.

Our strategy involves several key steps, each contributing to the effective preparation of the CSI amplitude data for DL model training. Below is a step-by-step high-level description of the logic behind the threshold-based padding strategy:

1. **Loading and Processing the Data:** The CSI amplitude data signal from CSV files is loaded into an array. This array is used to extract frames based on a specified `frame_size` and `hop_length`. The `frame_size` is the predefined number of samples in each frame (e.g., 200 samples), and the

hop_length is the number of samples to shift for the next frame (e.g., 40 samples), which results in overlap between consecutive frames. Each frame, referred to as `CSIamp_frame`, can be complete or incomplete.

2. **Frame Extraction:** Frames (`CSIamp_frame`) are extracted using the predefined `frame_size` (e.g., 200 samples) and `hop_length` (e.g., 40 samples) to ensure overlap. This overlap is crucial for capturing continuous signal characteristics across frames.
3. **Handling Incomplete Frames:** Incomplete frames are those with fewer samples than the required `frame_size`. To address this, we calculate the percentage of real samples in the remaining_samples part of the frame. This percentage guides the decision to pad or discard the frame based on a predefined threshold.
   — **Calculation of remaining_samples:** Remaining samples are equivalent to the hop_length, indicating the number of new samples needed to complete a frame (non-overlapping part).
   — **Calculation of real_samples_in_remaining:** For an incomplete frame (`CSIamp_frame`) starting at `start_index` and ending at `real_frame_end`:

   $$\text{real_samples_in_remaining} = \text{real_frame_end} - (\text{start_index} + (\text{frame_size} - \text{hop_length}))$$

   This calculation determines the actual signal samples in the remaining (non-overlapping) portion of the incomplete frame.
   — **Threshold-Based Padding Decision:** The percentage of real samples in the remaining_samples part of the frame is calculated as follows:

   $$\text{real_samples_percentage} = \left(\frac{\text{real_samples_in_remaining}}{\text{remaining_samples}} \right) \times 100$$

   If this percentage exceeds the remaining_samples_threshold (e.g., 60%), the frame is padded; otherwise, it is discarded. This threshold is set based on empirical observations and domain-specific knowledge.
4. **Padding Incomplete Frames:** When the percentage of real samples is above the threshold, the frame is padded using the mean value of the multidimensional signal, ensuring minimal disruption to the signal's characteristics. The padding size is calculated as:

   $$\text{pad_size} = \text{frame_size} - \text{LengthOf}(\text{CSIamp_frame})$$

   where frame_size is the desired frame size, and CSIamp_frame is the current incomplete frame. Padding is achieved by appending values to the incomplete frame until it matches the desired frame size. Specifically, we utilize the mean of the multidimensional signal as the padding value. By using the mean to pad, we fill the incomplete data in such a way that it doesn't introduce significant distortion. This method is superior to zero-padding as it maintains the signal's inherent distribution, balancing data integrity and representativeness. Using the mean helps ensure that the added values do not skew the overall data distribution [García-Laencina, 2010].

Algorithm 1 encapsulates the logic behind the threshold-based padding strategy optimization:

Providing concrete examples helps in understanding the practical application of the threshold-based padding strategy:

— **Complete Frames:** Given a `frame_size` of 200 and a `hop_length` of 40, frames such as these are complete, each containing exactly 200 samples:

Algorithm 1 Optimize Data Integrity with Threshold-Based Padding Strategy

Require: `CSI_amplitude_data`, `frame_size`, `hop_length`, `remaining_samples_threshold`
Ensure: Optimized data integrity and representativeness of CSI frames

```
for each CSV file do
  Load CSI_amplitude_data into an array multidimensional_signal
  Initialize frames, real_frame_ends
  signal_mean ← mean(multidimensional_signal)
  for j = 0 to length(multidimensional_signal) step hop_length do
    Extract CSIamp_frame from multidimensional_signal[j:j + frame_size]
    real_frame_end ← j + length(CSIamp_frame)
    if length(CSIamp_frame == frame_size) then
      Append CSIamp_frame to frames
      Append real_frame_end to real_frame_ends
    else
      pad_size ← frame_size - length(CSIamp_frame)
      remaining_samples ← hop_length
      real_samples_in_remaining ← real_frame_end - (j + (frame_size - hop_length))
      real_samples_percentage ← (real_samples_in_remaining / remaining_samples) * 100
      if real_samples_percentage ≥ remaining_samples_threshold then
        Append real_frame_end to real_frame_ends
        Pad CSIamp_frame with signal_mean to match frame_size
        Append CSIamp_frame to frames
      end if
    end if
  end for
end for
return Optimally processed CSI frames
```

— Start: 0, End: 200, Real End: 200

— Start: 40, End: 240, Real End: 240

— Start: 80, End: 280, Real End: 280

— **Incomplete Frame with Sufficient Real Samples:** For a frame starting at 360 and ending at 559:

— Remaining Samples: 40 (hop_length)

— Real Samples in Remaining: $559 - (360 + (200 - 40)) = 559 - 520 = 39$

— Real Samples Percentage: $\left(\frac{39}{40}\right) \times 100 = 97.5\%$

Various thresholds (40%, 50%, and 60%) were evaluated, and a 60% threshold was selected as optimal. Since 97.5% exceeds the remaining_samples_threshold, this frame is padded with 1 sample and included.

— **Incomplete Frame with Insufficient Real Samples:** For a frame starting at 360 and ending at 550:

— Remaining Samples: 40 (hop_length)

— Real Samples in Remaining: $550 - (360 + (200 - 40)) = 550 - 520 = 30$

— Real Samples Percentage: $\left(\frac{30}{40}\right) \times 100 = 75\%$

Since 75% exceeds the remaining_samples_threshold, this frame is padded with 10 samples and included. This example illustrates how the strategy handles a frame that is less complete but still meets the threshold.

— **Incomplete Frame Below Threshold:** For a frame starting at 360 and ending at 530:

— Remaining Samples: 40 (hop_length)

— Real Samples in Remaining: $530 - (360 + (200 - 40)) = 530 - 520 = 10$

— Real Samples Percentage: $\left(\frac{10}{40}\right) \times 100 = 25\%$

Since 25% is below the remaining_samples_threshold, this frame is discarded.

This threshold-based padding strategy ensures that only frames with a significant amount of real data are retained, preserving data integrity and representativeness. By accurately calculating the real samples in the remaining part of an incomplete frame, we make informed decisions to pad or discard frames, thus enhancing the dataset quality for DL model training. This approach balances the need to preserve valuable data while avoiding the inclusion of predominantly padded frames, maintaining the dataset's intrinsic characteristics. This balance helps in achieving optimal model performance by using frames that are sufficiently representative of genuine activity samples.

4 ActiNet Flexible CNN Classifier

The processing stage in the HFD-Adv methodology focuses on developing an advanced, flexible DL model to ensure precision and efficiency in fall detection. This stage is crucial for creating a robust classifier that can generalize well to new data and adapt to various HAR tasks. Key strategies in this stage include the development of an adaptive classifier, ActiNet, with overfitting prevention techniques, loss function optimization using GAs, and learning rate optimization.

At the core of this stage is the proposed ActiNet, a CNN classifier. ActiNet is designed to be adaptable, allowing it to handle various HAR tasks, whether binary classification, like distinguishing between fall and non-fall activities, or multi-class classification involving several types of activities. This flexibility marks a paradigm shift towards task-adjustable neural network architectures, serving as a foundational model for diverse CSI-based HAR tasks.

ActiNet can be tailored for different HAR tasks by modifying both the input and output layers according to the specific characteristics of each dataset and the number of classes in the classification task. The input dimensions are dynamically determined by the dataset's configuration and collection processes, structured in frames representing periods of Wi-Fi CSI data. Each frame comprises time steps (*frame_size*) and data subcarriers (*num_subcarriers*), with an added depth dimension for compatibility with CNNs, making the input a 3D tensor (*frame_size*, *num_subcarriers*, *depth*) batched during model training. This ensures that the model can adapt to the unique temporal and spectral resolutions of different CSI datasets. For binary classification tasks, such as distinguishing between fall and non-fall activities, the output layer is configured with a single neuron and a sigmoid activation function. For multi-class classification tasks, the output layer is adapted with multiple neurons and a softmax activation function.

Despite its adaptability, ActiNet maintains a consistent underlying architecture across various tasks, providing a familiar starting point for researchers and practitioners. This consistency ensures a streamlined research process and more standardized results. Moreover, ActiNet's efficiency is preserved, ensuring that real-time activity recognition is feasible without compromising on speed or accuracy. With a consistent architecture, research efforts can focus on refining and fine-tuning the model rather than building from scratch, saving time and enhancing the reliability of findings.

The key strategies mentioned above, development of the adaptive classifier ActiNet, overfitting prevention techniques, loss function optimization using GAs, and learning rate optimization, are detailed in the following subsections.

4.1 Architecture for Fall Detection

In this study, ActiNet is specifically configured to address the binary classification task of HFD. The architecture of ActiNet is meticulously crafted to ensure high accuracy and efficiency in distinguishing fall events from other activities. The following components make up the architecture:

— **Input Layer**: This layer processes 3D tensors with dimensions $(200, 52, 1)$, where 200 represents the *frame_size*, 52 corresponds to the *num_subcarriers*, and 1 is the added *depth* dimension to ensure compatibility with CNNs. This structure allows the model to capture both temporal and spectral variations in the CSI data.

— **Padding Layer**: This layer utilizes zero padding to maintain consistent spatial dimensions through subsequent convolutional layers, ensuring that edge information is not lost during convolution operations.

— **Convolutional Layers**: The core of the feature extraction process involves three convolutional layers. Each layer employs 3×3 and 2×2 kernels with 32 filters to capture intricate patterns in the data. These layers use a Leaky ReLU activation function for better gradient flow, MaxPooling to reduce the dimensionality, and BatchNormalization to stabilize activations and accelerate convergence.

  — **First Convolutional Layer**: Uses 3×3 kernels, 32 filters, followed by Leaky ReLU activation ($\alpha = 0.1$), MaxPooling with 3×3 window, stride 2×2, and BatchNormalization.

  — **Second Convolutional Layer**: Also uses 3×3 kernels, 32 filters, followed by Leaky ReLU activation ($\alpha = 0.1$), MaxPooling with 3×3 window, stride 2×2, and BatchNormalization.

  — **Third Convolutional Layer**: Uses 2×2 kernels, 32 filters, followed by Leaky ReLU activation ($\alpha = 0.1$), MaxPooling with 2×2 window, stride 2×2, and BatchNormalization.

— **Flatten Layer**: This layer transforms the 3D output from the convolutional layers into a 1D vector, preparing it for the fully connected layers.

— **Fully Connected Layer**: This Dense layer comprises 64 units with a LeakyReLU activation function to introduce non-linearity. A Dropout layer with a rate of 0.3 is included to prevent overfitting by randomly dropping units during training.

— **Output Layer**: The final layer is a single neuron that calculates the weighted sum $z = \sum_{i=1}^{n} w_i x_i + b$, where w_i are the weights, x_i are the activations from the previous layer, b is the bias term, and n is the number of inputs. A sigmoid activation function $\sigma(z) = \frac{1}{1+e^{-z}}$ is applied to obtain the predicted probability $\hat{y}$ of a fall. The threshold for fall detection is set at $\hat{y} = \sigma(z) \geq 0.5$, marking a frame as a fall if this condition is met.

To prevent overfitting, L2 regularization is applied to the convolutional and dense layers. This adds a penalty to the loss function, expressed as $\text{Loss} = \text{Original Loss} + \lambda \sum w_i^2$, where λ is the regularization strength and w_i are the model parameters. L2 regularization is preferred over L1 due to its stability in maintaining model weights [Goodfellow, 2016]. For a visual representation of the ActiNet architecture, see FIGURE 4.4.

4.2 Weighted Loss Function Optimization using Genetic Algorithm

In compiling ActiNet, the optimization process involves defining key hyperparameters: the loss function, optimizer with its learning rate, and training metrics. Given the critical nature of fall detection, where FNs can have severe consequences compared to false positives (FPs), we implement a tailored loss function. This weighted binary cross-entropy loss function places a higher penalty on FNs to enhance

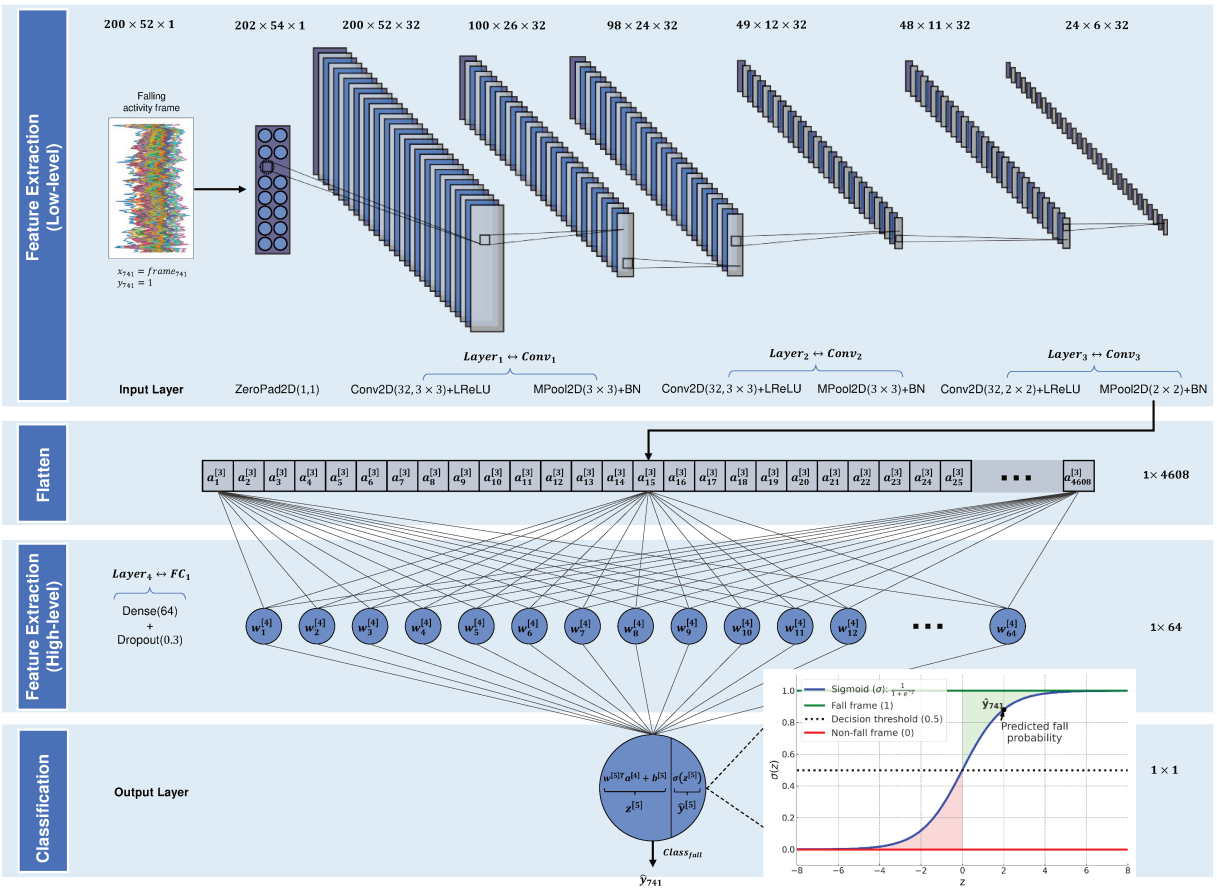


FIGURE 4.4 – Architecture of ActiNet fall detector: This figure illustrates the architecture of the ActiNet model specifically configured for fall detection. It includes the input layer processing 3D CSI data tensors, three convolutional layers for feature extraction, a flatten layer, a fully connected layer with dropout for overfitting prevention, and an output layer with a sigmoid activation function to predict fall probabilities. The architecture ensures robust feature extraction and efficient classification for real-time activity recognition, particularly focusing on fall events.

the predictive accuracy of the fall detection model. The standard binary cross-entropy loss is expressed as [Goodfellow, 2016]:

$$L_{\text{binary_crossentropy}} = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) \quad .$$

Here, y represents the true label (1 for fall, 0 for non-fall), and $\hat{y}$ denotes the predicted probability of a fall event. To emphasize the cost of FNs over FPs, we introduce weights w_1 and w_0 for fall and non-fall classes, respectively, leading to the modified loss equation:

$$L_{\text{w_binary_crossentropy}} = -(w_1 \cdot y \log(\hat{y}) + w_0 \cdot (1 - y) \log(1 - \hat{y})) \quad .$$

In this formula, w_1 and w_0 are the weights for the fall and non-fall classes. For a positive class ($y = 1$), the loss calculation involves w_1, while for a negative class ($y = 0$), it involves w_0. By adjusting w_1 relative to w_0, a higher penalty is assigned to FNs by increasing w_1. This weighted loss function $L_{\text{w_binary_crossentropy}}$ differentiates the importance of each class, which is crucial in scenarios where misclassifying a fall (positive case) is more consequential than a non-fall (negative case) [He, 2009].

A significant limitation within the domain of Wi-Fi CSI-based HAR pertains to overfitting training

data, where the model tends to fit noise in the training data, leading to poor generalization on new data. This falls under the category of Processing Technique Concerns. Overfitting is critical as it compromises the model's ability to perform reliably on unseen data. The use of a weighted loss function and GAs helps to fine-tune the model, reducing the impact of overfitting. By assigning different weights to fall and non-fall classes, the model is more penalized for FNs, which helps it to generalize better and not overfit the training data.

Another limitation addressed is hyper-parameter sensitivity, also under Processing Technique Concerns, where heavy dependence on the choice of parameters can lead to suboptimal performance and reliability issues. Manual tuning and cost-based adjustments of these weights depend on domain knowledge but are often ineffective when the exact cost of misclassification is uncertain. To address this, we utilize optimization techniques, particularly GAs, which offer an efficient way to automate the tuning process through simulated natural selection, allowing for the identification of optimal weights after multiple generational trials [Yang, 2020]. GAs help mitigate hyper-parameter sensitivity by automating the tuning process, thus enhancing model reliability and performance.

We employ the GA for fine-tuning the weights w_1 and w_0 of the weighted loss function $L_{\text{w_binary_crossentropy}}$. The GA evolves a population of chromosomes, each representing a weight pair, through the following steps:

1. **Initialization**: Start with a diverse population of 20 chromosomes, each encoding a random weight pair within [0, 5].
2. **Fitness Evaluation**: The fitness function is a critical component in our GA, playing a central role in the selection process and significantly influencing the algorithm's overall performance. The evaluation of each chromosome's fitness involves integrating its weights into the loss function of ActiNet and then assessing the model's performance on a validation set. The fitness function is specifically designed to strike a balance between detecting actual falls (recall) and minimizing false alarms (FAR), crucial for ensuring patient safety and system reliability. The fitness function calculates a weighted score based on:

$$\text{fitness} = w_1 \times \text{recall} - w_2 \times \text{FAR}$$

   where $w_1 = 0.7$ and $w_2 = 0.3$ reflect the relative importance of maximizing recall and minimizing FAR, respectively. This strategic weighting emphasizes recall to reduce missed falls while considering the impact of false alarms, aligning with our goal of optimizing fall detection through a nuanced understanding of system performance needs.
3. **Selection**: Choose the fittest chromosomes based on their fitness scores for reproduction.
4. **Crossover**: Generate offspring by mixing genetic material from parent pairs, controlled by a crossover rate of 0.8.
5. **Mutation**: Introduce random weight variations in the offspring, with a mutation rate of 0.1, to explore new solution spaces and maintain genetic diversity.
6. **New Generation Formation**: Form a new generation by combining fit parents and offspring.
7. **Termination**: Repeat the cycle for 10 generations or until improvements are minimal, indicating convergence.

Using these parameters, the GA progressed over generations, ultimately yielding optimal weights for the loss function: $w_1 = 3.6336$ and $w_0 = 0.4980$. The detailed steps of the GA optimization are presented in Algorithm 2.

Algorithm 2 Optimize Loss Function Weights with GA

Require: `training_set`, `validation_set`, `population_size`, `num_generations`
Ensure: Optimized weights w_1 & w_0
  Procedure `GA_Optimization`
  Initialize population with chromosomes having random weights (w_1, w_0)
  for each generation in `1` to `num_generations` **do**
    for each chromosome in population **do**
      Compute fitness:
        create new model
        Compile model with $L_{\texttt{w_binary_crossentropy}}$
        Train model on `training_set`
        Evaluate model on `validation_set`
        Compute `recall` and `FAR`
        Fitness $\leftarrow 0.7\times$ `recall` $-0.3\times$ `FAR`
    end for
    Sort population by fitness in descending order
    Select top half of population as parents
    `Offspring` $\leftarrow$ Empty set
    while `offspring_size` $<$ `parents_size` **do**
      Select `parent1`, `parent2` from `parents`
      `child1`, `child2` $\leftarrow$ `Crossover(parent1, parent2, crossover_rate)`
      `Mutate(child1, mutation_rate)`
      `Mutate(child2, mutation_rate)`
      Add `child1`, `child2` to `Offspring`
    end while
    Replace least fit half of population with `Offspring`
    if best fitness has not improved significantly **then**
      break
    end if
  end for
  return Optimal weights of best chromosome
  End Procedure

4.3 Learning Rate Optimization with LRFinder

In ActiNet, the Adam optimizer, combining the strengths of AdaGrad and RMSProp, is utilized to refine neural network attributes, focusing on loss minimization. Central to its efficacy is the Learning Rate (LR), which is crucial for adjusting weights based on error estimates. To avoid the pitfalls of prolonged training due to a low LR or missing the optimal solution with a high LR, we introduce the 'LRFinder' mechanism. This mechanism aligns with well-regarded methods of dynamically modulating LRs to match training evolution and loss patterns, as noted in LR optimization literature [Goodfellow, 2016 ; Kingma, 2014]. LRFinder effectively identifies the optimal LR, promoting efficient training outcomes and enhancing the robustness of the fall detection model.

In the previous subsection, we addressed hyper-parameter sensitivity by optimizing the weights of the loss function using Genetic Algorithms (GAs). Another crucial hyper-parameter impacting model performance is the learning rate. Hyper-parameter sensitivity, a key concern under Processing Technique Concerns, can lead to suboptimal performance and reliability issues if not managed properly. The LR-Finder mechanism is specifically designed to tackle this by automating the tuning of the learning rate, thereby reducing the dependence on manual hyperparameter tuning and improving overall performance and reliability.

Additionally, the LRFinder mechanism addresses the limitation of long training periods. This also falls under Processing Technique Concerns. Significant time and resources required for model training

limit rapid prototyping and experimentation. By identifying the optimal learning rate quickly, LRFinder accelerates the convergence of the model, potentially reducing the overall training time.

The LRFinder mechanism functions as follows:

1. **Initialization**: LRFinder begins by establishing a range for the LR, defined by a minimum and maximum value. It also considers the number of steps per epoch and the total number of epochs for the training process.
2. **Dynamic Adjustment**: During training, the LR is dynamically adjusted. Starting from the minimum value, it incrementally increases towards the maximum. This adjustment is linear and correlated with the training progress, ensuring a comprehensive exploration of the LR spectrum.
3. **Tracking and Updating**: After each training batch, the LR and corresponding loss values are recorded. The mechanism then updates the LR for the subsequent batch, maintaining a balance between learning speed and accuracy.
4. **Optimal LR Identification**: The key aspect of this mechanism is identifying the most effective LR. It does this by analyzing the trend of loss values relative to different learning rates. The optimal LR is typically found just before the point where the loss starts to escalate rapidly. This rate ensures efficient training without compromising the model's ability to converge to a solution.

By continuously adjusting and monitoring the learning rate relative to loss, our refined strategy for learning rate optimization aims to enhance the training process. This method offers a data-driven approach to selecting the LR, reducing the risks of suboptimal model performance due to arbitrarily chosen hyperparameters. As illustrated in FIGURE 4.5, LRFinder yielded an optimal LR of 7.5×10^{-3}, aligning with the lowest loss values. Algorithm 3 encapsulates the logic behind the LRFinder optimization.

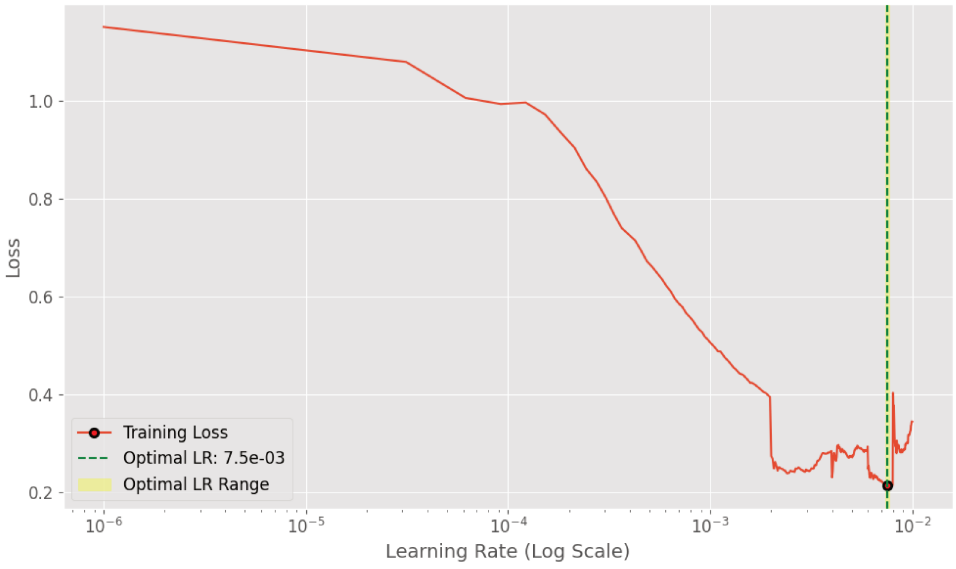


FIGURE 4.5 – Plot of loss values versus learning rates determined by LRFinder: This figure illustrates the relationship between the training loss and various learning rates on a logarithmic scale, as determined by the LRFinder mechanism. The red line represents the training loss across different learning rates, while the green dashed line marks the optimal learning rate of 7.5×10^{-3}. The shaded yellow region indicates the optimal learning rate range, providing a clear visual representation of where the training loss is minimized before it starts to increase rapidly. This plot is crucial for identifying the most effective learning rate that ensures efficient and stable training of the ActiNet model.

Algorithm 3 Optimize LR with LRFinder

Require: `min_lr`, `max_lr`, `steps_per_epoch`, `epochs`
Ensure: Optimized LR
  Initialize `current_lr` ← `min_lr`
  Define `lr_increment` ← $\frac{\texttt{max_lr} - \texttt{min_lr}}{\texttt{steps_per_epoch} \times \texttt{epochs}}$
  for `epoch` = 1 to `epochs` **do**
    for `step` = 1 to `steps_per_epoch` **do**
      Update `current_lr` ← `min_lr` + (`lr_increment` × ((`epoch` - 1) × `steps_per_epoch` + `step`))
      Train model with `current_lr`
      Record `current_lr` and corresponding loss
      if loss escalates rapidly **then**
        break
      end if
    end for
    if optimal LR identified **then**
      break
    end if
  end for
  return Optimal LR just before rapid loss escalation

4.4 Hyperparameter Optimization and Training Strategies

In the development of ActiNet, we selected the hyperparameter optimization strategies based on the complexity and requirements of each optimization task. For learning rate optimization, the LRFinder method was chosen for its efficiency and minimal computational demands, aligning well with the task's straightforward nature. In contrast, the intricate challenge of optimizing our loss function for precise fall detection necessitated the use of GAs. Despite their higher computational intensity, GAs provided the required sophistication for fine-tuning the loss function weights. This approach underscores the importance of carefully choosing optimization strategies that are not only effective but also proportionate to the task's complexity, ensuring optimal model performance while maintaining computational efficiency.

In the training of ActiNet, various callbacks were employed for monitoring and optimizing the training process. The EarlyStopping callback halts training when there's no improvement in validation loss for a set number of epochs. The ReduceLROnPlateau callback decreases the learning rate upon a plateau in validation loss, aiding in more precise weight adjustments and potential model enhancements. Additionally, the ModelCheckpoint callback preserves the best-performing model based on validation accuracy. Hyper-parameter selections for model compilation & training phases are detailed in TABLE 4.1.

TABLE 4.1 – Summary of Hyperparameters and Training Strategies for ActiNet

Hyper-parameter	**Chosen Value**
Optimizer	Adam
Learning Rate	7.5×10^{-3}
Loss Function	Weighted Binary Cross-Entropy
Batch Size	64
Number of Epochs	150
EarlyStopping Patience	10 epochs
ReduceLROnPlateau Factor	0.5
ReduceLROnPlateau Patience	5 epochs
ReduceLROnPlateau Min LR	0.001

5 Transfer Learning for Model Adaptability

In our novel approach, HFD-Adv, we introduce a strategic enhanced-processing stage that extends the initial processing phase. This stage is specifically designed to improve the model's adaptability and generalization capabilities when encountering new, unseen users. By leveraging transfer learning, we aim to calibrate the system for each new user, ensuring robust performance and reliability across diverse real-world conditions.

5.1 Role of Transfer Learning in Enhancing Model Adaptability

In Wi-Fi CSI-based HAR, adapting DL models to new users and environments is challenging due to CSI's sensitivity to even slight user or environmental changes, which affect data consistency despite identical activities [Chen, 2023a]. This sensitivity poses a significant limitation in terms of data diversity, as the lack of representativeness across various demographics, environments, scenarios, and hardware configurations affects the model's generalizability in diverse real-world conditions. This limitation falls under the category of Data-Related Concerns.

To address this limitation, we initially employed dimensionality reduction techniques such as PCA to integrate data from various sources, leading to a more generalized model capable of adapting to different conditions. As discussed in Section 3.4, by integrating the 'empty' activity data into our training and testing sets, we enhanced the model's ability to discern between human activities and environmental noise, improving its robustness and accuracy.

However, it is impractical to train the model on every possible real-world condition due to the vast variability and dynamic nature of CSI data. Therefore, as a complementary approach to further mitigate the impact of limited data diversity, we propose a refined transfer learning approach centered on calibrating our system for each new user. This crucial adjustment, essential in practical scenarios, involves fine-tuning our model with specific instances from new users. This strategy helps to enhance the model's adaptability and generalization capabilities.

Furthermore, this refined transfer learning approach also addresses the limitation of overfitting training data, which falls under the category of Processing Technique Concerns. By fine-tuning the model with data from new, unseen users, we expose the model to diverse patterns, which helps mitigate overfitting and improves its generalization on new data. This ensures that the system adapts to the distinct effects of user characteristics and environmental variables on CSI data, thus enhancing detection precision. This approach ensures that the model remains robust and effective across various demographics and environments, ultimately improving its performance and reliability in real-world applications.

5.2 Transfer Learning Strategy and Implementation for Model Fine-Tuning

Our methodology involves initially training ActiNet, a CNN-based binary classifier, on CSI amplitude data from several users performing seven daily activities (bending, falling, lying down, running, sitting down, standing up, walking, and an 'empty' class representing environmental baseline). To enhance its adaptability, we fine-tune ActiNet with data from a 'left-out' user (designated as $user_3$ in our study), who was not included in the initial training, to simulate real-world calibration when a new user begins using our system. This involves the user performing each of the six non-fall activities and, if possible, imitating a fall.

Key to our transfer learning strategy is the preparation of instances from this 'left-out' user. The process randomly selects one instance (a CSI amplitude data signal) for each of the six non-fall activities as well as for the fall activity. Each chosen instance includes all frames derived from the original activity signal. This selection emulates the real-world scenario where a user performs each activity once during the

calibration phase. To transparently validate our approach, we ensure the selection of only one instance per activity, leaving 19 other instances unexposed to the model. This preparation process enables testing whether the transfer learning improves ActiNet's ability to generalize to this new user's unique data patterns.

We focus on fine-tuning ActiNet's latter layers—Flatten, Dense, Dropout, and Output—while preserving early layers that capture universal features, adapting later ones to the specific patterns of the new user [Goodfellow, 2016; Yosinski, 2014]. The fine-tuning process is implemented through the following steps:

1. **Loading Pre-trained Model**: The pre-trained ActiNet model is first loaded, ensuring that the learned weights and biases are preserved.
2. **Freezing Layers**: All layers except the last four (Flatten, Dense (64 units), Dropout (rate = 0.3), and Output) are frozen, thereby retaining the learned features while enabling the adaptation of the model's higher-level feature detectors to the unique data patterns of the new user.
3. **Replacing Output Layer**: Despite the pre-trained model sharing the same binary classification task, the model's output layer is replaced. This replacement recalibrates the model's focus, adjusting its sensitivity to the specific characteristics of the unseen user. This ensures the model remains robust and accurate, irrespective of whether it's being adapted to a similar task or a different HAR classification with distinct class structures.
4. **Compiling Prepared Model**: Post-modification, the model is recompiled to effectuate the changes. For the fine-tuning phase, the binary cross-entropy loss function and the Adam optimizer (with a learning rate of 0.01) are used.
5. **Training Prepared Model**: The retraining of the classifier is executed over 150 epochs with a batch size of 64. To address potential class imbalances in the retraining data, we compute and apply sample weights, ensuring equitable learning from both classes. Callbacks are employed to enhance training efficiency, including model checkpointing and early stopping based on validation accuracy and loss metrics.

This refined transfer learning strategy enhances ActiNet's adaptability and generalization, improving its performance across various real-world conditions. By retraining on limited, user-specific data, it becomes proactive in adapting to individual user patterns. Essential in dynamic environments, this adaptive learning process fine-tunes the model to diverse conditions, bolstering its versatility and reliability in HFD.

6 Advanced Decision Mechanism for Classifier Outputs Reassessment

6.1 Introduction to the Advanced Decision Mechanism

The concluding segment of the HFD-Adv methodology, designated as the post-processing stage, aims to refine the fall detection process, particularly for vulnerable groups such as the elderly. This stage is critical for enhancing the ability of binary classifiers to accurately distinguish falls from non-falls. A significant issue it addresses is the failure of classifiers to detect actual falls (FNs, Type II error), which poses significant risks by potentially delaying necessary assistance. By implementing a post-processing mechanism, this stage enhances the overall effectiveness and personalization of the fall detection system.

A significant limitation identified in this stage is the absence of an effective post-processing mechanism, which falls under the category of Post-processing Mechanism Concerns. The lack of decision mechanisms

hinders the enhancement of primary processing results, leading to inadequate detection accuracy. Additionally, the static nature of existing post-processing mechanisms limits their sensitivity to various real-world conditions, further affecting the system's reliability.

To address these limitations, we introduce the Advanced Decision Mechanism (ADM), which is designed to reassess the outputs of binary classifiers, making the fall detection process more accurate and responsive. By incorporating demographic-sensitive adjustments, the ADM tailors its sensitivity to one of the most important real-world variables: age. This ensures that the system can better handle the diversity of real-world conditions and user demographics.

Post-Classification Refinement via Majority Voting. In time-series classification, many post-classification refinement strategies exist to further improve prediction quality after the primary model (e.g., a DL classifier) has assigned frame-level labels. Among these strategies are 'temporal smoothing' and 'majority voting' (MV) [Ganaie, 2022; Asghari Oskoei, 2007; Wahid, 2020], both commonly used to reduce noise and enhance final decisions.

Majority Voting (MV) is broadly understood as a decision-making strategy in which the class that appears most frequently (or surpasses a certain proportion) among multiple outputs is selected. In *sequence-level MV*, a single model's predictions are aggregated across frames (or time windows) of a single activity or event, and a final label is assigned to the entire sequence. This contrasts with *ensemble MV*, where multiple models vote on the same frame (or instance). Studies such as [Wahid, 2020] demonstrate how majority voting helps smooth frame-level noise in time-series tasks (e.g., EMG-based gesture recognition), ultimately improving overall recognition accuracy.

In our HFD-Adv methodology, we adapt the well-known principle of sequence-level majority voting by introducing an *activity-aware, proportion-based variant of MV* in the post-processing stage. Concretely, we collect the frame-level outputs of our fall detector (ActiNet) for each entire activity signal (a "sequence") and then re-evaluate them collectively. Unlike standard MV (which typically applies a fixed $> 50\%$ rule), our proportion-based approach uses a configurable threshold that can be domain-adapted. In this study, we tie that threshold to age-specific fall risk (e.g., 55% for elderly users), which allows the system to refine frame-level decisions by reclassifying the entire activity as a fall if the proportion of fall frames surpasses the chosen threshold.

From a methodological standpoint, no hard-and-fast universal rule dictates exactly what that threshold must be. Many works tune it empirically on validation data, while others justify it logically based on domain factors such as the severe risk of missing a fall in an elderly user. Hence, we keep the threshold strictly above 50% to retain a majority-based rationale but adjust it *lower or higher* depending on risk—ensuring that false negatives are minimized where they would be most critical. To our knowledge, this study is the first time a sequence-level proportion-based MV approach has been adapted specifically to the Wi-Fi CSI-based HAR domain. The resulting post-processing step is straightforward to implement, computationally light, and easily adjusted (or "tuned") for evolving demographics or application requirements, making it both practical and theoretically grounded in broader time-series classification concepts.

Linking to CSI Data and ActiNet Outputs. Technically, the ADM operates on CSI amplitude data signals segmented into overlapping activity frames, reassessing each frame initially classified as either 0 (non-fall) or 1 (fall). Specifically, each performed activity *act* yields a raw CSI data tensor, $\mathrm{CSI}^{(\mathrm{act})} \in \mathbb{C}^{\mathcal{T} \times N_r \times N_t \times N_s}$, from which we leverage the amplitude component $\left|\mathrm{CSI}^{(\mathrm{act})}\right|$. This amplitude data is then split into overlapping frames $\mathbf{X}_i^{(\mathrm{act})}$ of size $(\texttt{frame_size}, \texttt{num_subcarriers})$. For the fall detection task, ActiNet assigns a probability $\hat{y}_i \in [0, 1]$ to each frame via a single sigmoid output neuron. If $\hat{y}_i \geq 0.5$, the i-th frame is labeled as a fall; otherwise, it is labeled as non-fall.

In the post-processing stage, the ADM reviews these frame-level labels over the entire activity signal. Whenever the fraction of frames labeled as falls exceeds a threshold $\tau_{\text{fall_class_percentage}}$, the ADM re-classifies all frames to fall. This mechanism is particularly useful for mitigating Type II errors, as small misclassifications at the frame level can be corrected when the overall fall likelihood is high.

6.2 Age-Sensitive Threshold Adjustment

A key innovation in this stage is the demographic-sensitive adjustment of the $\tau_{\text{fall_class_percentage}}$. Acknowledging that the probability of falls varies with age [World Health Organization, 2021], we tailor this threshold to align with age-specific fall risks. This adjustment is strategic: it lowers the threshold for higher-risk age groups, enabling the system to identify falls in signals with mixed labels more effectively. This approach is pivotal in reducing the occurrence of FNs, ensuring a more accurate and responsive FD methodology.

The number of falling incidents varies significantly across different age groups, influenced by factors such as health, environment, and lifestyle [World Health Organization, 2021 ; Centers for Disease Control and Prevention, 2023a]. Below is a succinct overview of the likelihood of falls by age group, tailored specifically for healthy individuals in standard environments, thus excluding the elevated risks associated with more hazardous conditions.

— **Infants (0-2 years):** Low likelihood. Limited mobility and supervision reduce fall risk [Chaudhary, 2018 ; Disease Control, 2023].

— **Children (3-12 years):** High likelihood. Increased activity and play heighten fall risk [World Health Organization, 2021 ; Centers for Disease Control and Prevention, 2021].

— **Teenagers (13-19 years):** Moderate likelihood. Prone to risk-taking behaviors [Kahn, 2020].

— **Adults (20-49 years):** Low likelihood. Peak physical condition reduces fall frequency [Klin, 2018].

— **Middle-aged (50-64 years):** Moderate likelihood. Increased risk associated with aging [World Health Organization, 2021].

— **Elderly (65+ years):** High likelihood. The highest risk level is due to physical, sensory, and cognitive changes associated with aging [World Health Organization, 2021].

Based on fall likelihood, we suggest the following age-specific thresholds for our ADM:

— **High Likelihood (Children 3-12 / Elderly 65+):** Threshold at 55%.

— **Moderate Likelihood (Teenagers 13-19 / Middle-aged 50-64):** Threshold at 65%.

— **Low Likelihood (Young Adults 20-49 / Infants 0-2):** Threshold at 75%.

In practical terms, the ADM offers tailored flexibility by allowing users to select an age group, such as 'Elderly' or 'Teenagers'. This choice triggers the *determine_threshold_by_age_group* function to adjust the $\tau_{\text{fall_class_percentage}}$ accordingly, ensuring precise and age-aligned fall detection. This adaptive, user-centric mechanism bolsters the HFD methodology's effectiveness and versatility for managing fall risks among various age groups.

This step is pivotal, as it fine-tunes the sensitivity of our fall detection methodology to align precisely with the fall likelihood specific to each user's age group. Such an approach ensures enhanced accuracy in fall detection, effectively reflecting the varied risk profiles across different age groups. This entire process, including the core logic of the ADM mechanism, is comprehensively detailed in Algorithm 4.

Illustrative Example for Reassessment. Consider a fall activity trial segmented into 17 overlapping frames, each assigned a binary label by ActiNet's sigmoid output. For instance, ActiNet might predict 13 frames as fall and 4 as non-fall, yielding a $fall_class_percentage \approx 76\%$. If the user belongs to an age

group with a $\tau_{\text{fall_class_percentage}}$ of 55%, the ADM will reclassify *all* frames to fall, thereby preventing a missed detection in scenarios where the majority of frames indeed reflect a fall event but a small subset is misclassified. For younger adults with a higher threshold, these same frame-level outputs might or might not trigger a full reclassification, highlighting how age-adaptive sensitivity can shape the final decision.

Algorithm 4 Enhance Fall Detection Capability with Age-Sensitive ADM

Require: `test_file_names`, `predicted_labels`, `Age category`
Ensure: Adjusted fall detection predictions
  $\tau_{\texttt{fall_class_percentage}} \leftarrow$ *determine_threshold_by_age_group*(`Age category`)
  Initialize `adjusted_predictions` ← `predicted_labels`
  for each `file_name` in unique(`test_file_names`) **do**
    Identify indices `file_idx` of `file_name` in `test_file_names`
    Extract predictions `file_preds` ← `predicted_labels[file_idx]`
    Calculate `fall_class_percentage` ← mean(`file_preds`)
    if `fall_class_percentage` $\geq \tau_{\texttt{fall_class_percentage}}$ **then**
      Set all `adjusted_predictions[file_idx]` to 1 (fall)
    end if
  end for
  return `adjusted_predictions`

7 Results and Evaluation

7.1 Evaluation Metrics and Benchmarking

This section assesses the proposed HFD-Adv methodology's performance, using the testing set prepared in Section 3. Our primary objective is to overcome challenges in fall detection with new user data. Accordingly, we employ the user-specific train-test split S_{user_3}, which isolates instances related to a specific user ($user_3$) from the training set to evaluate the methodology's generalization to new users. This user is referred to as left-out $user_3$.

Note on Scalability: Although the experiments here focus on $user_3$, we also performed the same set of experiments for another left-out user ($user_2$) to demonstrate the methodology's consistent effectiveness regardless of which user is chosen. The corresponding results for S_{user_2} are provided in Appendix 9, reaffirming that HFD-Adv adapts effectively to different users.

For evaluation, seven experiments were conducted:

1. **Experiment 1a:** Evaluates the base CNN (fall & non-fall) binary classifier (ActiNet).
2. **Experiment 1b:** Implements the ActiNet model using the weighted binary cross-entropy loss function, with optimized weights $w_1 = 3.6336$ and $w_0 = 0.4980$.
3. **Experiment 2:** Enhances 'Experiment 1a' by incorporating the ADM, aiming to improve fall detection by adjusting the frames of activity signals with the dominant fall class. To validate the proposed mechanism in real-world conditions, we conducted this experiment on the left-out test subject ($user_3$). The ADM received the following inputs:
   — File names of the testing activity signals,
   — Initial binary predicted labels by ActiNet,
   — The age group of the user.

   Given the lack of specific age information in the 1st dataset [Fard Moshiri, 2021], we assumed both $user_3$ (in this subsection) and $user_2$ (in Appendix 9) to be elderly. Consequently, the ADM dynamically assigned a $\tau_{\text{fall_class_percentage}}$ of 55%, consistent with the high fall likelihood threshold.

This assumption aligns with our research focus on the elderly demographic. The user's age group selection thus triggers the system to assign the corresponding threshold, effectively reflecting the varied risk profiles across different age groups.

4. **Experiment 3:** Applies transfer learning, retraining ActiNet using one instance (CSI data signal) from each activity of left-out $user_3$. Selected activity frames from S_{user_3} testing set were used for fine-tuning the 'Experiment 1a' classifier.
5. **Experiment 4:** Utilizes the 'Experiment 3' model, reapplying ADM for enhanced HFD.
6. **Experiment 5:** Replicates 'Experiment 3' steps with left-out $user_3$ instances but incorporates time-series DA techniques (previously validated in Chapter 3) on the selected activity frames for classifier retraining.
7. **Experiment 6:** Employs 'Experiment 5' model with the ADM decision mechanism, representing the comprehensive application of the HFD-Adv methodology, integrating and testing all facets of the proposed approach from preprocessing through enhanced processing to post-processing, as delineated in Subsection 2.2.

These experiments aimed to determine the optimal strategy for efficient HFD. Additionally, the study compares HFD-Adv performance against state-of-the-art models proposed in "Wi-Fi CSI-based HAR" domain of study. The evaluation metrics employed for evaluating our methodology include:

— **Weighted Accuracy:** Measures the overall effectiveness of the system in correctly identifying both fall and non-fall events, adjusting for class imbalance by giving different weights to the different classes.

— **Fall Detection Rate (FDR):** Emphasizes the system's ability to correctly identify actual fall events:

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

where TP is the number of True Positives.

— **False Alarm Rate (FAR):** Reflects the system's tendency to mistakenly identify non-fall activities as falls:

$$\text{FAR} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

where TN is the number of True Negatives.

— **Response Time:** Measures the system's prediction latency for each activity frame within the testing set, reported in milliseconds (ms).

7.2 Effectiveness, Efficiency, and Versatility of HFD-Adv

In evaluating the HFD-Adv methodology, the performance enhancement brought forth by the ActiNet fall classifier was meticulously assessed through a series of seven experiments, summarized in TABLE 4.2.

1. **Base Classifier (Experiment 1a and 1b):** 'Experiment 1a' revealed ActiNet's initial capability with a weighted accuracy of 89.43%, highlighting the robust architecture of our model. 'Experiment 1b' refined the weighted accuracy to 90.75% and uplifted the FDR to 97.12% by implementing a specialized weighted binary cross-entropy loss function. This adjustment exemplifies our strategic balance between enhancing fall detection reliability and slightly increasing the FAR, affirming our method's capability to optimize for both recall and precision in fall detection scenarios.

TABLE 4.2 – Performance of ActiNet Across Experiments

Metric \ Experiment	1a	1b	2	3	4	5	6
Weighted Accuracy (%)	89.43	90.75	91.27	95.21	95.40	97.40	**98.28**
Fall Detection Rate (%)	94.81	97.12	**100**	97.98	**100**	97.98	**100**
False Alarm Rate (%)	15.96	15.61	17.45	7.56	9.20	**3.18**	3.43
Response Time (ms)	**0.136**	**0.136**	0.1487	**0.136**	0.1487	**0.136**	0.1487

2. **Incorporating the ADM (Experiment 2):** Introducing ADM in 'Experiment 2' increased the weighted accuracy slightly to 91.27% and achieved a perfect FDR of 100%. This demonstrates the ADM's effectiveness in minimizing FNs, significantly enhancing fall identification accuracy.
3. **User-specific Retraining (Experiments 3 & 4):** Adapting to new user data through retraining in 'Experiment 3' resulted in a remarkable increase in weighted accuracy to 95.21%. 'Experiment 4', which re-applied ADM, confirmed the model's adaptability, maintaining high weighted accuracy and achieving a perfect FDR of 100%, while reducing the FAR to 9.20%.
4. **Data Augmentation in Retraining (Experiments 5 & 6):** 'Experiment 5' augmented the selected instances from the left-out user, showing significant improvements, especially in reducing the FAR to an impressive 3.18%. The culmination of our methodology in 'Experiment 6', integrating all stages from preprocessing through post-processing, achieved the highest weighted accuracy of 98.28%, along with a 100% FDR and a low FAR of 3.43%. This exemplifies the HFD-Adv methodology's comprehensive approach, proving its efficacy in offering a highly accurate, user-adaptable, and efficient HFD solution.

7.3 Benchmarking Against State-of-the-Art Models

To comprehensively assess the effectiveness and efficiency of our HFD-Adv methodology within the realm of CSI-based HFD, we employed the specific architectures of LSTM and BLSTM DL models, as delineated in Moshiri et al.'s study [Fard Moshiri, 2021]. These models were selected due to their broad applicability across diverse methodologies and their documented effectiveness in HFD contexts, as evidenced by prior research. Notably, the LSTM architecture has been extensively applied in HFD, as highlighted by Schäfer et al. [Schäfer, 2021b] and Cardenas et al. [Cardenas, 2023]. Additionally, the BLSTM model, as utilized in the work of Guo et al. [Guo, 2023], serves as a critical benchmark.

The comparative evaluation of HFD-Adv against LSTM and BLSTM models, as shown in TABLE 4.3, highlights HFD-Adv's dominance with 100% FDR, establishing a new benchmark in CSI-based HFD. While the LSTM and BLSTM models deliver respectable performances with FDR of 89.05% and 92.01% respectively, they are markedly outperformed by HFD-Adv's flawless metrics and faster response time of 0.1487 ms. This not only underscores HFD-Adv's effectiveness and efficiency in fall detection but also its practical applicability in emergency response systems. The results set HFD-Adv as a significant advancement over current state-of-the-art models.

TABLE 4.3 – Results on testing set derived from 1st and 3rd datasets for HFD-Adv vs. State-of-the-art

Fall Detector / Metric	LSTM	BLSTM	Base Classifier (ActiNet)	HFD-Adv
Weighted Accuracy (%)	87.53	88.78	90.75	**98.28**
Fall Detection Rate (%)	89.05	92.01	97.12	**100**
False Alarm Rate (%)	14.00	14.44	15.61	**3.43**
Response Time (ms)	0.307	0.609	0.136	**0.1487**

8 Conclusions

8.1 Summary of Key Findings

This chapter presents a comprehensive evaluation of the Advanced Human Fall Detection (HFD-Adv) methodology, designed to enhance the state of the art in Wi-Fi CSI-based Human Activity Recognition (HAR). The primary objective was to develop a robust, efficient, and versatile fall detection system that addresses existing limitations and enhances performance across diverse scenarios. The key findings from this study are summarized as follows:

1. **Effectiveness of HFD-Adv**:
   - **Fall Detection Rate (FDR)**: The HFD-Adv methodology achieved a perfect 100% FDR across various experiments, significantly outperforming existing methods and ensuring reliable fall detection in real-world scenarios.
   - **False Alarm Rate (FAR)**: The implementation of the Advanced Decision Mechanism (ADM) and optimized classifier models reduced the FAR to as low as 3.43%, demonstrating the system's capability to minimize false positives effectively.
2. **Efficiency Improvements**:
   - **Training Period Reduction**: The use of the LRFinder mechanism for learning rate optimization accelerated model convergence, reducing the overall training period and facilitating quicker iterations.
   - **Response Time**: The HFD-Adv system demonstrated a rapid response time of approximately 0.1487 milliseconds, ensuring timely detection and response crucial for emergency scenarios.
3. **Adaptability and Generalization**:
   - **User-Specific Adaptability**: The refined transfer learning approach enabled the model to adapt to new users and environments with limited initial data, ensuring precise fall detection across diverse conditions. Experiments showed significant improvements in weighted accuracy and FDR when adapting to new users.
   - **Data Diversity and Generalization**: Techniques such as PCA for dimensionality reduction and data augmentation improved the model's generalizability, addressing the challenge of limited data diversity and enhancing performance across different demographics and environments.
4. **Innovative Methodologies**:
   - **ActiNet Flexible CNN Classifier**: The development of ActiNet, a versatile CNN tailored for varied HAR tasks, proved to be a significant advancement. ActiNet's adaptable architecture accommodated both binary and multi-class classification tasks effectively.

— **Optimization Techniques**: The use of weighted loss function optimization via Genetic Algorithms (GAs) and learning rate optimization with the LRFinder mechanism demonstrated innovative approaches to enhancing model training and performance.

These findings underscore the potential of the HFD-Adv methodology to provide a reliable, efficient, and adaptable solution for fall detection. The implications for healthcare services and future research directions will be elaborated in the following subsections.

8.2 Benefits for Healthcare Services

The implementation of the HFD-Adv methodology offers significant benefits to healthcare services, particularly in enhancing emergency response systems for fall incidents. By leveraging Wi-Fi CSI data and advanced ML techniques, HFD-Adv provides a non-intrusive, accurate, and efficient solution for real-time fall detection. This capability is crucial for improving patient outcomes, reducing healthcare costs, and addressing the growing public health concern associated with falls among vulnerable populations such as the elderly.

8.2.1 Improving Emergency Response to Falls

Falls are a leading cause of injury and death among older adults, with significant implications for healthcare systems worldwide [World Health Organization, 2021; Centers for Disease Control and Prevention, 2023c]. Rapid detection and response to fall incidents are essential to mitigate severe injuries, reduce hospital admissions, and prevent fatalities [NHS England, 2023a; National Institute on Aging, 2022]. The HFD-Adv methodology enhances emergency response by providing:

— **Real-Time Fall Detection**: HFD-Adv's use of Wi-Fi CSI data enables continuous monitoring without the need for wearable devices, ensuring immediate detection of falls as they occur.

— **High Accuracy and Reliability**: With a 100% FDR and low FAR, HFD-Adv minimizes the risk of missed incidents and reduces unnecessary emergency interventions.

— **Non-Intrusive Monitoring**: The system operates using existing Wi-Fi infrastructure, preserving patient privacy and comfort, which is especially important for the elderly who may be reluctant to use wearable devices [Ge, 2023].

By accurately detecting falls promptly, HFD-Adv facilitates quicker medical assistance, which is critical in preventing complications such as prolonged immobilization, pressure ulcers, dehydration, hypothermia, and even death [National Institute on Aging, 2022].

8.2.2 Reducing Healthcare Costs and Burden

Falls among older adults contribute significantly to healthcare costs due to emergency services, hospitalizations, rehabilitation, and long-term care [Centers for Disease Control and Prevention, 2023c; Occupational Safety, 2022]. Implementing HFD-Adv can:

— **Decrease Hospital Admissions**: Early detection and intervention can prevent minor injuries from escalating into severe conditions requiring hospitalization.

— **Optimize Emergency Services**: Accurate fall detection reduces false alarms, ensuring emergency resources are allocated efficiently and are available for genuine emergencies.

— **Support Aging in Place**: By providing a safety net for older adults living independently, HFD-Adv enables them to remain in their homes longer, delaying or avoiding the need for assisted living facilities.

These benefits contribute to alleviating the financial strain on healthcare systems and improving the allocation of medical resources.

8.2.3 Enhancing Patient Safety and Quality of Life

The fear of falling can lead to reduced physical activity, social isolation, and decreased quality of life among older adults [National Institute on Aging, 2022]. HFD-Adv addresses these concerns by:

— **Providing Assurance and Confidence**: Knowing that falls will be detected promptly allows individuals to maintain an active lifestyle without the constant fear of being unattended after a fall.

— **Customizable Sensitivity through ADM**: The ADM adjusts fall detection sensitivity based on age groups, ensuring that high-risk individuals receive appropriate attention.

— **Non-Stigmatizing Technology**: As HFD-Adv operates invisibly within the environment, it avoids the stigma or discomfort associated with wearable fall detection devices.

By enhancing patient safety and promoting independence, HFD-Adv improves the overall well-being of individuals at risk of falls.

8.2.4 Facilitating Integration into Existing Healthcare Frameworks

The design of HFD-Adv allows for seamless integration into current healthcare services:

— **Utilization of Existing Infrastructure**: Leveraging Wi-Fi networks eliminates the need for additional hardware installations, simplifying deployment in residential and healthcare facilities [Ma, 2019].

— **Compatibility with Emergency Response Protocols**: The system can be configured to automatically notify caregivers, emergency contacts, or medical services upon detecting a fall, aligning with established response procedures.

— **Adaptability to Diverse Environments**: HFD-Adv's robust methodology ensures reliable performance across different settings and populations, enhancing its applicability in various healthcare contexts.

This ease of integration supports the broader adoption of advanced fall detection technologies within healthcare systems aiming to improve urgent and emergency care services [NHS England, 2023a].

8.2.5 Scenario: Application in Home-Based Care

Consider an elderly individual, Robert, living independently at home with a history of falls and mobility issues. The HFD-Adv system is installed within his home, continuously monitoring for fall incidents using Wi-Fi CSI data.

Scenario: One evening, Robert experiences a fall in his living room. The HFD-Adv system detects the fall immediately due to its high accuracy. Utilizing the ADM, the system recognizes Robert's age group and applies the appropriate sensitivity threshold, ensuring the fall is not overlooked.

An automatic alert is sent to Robert's designated emergency contact and healthcare provider. Emergency services are dispatched promptly, and Robert receives medical attention in a timely manner. The swift response prevents potential complications from prolonged immobilization and contributes to a positive health outcome.

This scenario demonstrates how HFD-Adv enhances emergency response in home-based care settings, providing safety and peace of mind for both the individual and their loved ones.

8.2.6 Advancing Public Health Outcomes

By addressing the significant public health issue of falls among older adults, HFD-Adv contributes to broader health outcomes:

— **Preventing Mortality and Morbidity**: Timely detection and intervention reduce the risk of severe injuries and fatalities associated with falls [World Health Organization, 2021].

— **Supporting Healthcare Policies**: The methodology aligns with initiatives aimed at improving urgent and emergency care services and reducing the burden on hospitals [NHS England, 2023a].

— **Data for Public Health Monitoring**: Aggregated data from HFD-Adv deployments can inform public health strategies by identifying trends and risk factors associated with falls.

The adoption of HFD-Adv thus plays a role in enhancing the quality of healthcare services and addressing a critical area of need in public health.

8.2.7 Summary of Benefits

The HFD-Adv methodology offers a comprehensive solution to the challenges posed by fall detection in healthcare services. By combining advanced technology with practical applicability, it enhances emergency response capabilities, reduces healthcare costs, improves patient safety, and supports public health initiatives. The system's non-intrusive nature and adaptability make it a valuable tool in the ongoing effort to provide high-quality care for vulnerable populations at risk of falls.

8.3 Future Directions

Building upon the successes of the HFD-Adv methodology, several avenues for future research can further enhance the effectiveness, adaptability, and practical deployment of advanced HFD systems using Wi-Fi CSI data.

Firstly, while the HFD-Adv methodology addresses the critical issue of fall detection with high accuracy, future work could focus on further reducing the FAR. Developing more sophisticated algorithms or incorporating contextual information could help in distinguishing falls from non-fall activities that may have similar signal patterns, thereby improving system reliability.

Secondly, enhancing the ADM to incorporate more personalized adjustments beyond age groups could be explored. Factors such as individual activity patterns, health conditions, or living environments might be integrated to tailor the sensitivity of fall detection, providing a more customized solution for each user.

Furthermore, integrating the HFD-Adv system with existing healthcare infrastructure and emergency response protocols could be investigated. Establishing seamless communication between the fall detection system and medical services would facilitate quicker interventions. Research into standardized interfaces and communication protocols would support this integration.

Lastly, conducting extensive real-world trials and longitudinal studies would be valuable in assessing the system's performance over time and in diverse settings. Such studies would provide critical feedback on user acceptance, system robustness, and the impact on patient outcomes, guiding further refinements and supporting the case for widespread adoption.

By pursuing these future directions, the capabilities of Wi-Fi CSI-based fall detection methodologies like HFD-Adv can be significantly enhanced, contributing to more effective and personalized healthcare services and improving safety and quality of life for individuals at risk of falls.

9 Appendix: Additional Validation on a Different Left-out User (user_2)

To further demonstrate the scalability and generalizability of our HFD-Adv methodology, we repeated the seven experiments described in Subsection 7.1 for a different left-out user, $user_2$. As with $user_3$, we isolated $user_2$'s data to create a user-specific train-test split S_{user_2}, ensuring that HFD-Adv was trained on the other users only. In TABLE 4.4, we present the experimental outcomes, emphasizing that the performance trends remain consistent with those observed for $user_3$.

TABLE 4.4 – Performance of ActiNet Across Experiments for Left-out $user_2$

Metric \ Experiment	1a	1b	2	3	4	5	6
Weighted Accuracy (%)	86.30	88.65	88.44	95.71	96.48	96.60	**97.22**
Fall Detection Rate (%)	73.70	81.04	77.98	93.27	**94.80**	93.58	**94.80**
False Alarm Rate (%)	1.11	3.74	1.11	1.85	1.85	**0.37**	**0.37**
Response Time (ms)	**0.2086**	**0.2086**	0.216	**0.2086**	0.216	**0.2086**	0.216

Discussion of Results:

— **Experiments 1a & 1b:** As seen in TABLE 4.2 for $user_3$, ActiNet again begins with a solid baseline (*Experiment 1a*), achieving a weighted accuracy of 86.30% and an FDR of 73.70%. Incorporating our weighted loss in *Experiment 1b* raises the FDR to 81.04% (albeit with a higher FAR) and boosts the weighted accuracy to 88.65%.

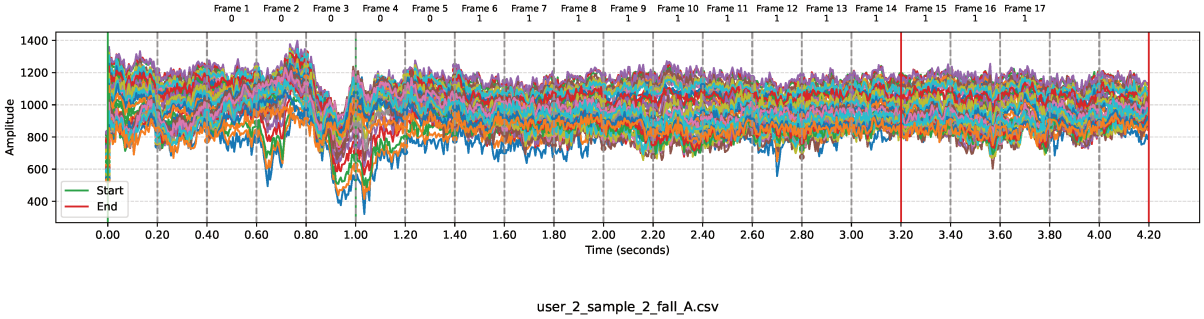
— **Experiment 2:** Building on *Experiment 1a*, the ADM further increases the FDR from 73.70% to 77.98%, while maintaining a low FAR of 1.11%. Although this final FDR remains lower than in *Experiment 2* for $user_3$, the gain for $user_2$ is still substantial, underlining the ADM's effectiveness in reducing FNs.

A visual example of this behavior is provided in FIGURE 4.6, which shows one representative trial (from the 20 fall trials of $user_2$) where ADM successfully reassigned four previously misclassified frames based on a fall-dominant ratio of 76%, exceeding the assigned threshold $\tau_{\text{fall_class_percentage}} = 55\%$. This trend is further illustrated in FIGURE 4.7

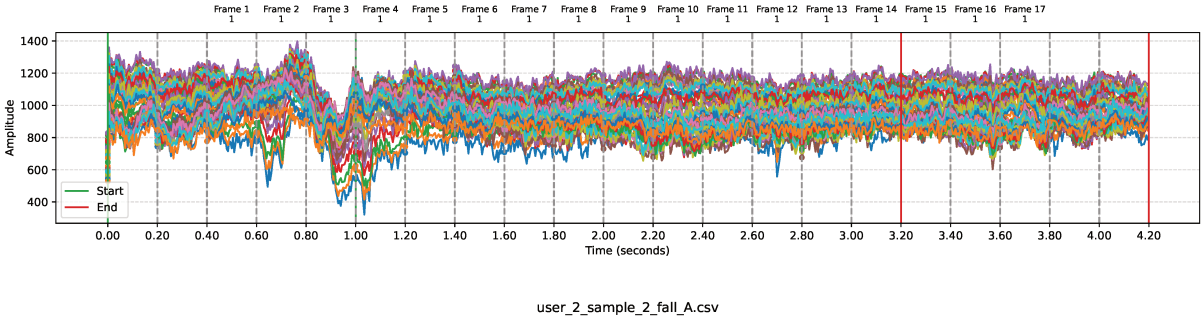
— **Experiments 3 & 4:** Fine-tuning ActiNet with one instance per activity from $user_2$ (*Experiment 3*) leads to a significant jump in weighted accuracy (from 86.30% in *Experiment 1a* to 95.71%), with the FDR rising sharply to 93.27%. When the ADM is reapplied (*Experiment 4*), both metrics improve further—weighted accuracy hits 96.48% and FDR reaches 94.80%.

— **Experiments 5 & 6:** By integrating time-series DA into transfer learning (*Experiment 5*), the FAR is further reduced to 0.37%, alongside an increase in weighted accuracy to 96.60%. Compared to *Experiment 3*, the FDR rises marginally from 93.27% to 93.58%, whereas the FAR drops from 1.85% to 0.37%—an approximate 80% decrease—showcasing how DA further enhances non-fall activity discrimination. In *Experiment 6*, integrating all stages of our methodology—from preprocessing through post-processing—culminates in our best performance for $user_2$, with a weighted accuracy of 97.22% and an FDR of 94.80%. While these metrics differ slightly from $user_3$'s results, the same overall improvement pattern is clearly observed, affirming HFD-Adv's adaptability.

As visually affirmed in FIGURE 4.8, the combined impact of personalization, DA, and ADM leads to minimal FNs and an exceptionally low FAR.

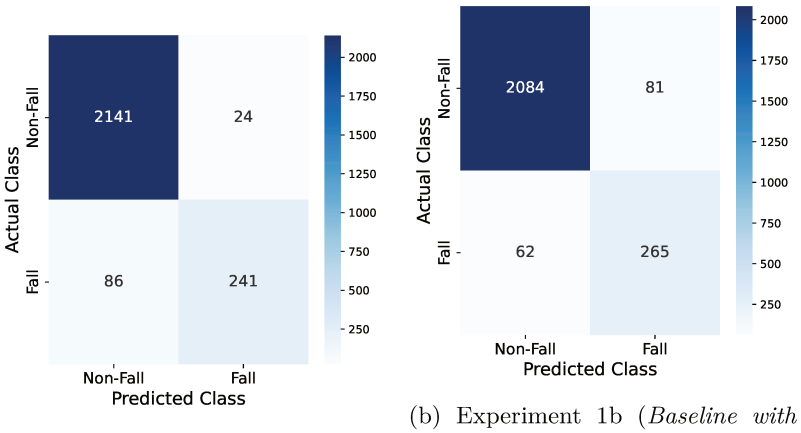


(a) Before ADM (*Experiment 1a*)



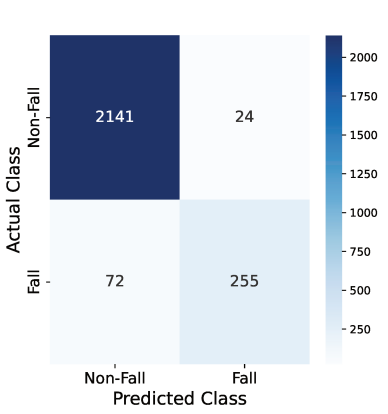
(b) After ADM (*Experiment 2*)

FIGURE 4.6 – Impact of ADM on Frame-wise Prediction for a Fall Activity Trial: This figure presents a comparison of frame-level predictions for a CSI amplitude signal (user_2_sample_2_fall_A.csv) before and after applying ADM. In (a), ActiNet misclassifies 4 out of 17 frames as non-fall. In (b), ADM corrects the labels by reclassifying all frames as fall, based on a dominant fall class ratio of approximately 76%, exceeding the threshold $\tau_{\text{fall_class_percentage}} = 55\%$.



(a) Experiment 1a (*Baseline*)

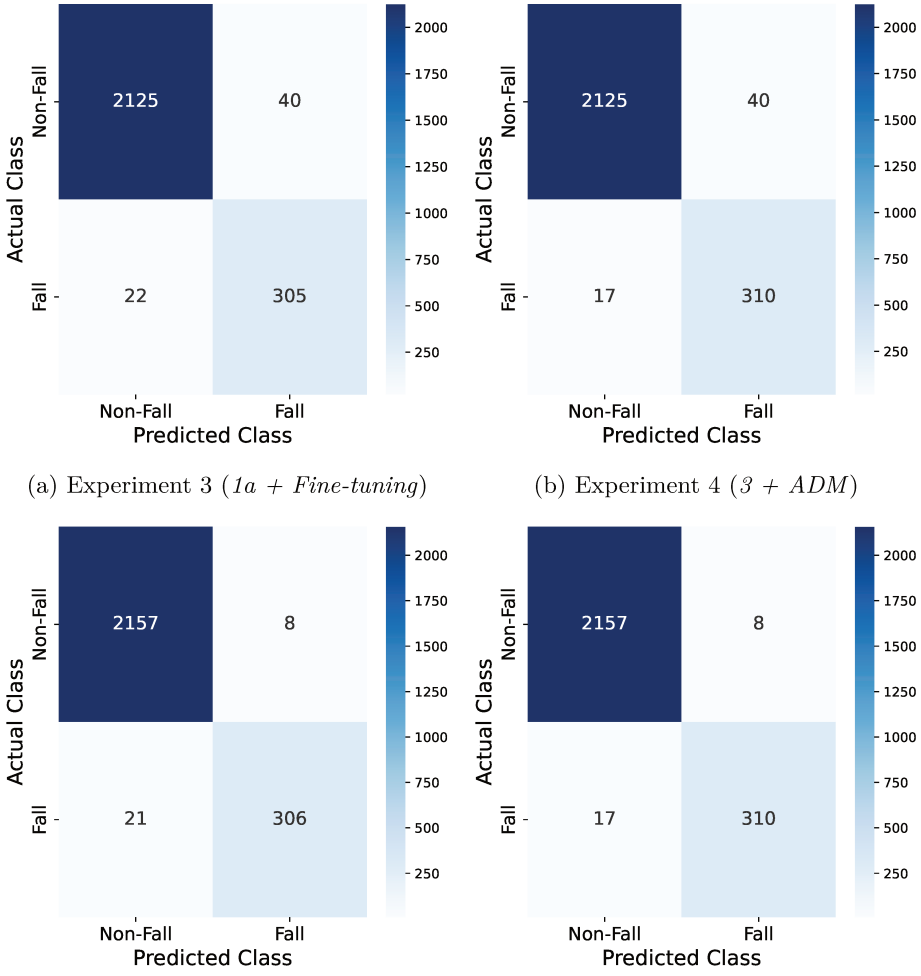
(b) Experiment 1b (*Baseline with Weighted Loss*)



(c) Experiment 2 (*1a + ADM*)

FIGURE 4.7 – Confusion Matrices for Experiments 1a–2: This figure displays the confusion matrices for the initial experiments, illustrating the progression from the base classifier (a), through weighted loss optimization (b), to the integration of ADM (c). Each matrix shows how specific enhancements influence fall detection accuracy.

These results reinforce that HFD-Adv's core improvements—weighted loss function optimization via Genetic Algorithms, refined transfer learning with data augmentation, and ADM post-processing—are not tied to a single user. Instead, they extend robustly to other unseen users (such as $user_2$), thereby confirming HFD-Adv's reliable generalization and scalability.



(a) Experiment 3 (*1a + Fine-tuning*)

(b) Experiment 4 (*3 + ADM*)

(c) Experiment 5 (*1a + Fine-tuning + DA*)

(d) Experiment 6 (*Full HFD-Adv*)

FIGURE 4.8 – Confusion Matrices for Experiments 3–6: This figure showcases the confusion matrices for later experiments that incorporate personalization through fine-tuning (a), ADM post-processing (b), time-series DA (c), and the complete HFD-Adv pipeline (d). Together, they demonstrate the compounding improvements in fall detection performance.

Chapter 5

CSI-Based Human Identity Recognition Approach Using a Transformer-Based Model for Personalized Healthcare

Objectives

This chapter presents *IdentNet*, a Transformer-based model for CSI-based identity recognition during an activity. By addressing intra-class similarity and long-term dependencies, the methodology ensures accurate and non-intrusive user identification, contributing to personalized healthcare monitoring and emergency response.

Chapter Outline

1 Introduction . . . **143**
1.1 Background and Importance of HIDR in HAR for Personalized Healthcare . . . 143
1.2 Defining the Problem Space and Its Challenges . . . 143
1.3 Literature Review on CSI-Based Human Identity Recognition . . . 145
1.4 Contributions . . . 147
1.5 Organization of the Chapter . . . 148
2 Methodology Overview . . . **149**
2.1 Introduction to the Methodology . . . 149
2.2 Objectives and Design Principles . . . 149
3 User CSI Data Preprocessing for User Identification . . . **151**
4 IdentNet Transformer User Identifier . . . **152**
4.1 Theoretical Background on Transformers and Their Use in Multidimensional Time-Series Analysis 152
4.2 Architecture of the Identifier . . . 153
5 Results and Evaluation . . . **156**
5.1 Evaluation Metrics and Benchmarking . . . 156
5.2 Effectiveness and Efficiency of IdentNet . . . 157
6 Conclusions . . . **161**
6.1 Summary of Key Findings . . . 161
6.2 Benefits for Healthcare Services . . . 162
6.3 Future Directions . . . 165
7 Appendix: Extending IdentNet to Other User Activities . . . **166**
7.1 Introduction and Experimental Hypotheses . . . 166
7.2 User CSI Data Preprocessing . . . 166
7.3 Experiments and Results . . . 167

7.4 Conclusion and Outlook 169

1 Introduction

1.1 Background and Importance of HIDR in HAR for Personalized Healthcare

The increasing integration of technology into healthcare systems underscores the vital role of Human Identity Recognition (HIDR) in delivering effective and personalized patient care. HIDR, which involves identifying individuals based on specific biometric or behavioral patterns, facilitates tailored healthcare interventions that meet each patient's unique needs. This approach is essential in environments where patient monitoring and data security are crucial, such as in-home care, hospitals, and rehabilitation centers. One promising technology for HIDR is the use of Channel State Information (CSI) data. CSI data captures the characteristics of wireless signal propagation between a transmitter and a receiver, offering a non-invasive and continuous method of monitoring individuals by analyzing variations in signal patterns caused by their movements and activities. Unlike traditional biometric systems, such as fingerprinting or facial recognition, CSI-based identification operates without direct contact or visible observation, providing a seamless and unobtrusive user experience [Chawla, 2013 ; Trusts, 2022 ; Duan, 2023].

CSI data is particularly effective for user-specific identification because it can detect subtle variations in human movement and behavior unique to each individual. These variations, captured as signal amplitude and phase shifts, enable the creation of distinctive user profiles for continuous monitoring and identification. In healthcare settings, CSI can track a patient's gait, posture, and daily activities, offering valuable insights into their health status and potential risks. For instance, identifying individuals based on their walking patterns has shown high accuracy, indicating the potential for reliable user identification in personalized patient monitoring systems [Wang, 2014 ; Custance, 2023 ; Deng, 2023 ; Tan, 2018 ; Khan, 2021]. The shift towards personalized healthcare has created an increasing demand for user-specific applications that adapt to the diverse needs of patients. Traditional healthcare models often rely on generalized approaches to treatment and care, which may not adequately address the unique circumstances of each patient. HIDR, powered by CSI data, addresses this gap by enabling healthcare providers to develop personalized assistance strategies that consider individual characteristics and behaviors.

User-specific applications in healthcare can lead to more effective interventions and improved patient outcomes. By accurately identifying patients and monitoring their activities, healthcare systems can detect early signs of health deterioration or emergencies, such as falls, and respond promptly with appropriate interventions [Wilson, 2023 ; Institute, 2016]. This personalized approach enhances patient safety and improves the overall quality of care by ensuring healthcare resources are used efficiently and effectively [Minvielle, 2021 ; Vicente, 2020 ; Chawla, 2013]. By leveraging CSI data, healthcare providers can continuously monitor patients without the need for intrusive wearable devices, reducing patient discomfort and increasing compliance. This continuous monitoring capability allows for early detection of health issues and timely interventions, enhancing patient outcomes and satisfaction.

Moreover, HIDR systems contribute to data security and privacy in healthcare settings. By accurately identifying and authenticating individuals, these systems help prevent unauthorized access to sensitive patient information, thereby safeguarding patient privacy and maintaining trust in healthcare services [Tan, 2018 ; Duan, 2023]. As research and technology continue to advance, the potential applications and benefits of HIDR in healthcare are expected to expand, offering new avenues for innovation and improvement in patient care.

1.2 Defining the Problem Space and Its Challenges

HIDR using CSI amplitude data for fall activity presents a significant opportunity to enhance personalized healthcare, particularly in environments where continuous and non-intrusive monitoring is crucial.

The specific problem addressed in this study involves accurately identifying individuals based solely on the subtle, user-specific patterns generated by their falling activity, as captured by CSI data. This approach leverages the sensitivity of CSI to detect nuanced variations in human movement, making it a promising method for personalized patient monitoring and intervention, especially in emergency scenarios like falls.

In the context of identity recognition during falls, two primary approaches can be considered: biometric traits and behavioral patterns. Biometric traits, such as fingerprints and facial features, are inherently static and unique to each individual. However, they require direct contact or clear visual capture, which may not be practical in healthcare settings, particularly during rapid-response scenarios like falls [Duan, 2023].

Behavioral patterns, particularly those exhibited during fall activity, offer a dynamic and non-intrusive means of identifying individuals. These patterns are valuable in healthcare scenarios where the response can be tailored based on the individual's identity. For example, knowing whether the person who has fallen is a child, an adult, or an elderly member can guide the urgency and type of intervention required [World Health Organization, 2021 ; Centers for Disease Control and Prevention, 2023a]. Behavioral patterns can be captured without requiring active user participation or specialized hardware, making them more suitable for real-time monitoring and emergency response [Duan, 2023].

Among the various technologies available for capturing behavioral patterns, such as vision-based systems (cameras) and sensor-based systems (wearables), CSI stands out for several reasons. Vision-based systems, while capable of providing detailed visual insights, face privacy concerns, the need for clear line-of-sight, and sensitivity to lighting conditions. These limitations are problematic in home environments, where continuous monitoring is necessary, but invasive surveillance is undesirable. Similarly, sensor-based systems require the user to wear devices, which can be intrusive and uncomfortable, potentially leading to non-compliance [Duan, 2023].

CSI technology, on the other hand, offers a non-intrusive, privacy-preserving alternative that leverages existing Wi-Fi infrastructure to capture detailed information about a person's movements, including falls. Unlike cameras, CSI does not require visual contact, and unlike wearable sensors, it does not depend on the user's cooperation [Duan, 2023]. The ability to identify who is falling in real-time allows for immediate, personalized responses, which are critical in situations where the speed and type of intervention can significantly affect outcomes.

However, there are several challenges to overcome when using CSI technology for identity recognition. One primary challenge is the intra-class similarity within the fall activity data. Different users may exhibit highly similar fall patterns, making it difficult to distinguish between individuals based on these movements alone. This task is inherently more complex than traditional activity classification, as the model must capture subtle, user-specific variations within the same type of activity, namely falling [Wang, 2021b]. These subtle differences may be masked by noise or other variations in the environment, complicating the recognition task.

In earlier chapters, we used Convolutional Neural Networks (CNNs) for tasks like activity recognition and fall detection. However, for this specific task of identity recognition based on fall activity, a Transformer-based architecture is more suitable. CNNs, while effective at capturing local dependencies through convolutional filters, often struggle to model long-range dependencies across the time series data, which is crucial in identifying subtle, user-specific fall patterns. Transformers, by contrast, use self-attention mechanisms to model global relationships within the sequence of CSI data. This allows the model to capture these fine-grained variations and dependencies across the entire sequence, making it more capable of distinguishing between users based on their falling activity [Vaswani, 2017 ; Zerveas, 2021 ; Bai, 2018 ; Goodfellow, 2016].

Moreover, Transformers are flexible in handling sequences of varying lengths, which is beneficial in real-

world scenarios where the number of CSI frames may differ across falls. Their ability to capture global temporal and spectral dependencies makes them ideal for this task, where the model must generalize across users despite potential variations in signal length and noise.

In summary, the problem of identifying individuals through their fall activity using CSI amplitude data involves challenges related to intra-class similarity. The decision to use a Transformer-based architecture is driven by the need to capture the global relationships and fine-grained patterns present in the CSI data. By leveraging transformers, our approach addresses these challenges, paving the way for more accurate and reliable identity recognition, particularly in emergency response scenarios where knowing the identity of the person falling can lead to more effective and personalized interventions.

1.3 Literature Review on CSI-Based Human Identity Recognition

HIDR using CSI has emerged as a pivotal approach due to its non-intrusive nature, high accuracy, and applicability in diverse environments. This section provides a detailed review of existing literature on CSI-based HIDR, discussing various approaches, techniques, and models, while highlighting the limitations that necessitate novel solutions, such as the Transformer-based method proposed in this chapter.

1.3.1 CSI-Based Human Identity Recognition Approaches

The use of Wi-Fi CSI for identity recognition has evolved as a promising alternative to traditional biometric and sensor-based methods, which often face challenges related to privacy, environmental dependencies, and the need for physical contact or line-of-sight. Early works in this domain, such as WiWho [Zeng, 2016] and WiFi-ID [Zhang, 2016], leveraged CSI data to distinguish between individuals based on their gait patterns. These approaches typically involved capturing CSI data through off-the-shelf Wi-Fi devices and applying signal processing techniques to extract features representative of individual movements. For instance, WiWho utilized peak-valley detection algorithms, while WiFi-ID employed Sparse Approximate Classification (SAC) methods, achieving considerable success in controlled environments [Bian, 2024].

Although these approaches were effective, they suffered from certain limitations, such as reliance on earlier hardware with fewer subcarriers, which constrained the system's sensing capabilities. Moreover, these methods struggled with the generalization of the models when faced with data imbalance and environmental variability, issues prevalent in real-world scenarios. For example, Bian et al. (2024) [Bian, 2024] tackled this by introducing Conditional Denoising Diffusion Probabilistic Models (C-DDPMs) for generating synthetic CSI data, improving identification accuracy, especially in underrepresented classes.

As technology advanced, the potential of deep Learning (DL) for HIDR became evident. The introduction of CNNs marked a significant shift in how CSI data was processed, with CNNs being particularly adept at extracting local spectral and temporal features from CSI signals. CNNs have been effective in capturing these features but consistently face limitations when it comes to handling long-range dependencies within the CSI data, which are crucial for distinguishing users based on subtle variations in activity. This is especially relevant in tasks where the model must capture subtle, user-specific nuances within similar activities, such as identifying individuals based on their fall patterns [Bai, 2018 ; Goodfellow, 2016].

1.3.2 Techniques for Robust CSI-Based Identity Recognition

The robustness of CSI-based HIDR systems is essential, particularly in dynamic environments where variations in user behavior and environmental conditions can affect the system's performance. Recent works have focused on improving the generalizability and adaptability of these systems. For instance,

Gu et al. (2021) [Gu, 2021] introduced WiONE, a system designed to enhance the sensitivity of CSI to subtle human behaviors like keystrokes using one-shot learning and Rician fading. However, this system still relied on CNN-based models, which face limitations in recognizing long-range dependencies crucial for distinguishing users in more complex environments.

Lin et al. (2023) [Lin, 2023] developed WiTL, combining ResNet and adversarial networks to extract environment-independent features from CSI data. This system achieved impressive accuracy, yet the reliance on CNNs meant it was still constrained by the model's ability to capture only local dependencies. Similarly, LW-WiID, proposed by Cao et al. (2021) [Cao, 2021], employed a lightweight CNN-based architecture designed to reduce model complexity. Although efficient for IoT devices, its ability to capture long-range dependencies was limited, making it less suitable for complex tasks like user identification based on falling activity, where subtle user-specific patterns need to be captured across the entire CSI sequence.

Notably, Transformer-based models have begun to emerge as a more suitable solution for tasks requiring the capture of global relationships in multidimensional time series data like CSI. The SimpleViTFi model introduced by Bian et al. (2024) [Bian, 2024] leveraged a Vision Transformer (ViT) architecture to improve identity recognition performance in imbalanced datasets. By incorporating the self-attention mechanism, this approach effectively modeled long-range dependencies in CSI data, addressing one of the key limitations of CNNs. Although the application of Transformers is relatively new in CSI-based identity recognition, their potential to significantly enhance accuracy, particularly in tasks involving subtle behavioral patterns like falling activity, is promising.

1.3.3 Applications of CSI-Based Identity Recognition

The applications of CSI-based HIDR extend beyond traditional security and surveillance. In vehicle security, for example, Zhang et al. (2022) [Zhang, 2022] introduced CarAu, a system that used CSI data to authenticate car key owners and identify individual users within a group. Similarly, Mo and Kim (2021) [Mo, 2021] focused on human identification in indoor environments, employing data augmentation (DA) techniques such as sliding windows and time warping to enhance the performance of CNN-based systems. While these approaches demonstrated high accuracy, they still struggled with the limitations of CNNs in handling long-range temporal dependencies, which are crucial in more complex tasks like user identification through falling activity.

Given the need for improved generalization and accuracy in such scenarios, Transformer-based architectures offer a promising avenue for future research. With their ability to model both local and global dependencies, Transformers hold significant potential for enhancing user identification across diverse applications, particularly those involving sequential data like falling activity.

1.3.4 Gaps and Limitations in Existing Methods

Despite the significant advancements in CSI-based HIDR, several limitations persist. Many existing systems, particularly those based on CNNs, are constrained by their inability to capture long-range temporal dependencies within CSI data. This limitation is especially problematic in complex tasks like fall-based user identification, where subtle differences between users may be distributed across the entire sequence. Moreover, systems like WiTL and CAUTION, while effective in controlled environments, still face challenges when dealing with diverse and overlapping human activities in real-world settings.

Additionally, the reliance on large amounts of labeled data is another challenge. DA techniques like sliding windows and time warping, proposed by Mo and Kim (2021), and the use of generative models such as C-DDPMs by Bian et al. (2024), have been effective in mitigating data sparsity and imbalance. However, these methods often require significant computational resources and may introduce biases if

not carefully implemented. Furthermore, CNN-based methods often struggle with generalizing to new environments, leading to reduced performance when faced with unseen scenarios or activities.

1.3.5 Positioning the Current Work

The approach presented in this chapter builds on the strengths of existing CSI-based HIDR systems while addressing their limitations. By transitioning from a CNN-based architecture to a Transformer-based model, this study aims to enhance the accuracy and robustness of user identification, particularly in tasks that require capturing long-range dependencies in CSI data. Unlike previous studies that primarily focused on static environments or specific activities, this approach is tailored for identity recognition based on falling activity, which are critical in personalized healthcare, but the model's flexibility allows it to be adapted to other human activities if needed.

To our knowledge, this study is the first to utilize CSI data specifically from falling activities to identify users, marking a significant advancement in the field. The novel application of a Transformer-based model allows for capturing both local and global dependencies, offering a significant improvement over traditional CNN-based methods.

In conclusion, while significant progress has been made in CSI-based HIDR, the challenges of environmental variability, data scarcity, and generalizability remain. The approach proposed in this chapter leverages the strengths of Transformers to address these gaps, offering a robust and scalable solution that holds promise for a wide range of applications, particularly in personalized healthcare.

1.4 Contributions

This chapter outlines the significant contributions of this research in the domain of Wi-Fi CSI-based HIDR. We introduce a novel Transformer-based model, IdentNet, which significantly enhances the state of the art in this domain. Our contributions are categorized into four primary areas:

1. **Introducing Novel Methodologies**
   - **Development of IdentNet Transformer-Based Model for HIDR**: A novel Transformer-based architecture named IdentNet was developed specifically for HIDR using CSI data during falling activity.
   - **Adaptability of the Methodology**: The proposed methodology is adaptable to other activities beyond falling activity. The Transformer user identifier can be retrained to identify users across various activities or applied in various settings, making it versatile for broader applications in HIDR and HAR.
2. **Addressing Key Limitations in CSI-Based HIDR**
   - **Addressing Intra-Class Similarity**: We tackled the challenge of intra-class similarity in falling activity data by leveraging the Transformer's self-attention mechanism. This allows the model to capture subtle, user-specific variations in CSI amplitude data that traditional CNNs might overlook.
   - **Handling Long-Term Dependencies**: We overcame the difficulty of modeling long-term dependencies in multidimensional time series data. The Transformer's ability to capture global temporal and spectral relationships enhances the model's capacity to distinguish between users based on their unique fall patterns.
3. **Enhancing Performance**
   - **Improved Performance over CNN Models**: We demonstrated that the Transformer-based IdentNet outperforms a CNN-based classifier in identifying users during falls.

— **Benchmarking Against State-of-the-Art Models**: We provided benchmarking results showing that IdentNet surpasses existing state-of-the-art models in CSI-based HIDR. This underscores the model's superior performance and its potential for real-world applications in personalized healthcare.

4. **Advancement in Personalized Healthcare**
   — **Facilitating Tailored Interventions**: By enabling accurate and non-intrusive user identification during critical events like falls, we contribute to the advancement of personalized healthcare services. This facilitates tailored and timely interventions, potentially improving emergency response outcomes.
   — **Utilization of CSI Data from Falling Activity**: To our knowledge, our work is the first to utilize CSI data specifically from falling activity to identify users. This marks a significant advancement, as previous studies had not focused on using fall activity data for HIDR.

These contributions collectively advance the field of CSI-based HIDR by addressing key challenges and introducing a novel, adaptable methodology. By leveraging the strengths of Transformer architectures, our work enhances the accuracy and reliability of user identification in healthcare patient monitoring systems. This not only improves the potential for personalized patient care but also sets new benchmarks for performance in HIDR applications.

1.5 Organization of the Chapter

The remainder of this fifth chapter (third chapter of the contributions part II) is organized as follows:

— Section 2 sets the stage for our Transformer-based HIDR approach, outlining the motivation behind user identification during fall activity and the key objectives and design principles that guide this methodology.

— Section 3 details the data preparation and preprocessing techniques tailored to user identification, focusing on how raw CSI signals are reorganized for the new classification task and labeled into "no_user" and user-specific classes.

— Section 4 introduces *IdentNet*, the proposed Transformer-based user identifier, and explains how self-attention mechanisms capture long-range dependencies and subtle inter-user variations in CSI data.

— Section 5 presents comprehensive performance evaluations of IdentNet. It compares effectiveness and efficiency metrics against both a prior CNN-based approach (developed and utilized in Chapters 3 and 4) and other state-of-the-art models, highlighting how the Transformer architecture excels in CSI-based user identification task.

— Section 6 concludes the chapter by summarizing the main findings, discussing the benefits of integrating HIDR into personalized healthcare services, and suggesting directions for future work that build on this Transformer-based methodology.

— Appendix 7 provides an additional set of experiments demonstrating how *IdentNet* can be adapted to a different daily activity (*running*). This extension showcases the methodology's potential for broader user identification tasks, especially when fall data is scarce or difficult to collect.

By the end of this final contribution chapter, we extend our CSI-based HAR research to identity recognition for personalized patient monitoring and emergency response services, completing the set of original methodologies proposed in this thesis. The following General Conclusion and Perspectives will synthesize our contributions and present overarching perspectives on the future of Wi-Fi CSI-based HAR in healthcare.

2 Methodology Overview

2.1 Introduction to the Methodology

HIDR using CSI data in healthcare has evolved significantly, particularly for personalized, non-intrusive patient monitoring. Building on these advances, the methodology proposed in this chapter focuses on identifying users during falls, an essential application in healthcare that enables tailored interventions and minimizes injury risks in emergencies [Duan, 2023 ; Tan, 2018 ; Vicente, 2020].

Unlike the previous chapters, which used CNNs for activity recognition and fall detection, our methodology adopts a Transformer-based approach to address user-specific fall activity patterns. By modeling long-range dependencies more effectively, Transformers capture subtle, individual nuances within multi-dimensional CSI data that are crucial for accurate HIDR.

The proposed Transformer-based model, denoted as IdentNet, employs a self-attention mechanism to capture relationships between time steps and subcarriers in CSI data, enabling it to distinguish between users based on their unique fall patterns. This capability is critical for overcoming the intra-class similarity challenge, where users' falling activity may exhibit overlapping characteristics that complicate recognition tasks. By focusing on global relationships within the entire CSI sequence, IdentNet excels at identifying the fine-grained variations needed for accurate user identification [Vaswani, 2017 ; Zerveas, 2021].

The methodology is structured into two main phases: data preparation and model training. During the data preparation phase, raw CSI data is processed and structured into activity frames for each user, ensuring that the data preserves the natural fluctuations inherent in wireless signal measurements. The transformer model is particularly well-suited for this task, as it can leverage the structured nature of the data to capture long-range dependencies across the CSI sequence.

In the model training phase, the IdentNet transformer model is trained using real CSI data. The goal is to enhance the model's ability to capture the global dependencies across time steps and subcarriers, ensuring robust identification of users even in challenging environments. By leveraging transformers' strengths in handling long-term dependencies, this methodology addresses the limitations of traditional CNN models, particularly in distinguishing subtle user-specific patterns during falling activity.

In summary, the methodology introduces a novel Transformer-based approach to HIDR using CSI data, specifically designed for user identification during falls. The self-attention mechanism within IdentNet enables the model to overcome challenges like intra-class similarity and long-term dependency handling, which are critical for accurate identification. This methodology represents a significant advancement in personalized healthcare monitoring, offering a scalable and adaptable solution for continuous patient monitoring based on falling CSI amplitude data.

2.2 Objectives and Design Principles

Throughout this chapter, we focus on tackling limitations specific to HIDR using CSI data. Although HIDR shares certain limitations with the broader domain of Wi-Fi CSI-based Human Activity Recognition (HAR), it introduces new challenges due to the complexity of identifying users based on subtle variations in activity signals. These challenges are particularly significant in the context of falling activity, where intra-class similarity and the need to capture user-specific patterns in the data are crucial.

The methodology is structured into two main phases: data preparation and model training. These phases are designed to tackle the specific challenges in CSI-based HIDR, ensuring the model has sufficient capacity to specialize in identifying individual users based on their unique falling activity patterns. The methodology addresses two key limitations within the broader domain of Wi-Fi CSI-based HAR and the specific challenges of Wi-Fi CSI-based HIDR:

— **Intra-class Similarity:** Intra-class similarity is a unique challenge in HIDR, where users exhibit highly similar fall patterns, making it difficult to distinguish between them. The proposed methodology addresses this limitation by employing a Transformer-based architecture, which excels at capturing subtle, user-specific variations across entire sequences. Through its self-attention mechanism, the model identifies global relationships within the CSI data, allowing it to focus on fine-grained differences between users' fall patterns [Vaswani, 2017 ; Zerveas, 2021]. This ability to model long-range dependencies helps distinguish between users, overcoming the challenge posed by intra-class similarity [Lv, 2020 ; Imanzadeh, 2024].

— **Difficulty in Handling Long-term Dependencies:** A common limitation in both HAR and HIDR is the difficulty in handling long-term dependencies in time series data. Traditional models, such as CNNs, often struggle to capture these dependencies across time, which is critical for accurately identifying users based on their falling activity. The Transformer architecture in this methodology is specifically designed to handle long-range temporal and spectral dependencies effectively. By leveraging the self-attention mechanism, the model can attend to both local and global features across time steps and subcarriers, making it highly suitable for tasks that require capturing complex temporal relationships in CSI data [Bai, 2018 ; Goodfellow, 2016]. This addresses the limitation of long-term dependency handling, ensuring the model can extract relevant features from fall activity signals to accurately identify users.

The design of the proposed methodology follows key principles to ensure robustness and adaptability for the task of identifying users based on fall activity:

— **Data Preparation:** The data preparation phase involves preprocessing the raw CSI amplitude data, structuring the activity frames for each user. This phase ensures that the CSI data is sufficiently structured, allowing the transformer model to specialize in identifying subtle variations in users' falling activity patterns.

— **Model Training:** The model training phase focuses on training the transformer model (IdentNet) using real CSI data. The transformer's self-attention mechanism is particularly adept at modeling long-term dependencies and capturing both temporal and spectral relationships within the CSI data. The training process includes fine-tuning the model's hyperparameters and applying regularization techniques to avoid overfitting, ensuring the model specializes in distinguishing between individual users based on their falling activity.

— **Global and Local Feature Extraction:** The self-attention mechanism within the Transformer architecture is designed to capture both global and local dependencies in the CSI data, ensuring that the model can effectively distinguish between users based on subtle variations in their fall patterns. This principle is critical for overcoming the challenge of intra-class similarity, allowing the model to focus on fine-grained differences between users' falling activity that might otherwise be missed by traditional models.

— **Adaptability:** While this methodology specifically focuses on identifying users based on falling activity, its Transformer-based architecture is highly adaptable to other daily activities and different environments if needed. Within the same environment, the model can be retrained to identify users across various activities (such as walking, standing up, or bending) using the same dataset. Furthermore, the methodology is versatile enough to be applied to multiple activities simultaneously, allowing it to classify users across several human activities at once. In other settings, with different users and activities, the methodology can be re-applied by training a new model on the CSI data specific to that environment. This flexibility makes the methodology suitable for broader applications in HAR, where it can be tailored to detect and identify users based on any chosen activity or set of activities across different scenarios, thereby extending its utility beyond the fall activity

context. In Appendix 7, we illustrate this adaptability by applying *IdentNet* to the running activity and showing how a small amount of fall data can be used to fine-tune the model for user identification in more critical scenarios.

In summary, the proposed methodology addresses key limitations in CSI-based HIDR, including intra-class similarity and long-term dependency handling, through the use of a Transformer-based architecture. By adhering to these design principles, the methodology ensures robust and accurate user identification during falling activity, contributing to the advancement of personalized healthcare monitoring.

3 User CSI Data Preprocessing for User Identification

In this chapter, we restructure the "First CSI Dataset" (described in Chapter 1) from the previous contributions to support the new task of user identification based on falling activity. This dataset, originally used for activity recognition and fall detection tasks in Chapters 2 and 3, is now adapted to isolate the falling activity for each user, enabling us to train a model for user identification.

The preprocessing steps follow similar procedures to those used in Chapter 3, particularly regarding the handling of incomplete frames and the integration of 'empty' (non-user) activity CSI amplitude data from the "Third CSI Dataset" (introduced in Chapter 1). However, the datasets are now specifically organized to support user identification based on falling activity, the new classification task.

After restructuring the datasets, we now have four classes for our identity recognition task: "no_user," "user_1," "user_2," and "user_3," representing the falling activity of three distinct users and the absence of any user (i.e., the "no_user" class).

Before preparing the training set, we applied an initial train-test split to ensure an unbiased evaluation on unseen data. The dataset was divided into two balanced subsets, allocating 85% of the data for training and 15% for testing. Out of 107 total instances (where each instance is a CSI amplitude data signal of either falling or empty activity recorded as a CSV file), 91 instances were assigned to the training set, while 16 were reserved for testing. This early split prevents data leakage, ensuring that no information from the testing data influences the training phase, a common issue when splitting is done after framing.

Following the train-test split, we performed overlapped framing on both the training and testing sets. This process extracts overlapping frames from the signals, with each frame consisting of 200 samples (corresponding to a 1-second snapshot of the recorded activity). A hop length of 40 samples was used to create the overlapping frames, capturing the temporal dynamics of the CSI signals.

Incomplete frames, which typically occur at the end of the signals, were handled using the same threshold-based padding strategy detailed in Chapter 3. Frames with 60% or more real samples were padded to the full frame size using the mean value of the signal, preserving signal integrity without introducing significant distortions that might affect classification.

Additionally, we applied PCA to reduce the number of subcarriers in the "no_user" data, as introduced from the 3rd dataset, from 256 to 52 subcarriers, matching the dimensionality of the primary dataset. This ensures that the "no_user" data is compatible with the data for the three users, allowing for a consistent analysis across all four classes.

We labeled the frames for each user using one-hot encoding, allowing the model to differentiate between the three users and the "no_user" condition. The one-hot encoded labels for the four classes were as follows:

- **no_user**: `[1, 0, 0, 0]`,
- **user_1**: `[0, 1, 0, 0]`,
- **user_2**: `[0, 0, 1, 0]`,
- **user_3**: `[0, 0, 0, 1]`.

After the framing and labeling processes, the training set contained $1,173$ activity frames, each of size $(200, 52)$, while the testing set comprised 216 frames of the same size.

The distribution of classes in the test data shows a slight class imbalance, as visualized in FIGURE 5.1. This imbalance was noted for its potential impact on model evaluation.

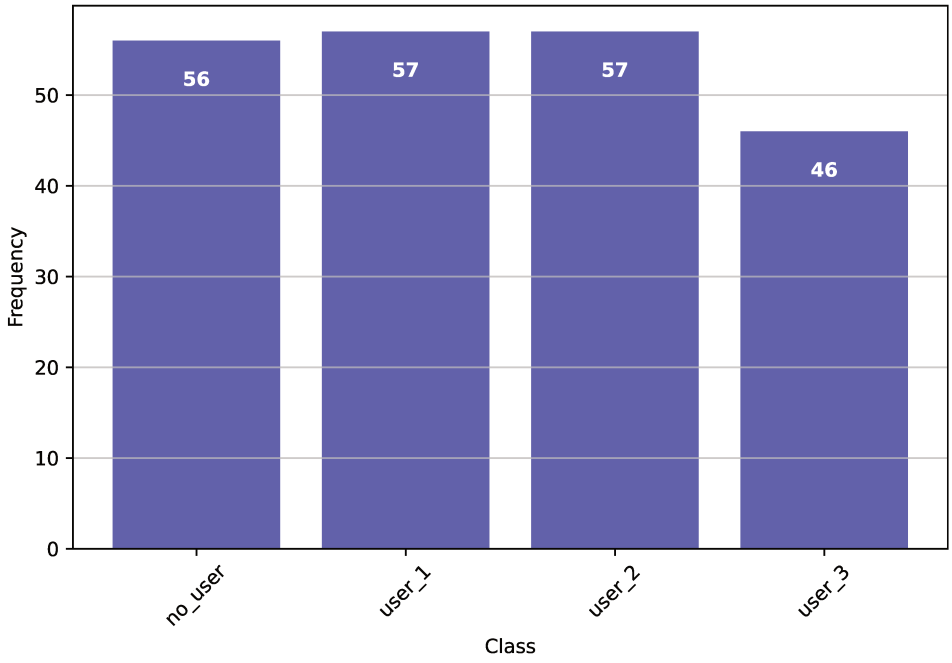


FIGURE 5.1 – Class Distribution in test data array: This figure shows the class distribution in the test data array, highlighting a slight imbalance across the four classes.

Following these preprocessing steps, we further split the training set into training and validation subsets using the holdout method, as detailed in Chapter 2. This stratified split ensures that each class is proportionally represented in both the training and validation sets. The final division allocates 85% of the training data for model fitting and 15% for validation. The final dataset shapes are as follows:

— **Training set**: 997 frames of size $(200, 52)$

— **Validation set**: 176 frames of size $(200, 52)$

This structured dataset is essential for training the Transformer-based model for user identification, ensuring that it can accurately distinguish between a small set of known users based on their falling activity while maintaining an unbiased evaluation through the use of an early train-test split.

4 IdentNet Transformer User Identifier

4.1 Theoretical Background on Transformers and Their Use in Multidimensional Time-Series Analysis

Transformers were first introduced by [Vaswani, 2017], offering a new architecture that effectively handles sequence-to-sequence tasks. Initially developed for natural language processing (NLP) tasks, Transformers have demonstrated their ability to outperform traditional sequence models like *Recurrent Neural Networks* (RNNs) and Long Short-Term Memory networks (LSTMs). The key innovation in Transformers is the self-attention mechanism, which enables the model to capture relationships across an entire

sequence without processing the data step-by-step, as RNNs do. This approach has proven highly effective in modeling long-range dependencies within data sequences, making Transformers applicable to a wide range of tasks beyond NLP.

At the core of the Transformer architecture is the self-attention mechanism, which allows each input element to attend to all others within the sequence. The attention mechanism computes the relationships between input tokens (or data points) as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V$$

Where Q (query), K (key), and V (value) are learned linear transformations of the input data, and d_k is the dimensionality of the key vectors. The self-attention mechanism assigns different levels of importance to each token in relation to the others, enabling the model to capture global dependencies within the data.

The Transformer architecture also uses multi-head attention, where several self-attention mechanisms are applied in parallel. This allows the model to capture diverse relationships in the data, enhancing its ability to model complex dependencies across multiple dimensions.

Transformers' strength lies in their ability to model global dependencies across entire sequences. In contrast to RNNs, which process data sequentially and can struggle with long-range dependencies, Transformers process all input tokens in parallel. This parallelization accelerates training, making Transformers scalable to larger datasets. Additionally, the ability to compute relationships between distant elements within the sequence allows them to capture intricate patterns, which is particularly advantageous for tasks involving multidimensional time series data.

Transformers have emerged as highly effective in analyzing multidimensional time series data, where complex dependencies exist between different dimensions and time steps. The self-attention mechanism's ability to model long-range temporal and spectral dependencies allows it to capture the global structure of such data. This is essential when working with tasks like HIDR, where subtle variations across time and different features can differentiate between users based on their unique activity patterns.

In our study, Transformers' ability to capture long-range dependencies makes them an ideal choice for modeling user-specific falling activity patterns in CSI amplitude data, as these patterns are often distributed across multiple subcarriers and time steps. Traditional models like CNNs may struggle to capture these global dependencies effectively, which is where their self-attention mechanism excels [Bai, 2018 ; Goodfellow, 2016].

4.2 Architecture of the Identifier

The IdentNet architecture, designed for HIDR using CSI, employs a Transformer-based model to process the activity patterns of different users. In this study, IdentNet classifies users based on their falling activity using CSI amplitude data, which is inherently multidimensional. The architecture capitalizes on the self-attention mechanism of Transformers, enabling it to model long-range dependencies between time steps and subcarriers, which are essential for distinguishing users based on their unique fall activity signatures. The self-attention mechanism is particularly adept at capturing both global temporal and spectral dependencies, making it well-suited for recognizing subtle variations in CSI data that are associated with each user's fall activity, which traditional models may overlook.

The architecture of IdentNet consists of several key components that work together to process the CSI data and classify users effectively. Below, we describe each component in detail:

1. **Input Layer**: This layer receives the preprocessed CSI amplitude data as a 3D tensor of the form $\mathbf{X} \in \mathbb{R}^{B\times T\times S}$. Here, B is the number of frames (sequences) in each training step (batch size), T is

the number of time steps in a single CSI frame (sequence length), and S is the number of subcarriers (features) per time step.

For the data used in our experiments, each CSI frame is 1 second of the user's activity sampled at 200 Hz, so $T = 200$. In addition, we have $S = 52$ subcarriers, each providing a distinct frequency response. Consequently, every input frame is represented by a matrix of size $(200, 52)$. By stacking these frames across the batch dimension, the model ultimately processes inputs of shape $(B, 200, 52)$. Positional embeddings are subsequently added to maintain temporal information, ensuring the Transformer can appropriately model the sequential nature of the data.

2. **Positional Encoding Layer**: This layer encodes the temporal sequence information, ensuring that the model recognizes the order of frames. Unlike Recurrent Neural Networks (RNNs) or CNNs, Transformers do not inherently capture sequence order, so additional embeddings are added to preserve this temporal relationship.

   In the original Transformer model [Vaswani, 2017], these positional values are computed using sinusoidal functions. However, in our implementation, we adopt a *learnable embedding layer* instead of the fixed sinusoidal approach. This means each time step is associated with a trainable embedding vector, allowing the model to discover an optimal representation of positional information directly from the data. The resulting embeddings are then added to the input to form a position-aware representation for the Transformer.

3. **Transformer Blocks**: The core of IdentNet consists of multiple Transformer blocks (four blocks in this implementation). Each Transformer block contains two main components:
   - *Multi-Head Attention*: This layer computes attention scores for different parts of the input sequence, allowing the model to attend to various subcarriers and time steps simultaneously.
   - *Feed-Forward Network (FFN)*: A simple neural network consisting of two Dense layers with a ReLU activation in between. The FFN processes the output of the multi-head attention layer, allowing the model to capture non-linear relationships between the features.

   The architecture allows each Transformer block to capture both local and global dependencies in the CSI data, which is crucial for recognizing user-specific patterns. The self-attention mechanism in particular is central to this process, as it can capture long-range dependencies across the entire input sequence, identifying important interactions between distant time steps and subcarriers. This ability to model global interactions makes the Transformer architecture highly effective in identifying subtle variations in the CSI amplitude data associated with each user's fall activity. The attention mechanism follows the formulation from [Vaswani, 2017]:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V$$

   Where:
   - Q, K, and V are the query, key, and value matrices, respectively.
   - d_k is the dimensionality of the key vectors.

   Here, $K^\top$ denotes the transpose of K, ensuring proper alignment in the dot product operation. Each block also includes Layer Normalization and Dropout layers to prevent overfitting [Ba, 2016].

4. **Global Average Pooling**: After the Transformer blocks, Global Average Pooling (GAP) is applied to reduce the dimensionality of the output. This step computes the average value for each feature map, summarizing the information across all frames and subcarriers, thereby retaining essential features while reducing computational complexity:

$$\mathbf{GAP}(X) = \frac{1}{T}\sum_{t=1}^{T} X_t$$

5. **Dense Layers**: Following the GAP, IdentNet includes a single fully connected (FC) Dense layer with 128 units and a ReLU activation. This layer refines the features extracted by the Transformer blocks, enabling the network to learn more complex patterns specific to each user's fall activity.
6. **Output Layer**: The final layer is a Dense layer with 4 output units, corresponding to the 4 classes of users (no_user, user_1, user_2, user_3). A softmax activation is applied to obtain class probabilities for multi-class classification.

The architecture of the IdentNet Transformer used for classifying users based on fall activity is illustrated in FIGURE 5.2.

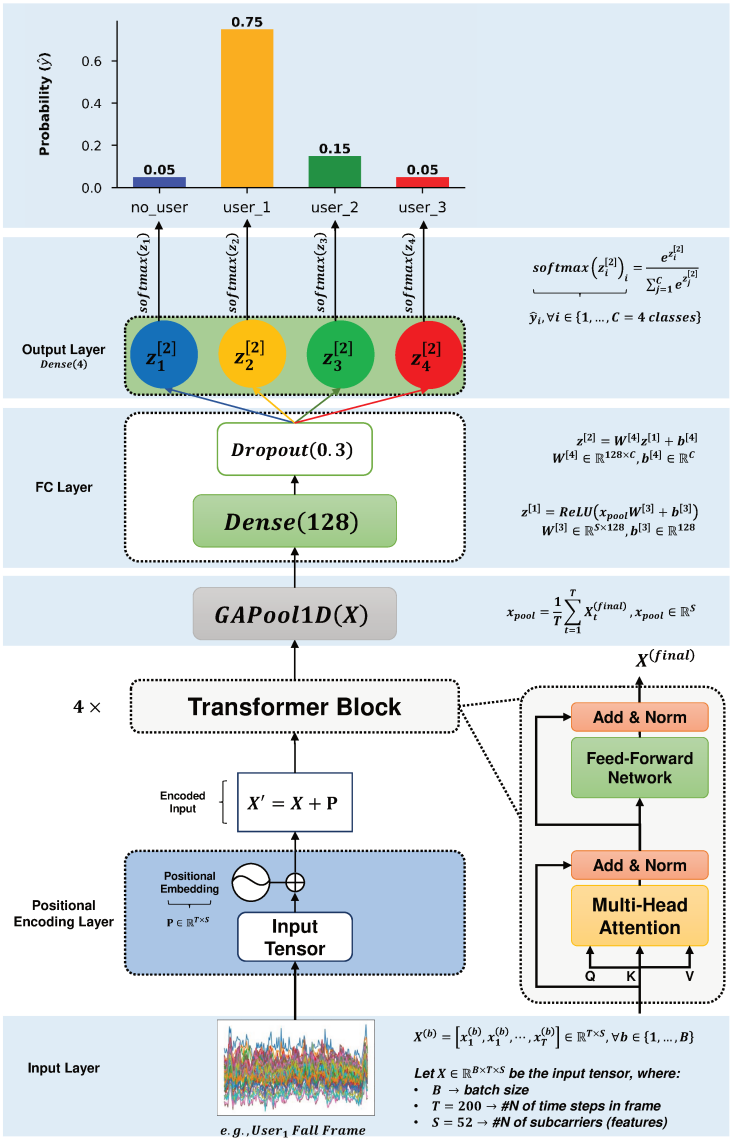


FIGURE 5.2 – Architecture of Transformer user identifier: This figure illustrates the core components of the IdentNet architecture, showcasing the Positional Embedding, Transformer Blocks, and Fully Connected Layers.

IdentNet is optimized using the AdamW optimizer [Loshchilov, 2017], which incorporates weight decay to prevent overfitting. The categorical cross-entropy loss function is used for multi-class classification, and early stopping and learning rate reduction are applied to ensure the model converges efficiently without overfitting. The key hyperparameters used in the IdentNet Transformer-based architecture are summarized in TABLE 5.1, which lists the most relevant configuration details.

TABLE 5.1 – Summary of Hyperparameters for IdentNet

Parameter	**Value**
Embedding Dimension	52
Number of Heads	4
Feed-Forward Dimension	128
Dropout Rate	0.3
L2 Regularization	0.001
Batch Size	64
Learning Rate	1e-4
Epochs	200

The IdentNet Transformer architecture is specifically designed to process the multidimensional CSI amplitude data associated with falling activity. By leveraging multi-head attention and self-attention mechanisms, the model captures both temporal and spectral dependencies, making it well-suited for the task of user identification. The self-attention mechanism is particularly effective at capturing global interactions between distant time steps and subcarriers, ensuring the model can recognize subtle variations in CSI data that distinguish one user's fall activity from another. Additionally, the inclusion of global average pooling and dense layers further refines the features, ensuring that the model can accurately distinguish between different users based on their unique CSI patterns during falls.

5 Results and Evaluation

5.1 Evaluation Metrics and Benchmarking

In this subsection, we introduce the key evaluation metrics used to assess the performance of the proposed Transformer-based user identifier, IdentNet, and compare it with the CNN-based model, as well as with state-of-the-art identity recognition models. All models are evaluated on the unseen testing set described in Section 3, ensuring a fair and consistent assessment of their performance. The primary goal of this evaluation is to provide a comprehensive understanding of the strengths and limitations of IdentNet in addressing the challenges of HIDR using CSI data during fall activity.

To evaluate the performance of the models, we use a combination of widely accepted evaluation metrics: accuracy, precision, recall, and F1-score. These metrics are chosen for their ability to provide a balanced view of the classifier's effectiveness, especially in multi-class classification problems. Due to the slight class imbalance in the testing data, we use the weighted versions of these metrics to ensure fair representation across all user classes. The following key evaluation metrics are considered [Sokolova, 2009]:

— **User Identification Accuracy**: This metric represents the proportion of correctly identified user classes out of the total number of instances. Specifically referred to as user identification accuracy, it provides insight into how well the model can differentiate between the users. The weighted accuracy adjusts for class imbalance by assigning a weight to each sample based on its class.

$$\text{W. Accuracy} = \frac{\sum_{i=1}^{n} w_i \cdot \delta(y_i, \hat{y}_i)}{\sum_{i=1}^{n} w_i}$$

where w_i is the sample weight based on the class distribution, and $\delta(y_i, \hat{y}_i)$ is the Kronecker delta function (1 if the prediction is correct, 0 otherwise).

— **Weighted Precision**: Precision evaluates the correct identification of positive results for each class, adjusted for the class distribution.

$$\text{W. Precision} = \sum_{k=1}^{K} \left(\frac{n_k}{N}\right) \cdot \text{Precision}_k$$

where K is the number of classes, n_k is the number of samples in class k, N is the total number of samples, and Precision_k is the precision for class k.

— **Weighted Recall**: Recall, or sensitivity, measures the model's ability to correctly identify all true instances of a class, weighted by class distribution.

$$\text{W. Recall} = \sum_{k=1}^{K} \left(\frac{n_k}{N}\right) \cdot \text{Recall}_k$$

where K is the number of classes, n_k is the number of samples in class k, and Recall_k is the recall for class k.

— **Weighted F1-Score**: It is the harmonic mean of weighted precision and recall, offering a balance between the two.

$$\text{W. F1-Score} = 2 \cdot \frac{\text{W. Precision} \cdot \text{W. Recall}}{\text{W. Precision} + \text{W. Recall}}$$

— **Response time**: It measures the latency between the input CSI activity frame and the model's output (predicted user class). This is critical for real-time applications, where quick decision-making is essential. The response time is reported in milliseconds (ms) per activity frame.

To ensure the robustness and relevance of our evaluation, we benchmark the Transformer-based IdentNet against existing state-of-the-art HIDR models. This comparison serves two main purposes:

1. **Assessing General Performance**: We aim to validate the performance improvements introduced by IdentNet compared to other models designed for identity recognition.
2. **Evaluating Versatility Across Metrics**: Benchmarking provides insights into how IdentNet handles the inherent challenges of CSI-based HIDR, including intra-class similarity and long-range dependencies in time series data.

The combination of user identification accuracy, precision, recall, and F1-score provides a well-rounded understanding of model performance, especially when supported by weighted versions of these metrics. Together with the response time metric, this evaluation process ensures that IdentNet's suitability for real-time, personalized healthcare applications is rigorously tested.

In the next subsections, we will present the comparative analysis of IdentNet against both the CNN-based models and state-of-the-art identity recognition models, illustrating the performance improvements achieved through the Transformer architecture.

5.2 Effectiveness and Efficiency of IdentNet

5.2.1 Comparison Against CNN-Based Model from Previous Contributions

In this subsection, we evaluate the performance of two DL models: the CNN-based classifier developed and utilized in Chapters 2 and 3, and the Transformer-based IdentNet, designed to identify users based on CSI data during falling activity. The CNN-based model, which focused on extracting local features

from CSI data, is compared with the Transformer architecture that aims to capture both local and global dependencies. This comparison highlights how transitioning to a Transformer-based approach improves the ability to identify subtle user-specific patterns that the CNN may overlook.

The CNN-based model achieved a user identification accuracy of 84.44%, with a precision of 85.19% and an F1-score of 82.78%. While it performs well in classifying certain users, such as *no_user* and $user_3$, its performance is notably weaker for $user_2$, with a recall of 54.4%. This demonstrates the CNN's limitations in differentiating between users with more overlapping fall dynamics due to its reliance on local feature extraction. As shown in FIGURE 5.3, the confusion matrix illustrates how the CNN struggles particularly with distinguishing $user_1$ and $user_2$, where substantial misclassifications occur.

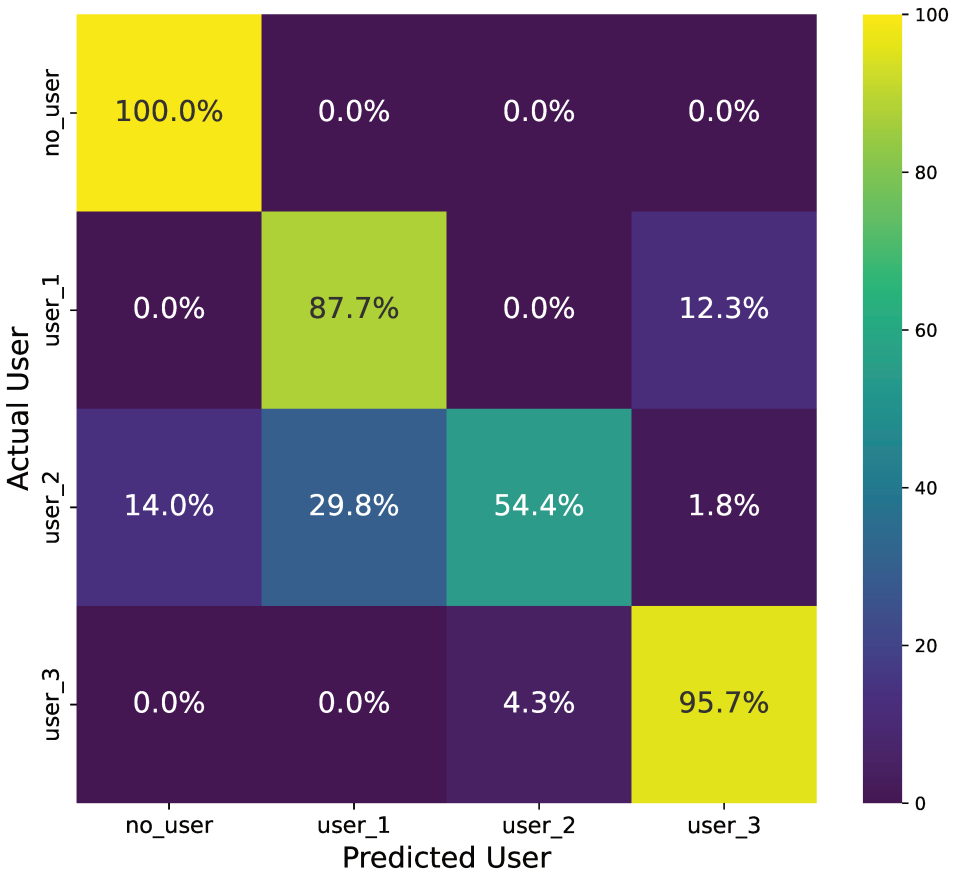


FIGURE 5.3 – Confusion Matrix for CNN-Based Classifier: This figure presents the confusion matrix, highlighting misclassifications particularly between $user_1$ and $user_2$, reflecting the CNN's limitations in handling subtle variations in CSI data.

The CNN's response time per frame was measured at 3.7 ms, reflecting its computational efficiency. However, its reduced capacity to model long-range dependencies limits its overall performance in distinguishing between users whose fall signatures are similar.

In contrast, the Transformer-based IdentNet outperformed the CNN model, achieving a user identification accuracy of 92.11%, with a precision of 93.67% and an F1-score of 91.45%. The Transformer's ability to capture both short- and long-range dependencies using its self-attention mechanism allowed it to better differentiate between users, even for challenging cases like $user_2$, where recall improved to 68.4%. FIGURE 5.4 shows the confusion matrix for the Transformer-based model, which demonstrates fewer misclassifications across all users, particularly for $user_1$ and $user_3$, where recall reached 100%.

Despite the Transformer's higher computational complexity, with a response time of 17.5 ms per frame,

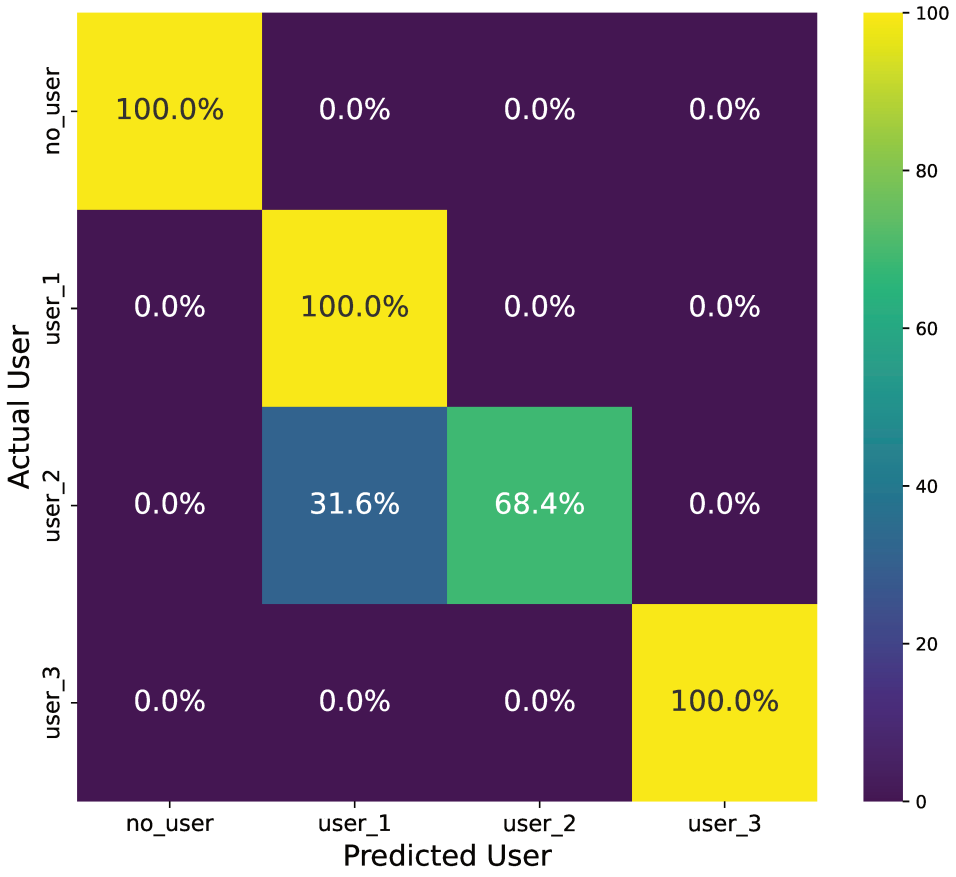


FIGURE 5.4 – Confusion Matrix for Transformer-Based IdentNet: This figure presents the confusion matrix, illustrating fewer misclassifications, particularly for $user_1$ and $user_3$, where recall is near perfect.

its significant improvement in classification accuracy makes it more suitable for personalized healthcare applications, where precise user identification is critical. TABLE 5.2 provides a comparative summary of the key evaluation metrics for both models.

TABLE 5.2 – Comparative Results of CNN-Based Classifier and Transformer-Based IdentNet

User Identifier / **Metric**	**CNN-Based Classifier**	**Transformer-Based IdentNet**
User Identification Acc. (%)	84.44	92.11
W. Precision (%)	85.19	93.67
W. Recall (%)	83.80	91.67
W. F1-Score (%)	82.78	91.45
Response Time (ms)	3.7	17.5

Overall, the Transformer-based IdentNet consistently outperforms the CNN-based classifier in all key evaluation metrics, with an improvement of 7.67% in user identification accuracy and 8.67% in precision. This superior performance, particularly in the case of $user_2$, demonstrates the Transformer's ability to model subtle variations across extended time sequences, which the CNN struggles with. While the Transformer incurs a higher computational cost, the trade-off is justifiable for applications where accuracy is paramount, such as real-time monitoring in healthcare settings. The CNN-based classifier, though faster in terms of response time, remains a viable option for scenarios where computational efficiency is more critical than high classification precision.

The comparison between the CNN-based classifier from earlier chapters and the Transformer-based IdentNet highlights the significant advancements achieved by leveraging Transformer architectures for user identification task. By capturing both local and global dependencies in CSI data, IdentNet offers superior user identification accuracy, albeit with an increased computational cost.

5.2.2 Benchmarking Against State-of-the-Art Models

In this subsection, we present a comparative evaluation of the proposed Transformer user identifier, IdentNet, against three prominent state-of-the-art models in CSI-based HIDR: LW-WiID developed by Cao et al. (2021) [Cao, 2021], CLSTM model proposed by Mo and Kim (2021) [Mo, 2021], and the SimpleViTFi introduced by Bian et al. (2024) [Bian, 2024]. These models were selected for their advanced approaches in handling CSI-based identity recognition and diverse architectural designs, encompassing CNN models, CNN-LSTM hybrids, and Transformer-based models. We aim to demonstrate that IdentNet surpasses these models in accuracy and robustness, particularly for user identification based on falling activity.

Our study compared these models, implemented according to the network configurations and hyper-parameter settings detailed in their respective papers, against our proposed model on the prepared unseen testing set. TABLE 5.3 summarizes the comparative performance of IdentNet and the selected models using the same dataset. All models were trained and evaluated under identical conditions, ensuring a fair and reliable comparison.

TABLE 5.3 – Comparison of IdentNet with State-of-the-Art Models in CSI-Based HIDR

User Identifier / **Metric**	**LW-WiID**	**CLSTM**	**SimpleViTFi**	**IdentNet**
User Identification Acc. (%)	84.69	50.00	87.72	**92.11**
W. Precision (%)	84.93	32.05	88.89	**93.67**
W. Recall (%)	84.72	47.22	87.04	**91.67**
W. F1-Score (%)	84.37	35.44	86.91	**91.45**
Response Time (ms)	4.67	4.21	8.89	17.5

The results clearly show that IdentNet consistently outperforms the other models across all key metrics. Specifically, IdentNet achieves the highest user identification accuracy of 92.11%, compared to 84.69% for LW-WiID, 50.00% for CLSTM, and 87.72% for SimpleViTFi. This superior performance can be attributed to IdentNet's Transformer-based architecture, which effectively captures long-range dependencies and subtle user-specific patterns in CSI data, particularly in complex activities like falling.

LW-WiID [Cao, 2021], which employs a CNN-based architecture optimized for low-resource environments, achieves an accuracy of 84.69%. While LW-WiID demonstrates competitive performance in identity recognition tasks, it is limited in its ability to capture long-range temporal and spectral dependencies, which are crucial for distinguishing subtle variations in fall dynamics. This limitation results in lower performance compared to IdentNet, which excels in modeling these dependencies through its Transformer-based architecture. IdentNet's self-attention mechanism enables it to better capture the nuanced user-specific falling activity patterns, significantly improving accuracy in distinguishing users with similar fall characteristics. However, LW-WiID offers a clear advantage in computational efficiency, with a lower response time of 4.67 ms per frame compared to IdentNet's 17.5 ms. This makes LW-WiID a strong candidate for real-time applications in resource-constrained environments, though it sacrifices some accuracy for speed.

The CLSTM model [Mo, 2021] integrates CNN for spatial feature extraction and LSTM for handling temporal dependencies. Despite its competitive architecture, the model struggles to perform well in user

identification during falling activity, yielding a relatively low accuracy of 50.00%. The hybrid architecture of CLSTM is limited in effectively capturing both spatial and long-term temporal features, which are critical in fall-based identity recognition. Additionally, its precision and F1-scores of 32.05% and 35.44% reflect the challenge of identifying users based on subtle fall-related patterns. However, it demonstrates a computational advantage with a response time of 4.21 ms, making it faster than the other models but less accurate for fall-based scenarios.

SimpleViTFi [Bian, 2024] is a lightweight Vision Transformer-based framework designed to handle imbalanced datasets using CSI-generated images. While SimpleViTFi performs well with an accuracy of 87.72%, its primary focus on class imbalance and general identity recognition results in lower performance compared to IdentNet in fall-specific scenarios. The model struggles with capturing fall-related identity traits, which IdentNet excels at due to its optimization for user identification during falling activity. However, SimpleViTFi offers a significant advantage in computational efficiency, with a response time of just 8.89 ms per frame, compared to 17.5 ms for IdentNet. This lightweight design, though computationally efficient, sacrifices some accuracy in fall-specific identification, demonstrating the trade-off between speed and precision in such models.

The development and use of SimpleViTFi highlight the emerging potential of Transformer-based architectures in CSI-based identity recognition. While traditionally dominated by CNNs, such as LW-WiID, the application of Transformers is relatively new but promising, as demonstrated by the results. Transformers excel in tasks involving subtle behavioral patterns, such as those seen in falling activity, due to their ability to capture long-range dependencies. Although SimpleViTFi achieves lower accuracy than IdentNet, its performance shows the potential of Transformers in this domain.

While IdentNet offers the highest performance across all accuracy-related metrics, it incurs a higher computational cost, with a response time of 17.5 ms per frame, compared to 4.67 ms for LW-WiID, 4.21 ms for CLSTM, and 8.89 ms for SimpleViTFi. This increased complexity is a trade-off for the transformer's ability to process the full sequence of CSI data and capture long-range dependencies. However, in healthcare applications, where accurate user identification and fall detection are critical, the trade-off between computational complexity and performance is justifiable.

The benchmarking results demonstrate that IdentNet significantly outperforms the selected state-of-the-art models across all key metrics for CSI-based identity recognition during falling activity. The ability of IdentNet to effectively capture long-range dependencies and subtle user-specific fall patterns makes it the most suitable model for real-time personalized healthcare applications. While IdentNet demands a higher computational cost, its superior accuracy and robustness make it a highly effective solution for identity recognition and monitoring in real-time scenarios.

6 Conclusions

6.1 Summary of Key Findings

This chapter presents a novel Transformer-based model, IdentNet, for HIDR using CSI data, focusing specifically on user identification during falling activity. The research introduces several significant advancements that address key challenges in CSI-based HIDR, and the following key findings have emerged:

1. **Introduction of IdentNet for CSI-Based HIDR**: The development of IdentNet marks a significant advancement in the field of CSI-based identity recognition, particularly for personalized healthcare applications. By leveraging the self-attention mechanism of transformers, IdentNet effectively captures both local and global dependencies in CSI data, enabling it to distinguish between users based on their unique fall patterns. This approach overcomes the limitations of traditional

CNN-based models, which struggle to handle long-term dependencies in multidimensional time series data.

2. **Superior Performance Over CNN-Based Models**: IdentNet demonstrated substantial improvements over the CNN-based model previously used for activity recognition and fall detection. The Transformer-based architecture excels in capturing subtle variations in user-specific falling activity, resulting in an overall accuracy improvement of 7.67%. These improvements are particularly evident in distinguishing between users with overlapping fall dynamics, where CNN models often face challenges due to intra-class similarity.
3. **Handling Intra-Class Similarity and Long-Term Dependencies**: One of the key challenges in CSI-based HIDR is the presence of intra-class similarity, where different users exhibit highly similar fall patterns. IdentNet addresses this challenge by utilizing the transformer's self-attention mechanism, which captures fine-grained, user-specific variations across the entire CSI sequence. Additionally, IdentNet's ability to handle long-term dependencies in the data ensures that it can accurately identify users even in complex scenarios where traditional models fall short.
4. **Benchmarking Against State-of-the-Art Models**: In addition to outperforming CNN-based models, IdentNet also demonstrated superior performance when benchmarked against state-of-the-art HIDR models, including SimpleViTFi, WiTL, and CLSTM. IdentNet achieved the highest accuracy, precision, and F1-score across all comparative models, highlighting its effectiveness in distinguishing between users based on CSI amplitude data. While IdentNet's increased computational cost is a trade-off for its enhanced performance, it remains well-suited for real-world applications in personalized healthcare, where accuracy is paramount.
5. **Advancement in Personalized Healthcare Monitoring**: IdentNet contributes to the advancement of personalized healthcare by enabling accurate, non-intrusive user identification during critical events like falls. This capability facilitates timely, tailored interventions, improving patient safety and healthcare outcomes. To our knowledge, this is the first work to utilize CSI data specifically from falling activity for identity recognition, marking a novel application of Transformer-based models in the field of healthcare monitoring.

These key findings collectively underscore the potential of Transformer-based models like IdentNet to revolutionize CSI-based HIDR, offering a robust, scalable, and accurate solution for patient monitoring and user identification in healthcare environments. The chapter demonstrates the significant improvements IdentNet brings to the domain, particularly in addressing the limitations of existing methods and advancing the state of the art.

6.2 Benefits for Healthcare Services

The integration of HFD and HIDR systems using CSI offers significant benefits to healthcare services, particularly in personalized patient monitoring and emergency response. These systems provide continuous, non-intrusive monitoring, enhancing the quality of care and improving patient outcomes. The ability to accurately detect falls and identify individuals during critical events is pivotal for timely and appropriate medical intervention.

6.2.1 Enhancing Personalized Emergency Response

CSI-based HFD and HIDR systems collectively contribute to the development of a personalized emergency response framework, especially in scenarios involving fall incidents. Falls are a major health concern, particularly among older adults, leading to serious injuries or fatalities if not addressed promptly [World Health Organization, 2021 ; Centers for Disease Control and Prevention, 2023c]. Traditional emergency

response frameworks often lack the precision to differentiate the needs of individuals, resulting in generalized and sometimes inadequate responses. By integrating user-specific identification into fall detection systems, healthcare providers can tailor interventions based on the individual's profile, ensuring a more effective and efficient response.

In Chapter 3, we introduced an advanced fall detection methodology, HFD-Adv, utilizing CSI data to detect falls with high accuracy. This methodology focuses on recognizing fall events and ensuring rapid intervention by alerting caregivers or emergency services. However, a critical question remains—how can the emergency response be further personalized? Knowing the identity of the person who has fallen allows healthcare systems to adapt responses according to the specific needs and health conditions of the individual. For instance, if an elderly person with known frailty issues is detected to have fallen, the response protocol can include immediate deployment of medical assistance, even if the fall does not initially appear severe. Conversely, if a younger, healthier individual is detected, the system might recommend a different intervention strategy, potentially optimizing healthcare resources.

The proposed HIDR methodology in this chapter addresses this need for personalization by accurately identifying the individual involved in the fall. By identifying users based on unique behavioral patterns captured through CSI data, healthcare systems can provide tailored interventions. This can range from sending alerts to caregivers to dispatching emergency services with specific instructions tailored to the patient's medical history. Integrating the HIDR methodology, IdentNet, with the fall detection system enhances the decision-making process, leading to quicker and more appropriate medical responses, particularly benefiting high-risk populations such as the elderly or those with chronic conditions [Vicente, 2020; Duan, 2023].

6.2.2 Protocol for Integrated HFD-Adv and IdentNet Deployment

To realize the full potential of our HFD and HIDR methodologies in healthcare, we propose a comprehensive deployment protocol. This protocol integrates the contributions from Chapter 3 and Chapter 5, enabling a seamless workflow from fall detection to user identification:

1. **Real-Time Fall Detection**: The system continuously monitors for fall events using CSI data, leveraging the HFD-Adv methodology from Chapter 3 to detect falls with high accuracy. It operates in real-time, ensuring immediate recognition of critical events.
2. **User Identification and Profile Analysis**: Upon detecting a fall, the system activates the user identifier, **IdentNet**, introduced in this chapter. This model identifies the individual based on their unique fall patterns captured in the CSI data. The identity information is cross-referenced with the user's health profile, including age, medical conditions, and specific needs.
3. **Tailored Intervention Strategy**: Once the individual is identified, the system automatically adjusts the intervention strategy based on the user's health profile. For example, if the identified user is an elderly person with a history of osteoporosis, the system may prioritize dispatching medical personnel even for minor falls, as such individuals are more susceptible to fractures.
4. **Adaptability to New Users**: While our models are initially trained on data from specific users, the system is designed to be adaptable. During deployment, the models can be fine-tuned or retrained with data from new users, ensuring accurate identification and monitoring across diverse populations. This adaptability is crucial for practical applications where the system must handle new users with varying characteristics.
5. **Feedback Loop for Continuous Improvement**: Post-event, the system records data on response times and intervention effectiveness, refining algorithms for future responses. This iterative process ensures the system remains adaptive, handling diverse real-world scenarios effectively.

This protocol leverages our proposed approaches, HFD-Adv and IdentNet, to create an integrated system capable of providing personalized emergency responses. By deploying this protocol, healthcare providers can enhance the efficiency and effectiveness of their emergency response service.

6.2.3 Scenario: Application in Residential Healthcare Settings

Consider a scenario in a residential healthcare setting where multiple individuals of varying ages and health conditions reside, such as an assisted living facility or a multigenerational family home. The facility is equipped with our integrated CSI-based fall detection and user identification system, utilizing HFD-Adv and IdentNet.

Scenario: An elderly resident, John, who has a known history of heart disease and is at high risk for severe injury from falls, experiences a fall in the living room. The HFD-Adv methodology detects the fall event within seconds and triggers an alert. Simultaneously, the IdentNet model processes the CSI data to determine that John is the individual who has fallen.

Upon identification, the system accesses John's medical profile and recognizes the urgency due to his age and health conditions. An immediate alert is dispatched to emergency medical services with detailed information about John, including his exact location within the residence and pertinent medical history. Simultaneously, notifications are sent to designated caregivers and family members.

This rapid, personalized response ensures that John receives timely medical attention, significantly reducing the risk of complications and improving his chances of a full recovery. Additionally, the system logs the incident for further analysis, contributing to ongoing health monitoring and risk assessment for John.

6.2.4 Advancing Personalized Healthcare

The integration of our HFD-Adv and IdentNet approaches exemplifies the advancement of personalized healthcare through technological innovation. By moving beyond generic monitoring to individualized identification and response, healthcare services can provide more effective and compassionate care. This approach aligns with the vision for future healthcare systems that prioritize patient-specific solutions and leverage data-driven technologies to improve health outcomes [Vicente, 2020 ; Minvielle, 2021].

While our models were initially trained on data from a limited number of users, the methodology is adaptable and can be extended to new users. In practice, deploying the system in a new environment involves a configuration phase where initial data is collected from the users. The models can then be fine-tuned or retrained with this data, ensuring high accuracy in both fall detection and user identification regardless of the specific individuals being monitored. This adaptability enhances the practicality and scalability of our system for widespread adoption.

Furthermore, the non-intrusive nature of CSI-based monitoring respects patient privacy and autonomy, essential considerations in modern healthcare delivery [Duan, 2023]. The ability to unobtrusively monitor and respond to patient needs without wearable devices or cameras enhances patient comfort and compliance.

6.2.5 Summary of Benefits

The integration of the HFD-Adv fall detection methodology and the IdentNet user identifier model into healthcare services offers substantial benefits by enabling personalized emergency response and tailored intervention strategies. By accurately detecting falls and identifying individuals during incidents, healthcare providers can deliver timely and appropriate responses, significantly improving patient experiences and outcomes.

The complementary nature of the contributions in Chapters 3 and 5 demonstrates the potential of combining advanced detection and identification technologies to enhance emergency response service. This integrated approach represents a significant step forward in realizing personalized, efficient, and effective healthcare delivery.

6.3 Future Directions

Building upon the advancements presented in this chapter, several future research directions can further enhance the effectiveness, adaptability, and practical deployment of Transformer-based HIDR systems using CSI data.

Firstly, while IdentNet has shown superior performance in identifying users based on falling activity, expanding the methodology to include a wider range of activities could enhance the system's applicability. Future work could involve training the model on additional daily activities such as walking, sitting, or running. This would allow the system to identify users in various contexts, improving its utility in comprehensive patient monitoring and personalized healthcare services.

Secondly, the current study is based on data from a limited number of users within a controlled environment. Collecting and incorporating larger and more diverse datasets with more participants, varied physical characteristics, and different environmental settings would improve the model's generalization capabilities. This expansion would enable the model to handle a broader spectrum of user behaviors and environmental conditions, enhancing its robustness and reliability in real-world applications.

Moreover, addressing the challenge of fall diversity is an important future direction. Falls can occur in various ways and contexts, and the model's ability to identify users accurately across different fall types is crucial. To tackle this, future research could explore the integration of adversarial learning techniques into IdentNet. Adversarial training can be employed to make the feature extractor invariant to specific fall patterns while enhancing user-specific features. This approach would involve training the model to focus on extracting user identity information that is robust to variations in fall motions. By deliberately encouraging the network to ignore fall-specific characteristics and emphasize user-specific features, the system can maintain high identification accuracy without requiring recalibration for new or unseen fall types. This enhancement would make the user identification results more robust to the diversity of falls, ensuring consistent performance even with limited fall data.

Additionally, addressing the challenge of multi-user scenarios is essential. In practical settings such as hospitals or assisted living facilities, the system may need to identify and monitor multiple individuals simultaneously. Developing algorithms capable of distinguishing between multiple users in the same environment would significantly broaden the system's applicability. Techniques such as advanced signal separation, multi-object tracking, or incorporating spatial information from multiple antennas could be explored to achieve this capability.

Another potential area of improvement is reducing the computational complexity and response time of the Transformer-based model. While IdentNet demonstrates high accuracy, its computational demands may limit real-time applications on resource-constrained devices. Future research could focus on optimizing the model through techniques like model pruning, quantization, or developing lightweight transformer architectures. These optimizations would facilitate deployment on edge devices, enabling real-time processing and expanding the system's scalability.

In addition to the fall-based user identification experiments presented in this chapter, we have conducted a preliminary study on the running activity to further demonstrate IdentNet's adaptability. Detailed results, along with a transfer learning setup leveraging minimal fall data, are provided in the Appendix. These findings underscore IdentNet's capacity to generalize beyond a single activity, offering a practical solution when collecting sufficient fall samples is difficult in real-world healthcare environments.

By pursuing these future directions, including the integration of adversarial learning to handle fall diversity, the capabilities of Transformer-based HIDR systems like IdentNet can be significantly enhanced. These efforts will contribute to more effective, adaptable, and personalized healthcare services, ultimately improving safety and quality of life for individuals under continuous monitoring.

7 Appendix: Extending IdentNet to Other User Activities

Note: This Appendix presents an extended set of experiments demonstrating IdentNet's applicability to a different daily activity (running) and its capacity for transfer learning with limited data from a more critical (but rarer) event, such as falling.

7.1 Introduction and Experimental Hypotheses

In these additional experiments, we employ the same dataset (detailed in Section 3) of three users each performing seven daily activities over multiple trials. As before, a "no_user" class is introduced using 'empty' (non-user) activity CSI amplitude data from an external source. Here, our principal focus is on the *running* activity, which is easier to collect and yields a consistent movement pattern.

We then investigate whether the proposed Transformer-based IdentNet model, trained on identifying users during running, can generalize to a second activity—*falling*—which is both critical and rare in real-world settings. Subsequently, we examine whether a small amount of fall data can be used to fine-tune IdentNet.

Our main experimental questions are:

— **Q1.** Can IdentNet effectively identify users based on running activity alone?

— **Q2.** How does IdentNet perform on unseen user fall activity with no retraining?

— **Q3.** Can minimal fine-tuning on limited fall data improve performance for fall-based user identification?

— **Q4.** Does this fine-tuning degrade performance on the original running data?

By systematically exploring these questions, we aim to demonstrate the feasibility of using a single activity (like running) for initial IdentNet model training and highlight how limited data from a rarer activity (falling) can be incorporated through fine-tuning.

7.2 User CSI Data Preprocessing

7.2.1 Run-based Training Set Preparation

We retain only the running activity from each user for the training set, aiming to train IdentNet to distinguish among "no_user," "user_1," "user_2," and "user_3" via 1-second frames (200×52). We then later examine how this run-trained model performs on falling data.

For training and evaluating IdentNet in the running scenario, the data split is as follows:

— **Training trials**: 17 trials of running per user + 40 trials of "empty" (no_user).

— **Testing trials**: 3 trials of running per user + 7 trials of empty (no_user).

After framing and preprocessing, the training set comprises :

— **no_user**: 320 frames

— **user_1**: 321 frames

— **user_2**: 305 frames

— **user_3**: 246 frames

7.2.2 Holdout Method for Robust Evaluation

To confirm performance stability, we adopt a repeated holdout method ($N = 3$ splits). In each split:

— 85% of the data is for training, and 15% is for validation.

— We randomly sample a seed for each split to ensure distinct partitions.

This yields three different training-validation sets, each used to train an instance of IdentNet. We then examine whether results remain consistent across splits.

7.2.3 Three Testing Sets Preparation

We create three testing sets to evaluate the trained models:

— **Run-based testing set**: 212 frames total (60 from user_1, 51 from user_2, etc.).

— **Fall-based testing set**: 216 frames total (57 from user_1, 57 from user_2, 46 from user_3, 56 no_user).

— **(Run+Fall)-based testing set**: 428 frames combining both run and fall user data plus no_user.

These sets allow us to test generalization to falling activity and to assess performance on both activities in a combined scenario.

7.3 Experiments and Results

7.3.1 Baseline: Identity Recognition through Running

We first verify if IdentNet accurately classifies users based solely on running. TABLE 5.4 shows the performance across the three holdout splits on the run-based testing set.

TABLE 5.4 – Performance of IdentNet on Run-based Testing Set

Metric \ Split	S_0	S_1	S_2	Average ($\bar{S}$)
User Identification Acc. (%)	97.78	97.22	97.78	97.59
W. Precision (%)	98.23	97.77	98.23	98.08
W. Recall (%)	98.11	97.64	98.11	97.95
W. F1-Score (%)	98.10	97.60	98.10	97.93
Response Time (ms)	10.80	10.51	10.37	10.56

IdentNet exceeds 97% accuracy across all splits, confirming that running alone can sufficiently characterize each user's unique motion signatures.

7.3.2 Direct Generalization to Falling User Data

Next, we test each run-trained IdentNet directly on the fall-based testing set, with no additional retraining. TABLE 5.5 summarizes results:

Although accuracy drops compared to the running scenario, IdentNet still achieves moderate user identification on fall data, indicating partial carryover of user-specific features even without fine-tuning.

7.3.3 Fine-Tuning on Limited Falling User Data

To improve performance on fall data, we fine-tune each run-trained model using a minimal subset of fall frames: just one fall trial per user (15–22 frames). We then retrain IdentNet for 5 additional epochs on these frames. TABLE 5.6 shows the resulting performance on the fall, run, and combined (run+fall) testing sets.

TABLE 5.5 – Performance of IdentNet on Fall-based Testing Set Before Fine-tuning

Metric \ Split	$\mathbf{S_0}$	$\mathbf{S_1}$	$\mathbf{S_2}$	**Average** ($\bar{S}$)
User Identification Acc. (%)	56.79	70.03	57.21	61.34
W. Precision (%)	66.48	82.07	68.94	72.50
W. Recall (%)	59.26	72.69	59.26	63.74
W. F1-Score (%)	53.97	69.71	53.73	59.14
Response Time (ms)	10.50	10.50	10.18	10.39

TABLE 5.6 – Performance of IdentNet After Fine-tuning

Metric \ Test set	Fall-based				Run-based				(Run+Fall)-based			
	$\mathcal{S}_0$	$\mathcal{S}_1$	$\mathcal{S}_2$	$\bar{\mathcal{S}}$	$\mathcal{S}_0$	$\mathcal{S}_1$	$\mathcal{S}_2$	$\bar{\mathcal{S}}$	$\mathcal{S}_0$	$\mathcal{S}_1$	$\mathcal{S}_2$	$\bar{\mathcal{S}}$
User Identification Acc. (%)	88.05	92.11	87.72	89.29	92.64	85.97	80.83	86.48	90.19	88.96	84.14	87.76
W. Precision (%)	89.69	92.86	91.94	91.50	94.01	87.90	84.09	88.67	90.60	90.30	86.26	89.05
W. Recall (%)	87.50	91.67	87.04	88.74	93.40	86.32	82.55	87.42	90.42	89.02	84.81	88.08
W. F1-Score (%)	87.41	91.74	87.55	88.90	93.31	86.29	81.63	87.08	90.45	89.17	85.24	88.29
Response Time (ms)	10.92	10.37	10.50	10.60	11.60	11.50	11.03	11.38	5.35	5.32	5.35	5.34

Key observations:

1. Substantial improvements on fall-based user identification (up to $\approx 89\%$ average accuracy).
2. Running-based user identification remains fairly high despite slight drops (likely due to "catastrophic forgetting," where fine-tuning shifts network parameters away from the original task).
3. Combined test (run+fall) still achieves 84–90% accuracy, indicating the model retains a significant portion of run-based knowledge after fine-tuning.

7.3.4 Discussion

Effectiveness of IdentNet. Baseline results confirm that IdentNet effectively learns user-specific signatures from running activity, exceeding 97% accuracy. When tested directly on fall data with no additional training, it still achieves moderate identification, indicating partial invariance of the learned user patterns (Q1 and Q2).

Fine-Tuning vs. Catastrophic Forgetting. Fine-tuning on just one user fall trial per user substantially boosts user identification for falls (Q3). However, this small amount of new data can lead to minor performance degradation on the original running data (Q4), reflecting a well-known challenge in DL called catastrophic forgetting. Although we only retrained for 5 epochs with a modest dataset, more advanced approaches—like freezing early layers or mixing a portion of the running data during fine-tuning—could further mitigate this issue.

Adapting Rare Activities. Our results show that even minimal calibration (one fall trial per user) can align a run-trained IdentNet to handle falls more effectively, demonstrating a practical pathway for extending the model to additional rare activities.

Real-World Potential. In practical healthcare scenarios, collecting run data is straightforward, while fall data is limited. Our results indicate that a handful of fall trials can help a run-trained IdentNet achieve strong identification on falling activity without discarding its knowledge of running.

7.4 Conclusion and Outlook

In summary, a Transformer-based model (IdentNet) trained solely on running data can achieve >97% user identification accuracy. When exposed to fall data without any retraining, its performance remains moderate, suggesting a transfer of user-specific patterns across different activites. However, fine-tuning on as few as one fall trial per user leads to a notable boost in fall-based user identification, with only modest drops in the running-based performance.

Next Steps.

— **Layer Freezing and Gradual Unfreezing.** More advanced transfer learning strategies could mitigate catastrophic forgetting, e.g., freezing early Transformer layers and incrementally unfreezing them for adaptation to new activities.

— **Beyond Running and Falling.** While we tested two exemplar activities, the same pipeline is applicable to other daily movements (walking, sitting, etc.). This extension would strengthen IdentNet's adaptability to various healthcare monitoring scenarios, eventually supporting multi-activity user identfication.

General Conclusion and Perspectives

The advancement of Human Activity Recognition (HAR) using Wi-Fi Channel State Information (CSI) presents a significant leap forward in leveraging ambient sensing technologies for healthcare applications. This thesis has focused on addressing critical limitations in CSI-based HAR, aiming to enhance the efficiency, accuracy, and applicability of HAR systems in real-world healthcare settings. By systematically tackling data-related and model-related concerns, the research has contributed three principal methodologies: HAR-LightCNN for efficient activity recognition, HFD-Adv for advanced fall detection, and IdentNet for human identity recognition during fall incidents. These contributions have advanced the state of the art in CSI-based HAR and laid a robust foundation for integrating these technologies into personalized healthcare services.

1 Conclusions and Discussions

The research undertaken in this thesis addresses two major problems hindering the practical implementation and widespread adoption of CSI-based HAR in healthcare applications: **Data-Related Concerns** and **Model-Related Concerns**. By developing innovative methodologies, the research contributes to overcoming specific limitations within these broader problems, thereby enhancing the performance, reliability, and applicability of HAR systems.

Addressing Data-Related Concerns: Data-related concerns, such as small-sized and imbalanced data, missing training data, absence of empty data, and limited data diversity, significantly impact the generalizability and reliability of HAR models. The research has addressed these concerns through various strategies:

— **Data Augmentation (DA):** In the development of HAR-LightCNN, time-series DA techniques were employed to enrich the training data, addressing small-sized and imbalanced datasets. This approach enhanced the model's ability to generalize from limited data, improving accuracy and robustness in recognizing single-user activities.

— **Data Integrity Optimization:** The HFD-Adv methodology introduced threshold-based padding to retain incomplete frames, mitigating the issue of missing training data. Additionally, the inclusion of "empty" activity data and the use of Principal Component Analysis (PCA) for data harmonization addressed the absence of empty data and limited data diversity, respectively, enhancing model robustness and adaptability.

— **Refined Transfer Learning:** HFD-Adv methodology incorporated refined transfer learning strategies to adapt models to new users and environments, even with limited initial data. This approach improved generalization across diverse demographics and scenarios, crucial for real-world healthcare applications.

Addressing Model-Related Concerns: Model-related concerns include limitations such as high model complexity, overfitting, hyper-parameter sensitivity, difficulty handling long-term dependencies, and the absence of effective post-processing mechanisms. The research tackled these issues through innovative model designs and optimization techniques:

— **Lightweight Model Design:** HAR-LightCNN was developed as a lightweight convolutional neural network (CNN), reducing computational complexity and memory usage. This design addressed the high model complexity limitation, making the model suitable for real-time deployment in resource-constrained environments.

— **Advanced Model Architectures:** The HFD-Adv methodology introduced ActiNet, a flexible CNN tailored for varied HAR tasks, and IdentNet, a transformer-based model utilizing self-attention mechanisms. These architectures effectively handled long-term dependencies and subtle variations in CSI data, overcoming limitations of traditional models.

— **Optimization Techniques:** Techniques such as Genetic Algorithms (GAs) for weighted loss function optimization and the LRFinder mechanism for learning rate optimization were employed. These strategies mitigated hyper-parameter sensitivity and overfitting, enhancing model precision and training efficiency.

— **Post-processing Mechanisms:** The introduction of the Advanced Decision Mechanism (ADM) in HFD-Adv addressed the absence of effective post-processing mechanisms. ADM improved detection accuracy by reassessing classifier outputs and incorporating demographic-sensitive adjustments.

Key Contributions and Findings: The three principal contributions of this thesis have collectively advanced the field of CSI-based HAR:

HAR-LightCNN

— **Efficiency and Effectiveness:** HAR-LightCNN demonstrated superior accuracy and efficiency in recognizing seven daily activities using Wi-Fi CSI data. Its lightweight design reduced computational demands, making it ideal for real-time applications.

— **Addressing Data Limitations:** Through DA techniques, the model overcame challenges of small-sized and imbalanced datasets, enhancing generalization and robustness.

— **Rigorous Evaluation:** The methodology included fair benchmarking and rigorous testing procedures, ensuring the reliability and validity of the results.

HFD-Adv Methodology

— **Advanced Fall Detection:** HFD-Adv achieved a perfect Fall Detection Rate (FDR) of 100% and a low False Alarm Rate (FAR), significantly outperforming existing methods.

— **Adaptability and Generalization:** The methodology's refined transfer learning and data harmonization strategies enhanced adaptability to new users and environments.

— **Efficiency Improvements:** Optimization techniques reduced training periods, accelerating model convergence and facilitating quicker iterations.

IdentNet Transformer Model

— **Innovative HIDR Solution:** IdentNet effectively identified individuals during falls by capturing subtle user-specific patterns in CSI data.

— **Handling Intra-Class Similarity:** The transformer-based architecture overcame challenges of intra-class similarity and long-term dependencies.

— **Benchmarking Excellence:** IdentNet outperformed state-of-the-art HIDR models, showcasing its potential for personalized healthcare applications.

Benefits to Healthcare Services: The integration of these methodologies into healthcare services offers substantial benefits:

— **Enhanced Patient Monitoring:** Non-intrusive monitoring systems like HAR-LightCNN improve patient care by accurately recognizing daily activities without the need for wearables.

— **Improved Emergency Response:** HFD-Adv enhances fall detection accuracy and response times, crucial for preventing severe injuries among vulnerable populations.

— **Personalized Healthcare:** IdentNet enables tailored interventions by accurately identifying individuals during fall incidents, advancing personalized patient care.

— **Cost-Effectiveness and Accessibility:** Leveraging existing Wi-Fi infrastructure reduces costs and simplifies deployment in various healthcare settings.

2 Perspectives

Building upon the foundational work presented in this thesis, several future research directions emerge that can further enhance the capabilities and impact of CSI-based HAR systems in healthcare. These perspectives are organized into **short-term**, **mid-term**, and **long-term** horizons, reflecting the progression from immediate enhancements to broader implications and visionary applications.

Short-Term Enhancements

Expanding Activity Recognition Scope: While HAR-LightCNN effectively recognizes seven daily activities, expanding the model to include a wider range of activities can enhance its applicability. Future work could involve:

— **Integrating Additional Activities:** Training the model on activities such as eating, drinking, or social interactions to provide a more comprehensive monitoring system.

— **Multi-User Activity Recognition:** Adapting the methodology to handle scenarios where multiple individuals are present, distinguishing between different users' activities.

Enhancing Model Robustness and Adaptability: Improving the adaptability of models to new users and environments remains a priority:

— **Collecting Diverse Datasets:** Incorporating larger and more varied datasets with different demographics, environments, and behaviors to improve generalization.

— **Advanced Data Augmentation Techniques:** Employing sophisticated methods like Generative Adversarial Networks (GANs) to generate synthetic CSI data, enriching training datasets.

Optimizing Computational Efficiency: Reducing computational demands without sacrificing accuracy is essential for real-time applications:

— **Model Compression and Optimization:** Exploring techniques such as model pruning, quantization, or developing lightweight architectures to facilitate deployment on edge devices.

— **Energy-Efficient Algorithms:** Designing algorithms that consume less power, crucial for battery-operated or resource-constrained devices.

Mid-Term Research Directions

Integrating Contextual and Environmental Information: Incorporating contextual data can enhance system performance:

— **Environmental Adaptation:** Developing methods to adapt models to changes in the environment, such as furniture rearrangement or varying signal conditions.

— **Context-Aware Systems:** Integrating data from other sensors or sources to provide context, improving activity recognition and fall detection accuracy.

Advancing Personalization in Healthcare Monitoring: Enhancing personalization can improve patient outcomes:

— **User-Specific Models:** Implementing algorithms that learn and adapt to individual user behaviors over time.

— **Health Condition Integration:** Incorporating health profiles and medical histories into the decision-making process for more tailored interventions.

Multi-Modal Data Fusion: Combining CSI data with other modalities can provide richer insights:

— **Sensor Fusion:** Integrating data from wearables, cameras, or environmental sensors to improve accuracy and reliability.

— **Cross-Domain Learning:** Applying transfer learning across different data types to enhance model robustness.

Long-Term Impact and Broader Implications

Establishing Industry Standards and Guidelines: As CSI-based HAR technologies mature, establishing standards is crucial:

— **Protocol Development:** Creating standardized protocols for data collection, model training, and evaluation to ensure consistency and interoperability.

— **Regulatory Compliance:** Addressing privacy, security, and ethical considerations in line with healthcare regulations.

Global Adoption and Implementation: Promoting widespread adoption can have significant societal benefits:

— **Scalable Deployment Strategies:** Developing frameworks for large-scale implementation in hospitals, assisted living facilities, and homes.

— **Collaboration with Healthcare Providers:** Partnering with healthcare organizations to integrate HAR systems into existing care models.

Exploring New Applications and Domains: Expanding beyond healthcare can open new opportunities:

— **Smart Environments:** Applying CSI-based HAR in smart homes, offices, or public spaces for enhanced safety and convenience.

— **Security and Surveillance:** Utilizing identity recognition capabilities for access control and security monitoring.

By pursuing these future directions, the research community can continue to advance CSI-based HAR technologies, contributing to more effective, personalized, and accessible healthcare services. The long-term vision is to integrate these systems seamlessly into everyday life, enhancing safety, well-being, and quality of life on a global scale.

Résumé long de la thèse

1 Contexte et motivations

La reconnaissance d'activités humaines (HAR) est devenue un domaine interdisciplinaire majeur, avec des applications variées en santé, dans l'Internet des Objets (IoT) et en sécurité. Sa capacité à fournir un suivi continu et non intrusif est particulièrement précieuse pour optimiser la prise en charge des patients, faciliter des interventions rapides et améliorer la précision des diagnostics. Cependant, les solutions traditionnelles reposant sur des caméras ou des capteurs portés soulèvent souvent des problèmes de confidentialité ou exigent la coopération des utilisateurs — autant de difficultés qui se révèlent d'autant plus critiques en contextes résidentiels ou cliniques, notamment auprès de populations fragiles comme les personnes âgées, les personnes handicapées ou les enfants.

Pour dépasser ces limites, cette thèse exploite la **Channel State Information (CSI) issue du Wi-Fi**, qui autorise une approche préservant la confidentialité et ne nécessitant aucun dispositif porté, en détectant les perturbations subtiles des signaux sans fil induites par le mouvement humain. **La Partie I (Chapitres 1 et 2)** du manuscrit présente les principes fondamentaux de la HAR fondée sur la CSI, puis conduit une **étude de cartographie systématique (SMS)** de la littérature. Cette analyse met en évidence des lacunes tant au niveau des données (disponibilité limitée, complexité du traitement) que des modèles (manque de post-traitement avancé, absence d'une évaluation rigoureuse), lesquelles freinent l'implémentation pratique de ces approches. La prise en compte de ces défis apportera des bénéfices considérables dans le **domaine de la santé**, notamment pour la **détection de chutes en temps réel** et la **reconnaissance d'identité humaine (HIDR)** à des fins de suivi *personnalisé* et d'interventions ciblées.

2 Problématique

Malgré son potentiel, la **HAR fondée sur la CSI** se heurte à deux grandes catégories de difficultés. Premièrement, les **contraintes liées aux données** — jeux de données réduits ou déséquilibrés, segments de données manquants et absence de signaux de référence ou classe '0' — compliquent la phase d'apprentissage et limitent la capacité de généralisation. Deuxièmement, les **contraintes liées aux modèles** — architectures complexes, surapprentissage, sensibilité aux hyperparamètres et manque de post-traitement rigoureux — nuisent à la fiabilité dans des conditions réelles. Ces écueils sont d'autant plus cruciaux dans le secteur de la santé, où les modèles doivent gérer des événements critiques (i.e. les chutes) tout en demeurant suffisamment légers pour s'exécuter **en temps réel** ou dans un environnement **à ressources limitées**, et suffisamment adaptables pour intégrer de nouveaux utilisateurs et contextes.

3 Questions de recherche

Compte tenu de ces problématiques, la thèse s'articule autour des questions-clés suivantes :

1. **Comment** concevoir des modèles de HAR capables de généraliser efficacement leur décision à partir de jeux de données CSI limités et déséquilibrés, tout en capturant des caractéristiques subtiles propres à chaque utilisateur (notamment pour l'HIDR) ?
2. **Quelles** architectures d'apprentissage profond et stratégies de calcul permettent une reconnaissance d'activités rapide et fiable, notamment dans des scénarios de détection de chutes urgentes ?
3. **Comment** développer et intégrer des systèmes de HAR personnalisés dans des applications de santé, de manière à faciliter des interventions ciblées ?

4 Objectifs et contributions

L'objectif général consiste à élaborer des solutions de HAR fondée sur la CSI qui soient à la fois **robustes, efficaces et adaptables** dans des contextes de santé. **La Partie II (Chapitres 3, 4 et 5)** répond aux limites relevées à la Partie I à travers trois contributions :

1. **HAR-LightCNN : une méthodologie de HAR par la CSI reposant sur un CNN allégé.**
   - **Problématique ciblée** : Les jeux de données restreints ou déséquilibrés, la complexité des modèles et l'absence d'évaluations cohérentes entravent la généralisation en temps réel de la HAR.
   - **Solution proposée** : Un réseau de neurones convolutifs (CNN) allégé, associé à une augmentation de données temporelles et à une évaluation normalisée, afin de gérer les bases de données limitées et le manque de standardisation dans les tests.
   - **Avantages et résultats** :
     - Haute précision en reconnaissance multi-activités avec un faible coût de calcul, adapté à la surveillance des patients en temps réel.
     - Atténuation du problème du « benchmarking inéquitable » grâce à des partitions entraînement/test strictes et une validation croisée.
     - Gains de performances notables par rapport à des modèles de référence, conciliant efficacité et robustesse.
2. **HFD-Adv : une méthodologie de détection de chutes en quatre étapes avancées.**
   - **Problématique ciblée** : Les segments manquants ou incomplets, l'absence de signaux de référence, le surapprentissage et l'absence de post-traitement entraînent une détection de chutes peu fiable.
   - **Solution proposée** : Un pipeline en 4 étapes intégrant un remplissage par seuil (pour les segments incomplets), la PCA (harmonisation des données), l'intégration de signaux sans activité, un réseau convolutif ActiNet muni d'une fonction de coût pondérée optimisée par algorithme génétique (pour pallier le surapprentissage et la sensibilité aux hyperparamètres), ainsi qu'un apprentissage par transfert pour les utilisateurs non vus et un ADM (Advanced Decision Mechanism) avec des seuils tenant compte de l'âge.
   - **Avantages et résultats** :
     - Taux de détection de chute (FDR) élevé dans plusieurs expérimentations, couplé à un faible taux de fausses alertes (FAR).

— Adaptation efficace à des utilisateurs totalement inconnus via l'apprentissage par transfert et l'augmentation de données, avec de bonnes performances sur plusieurs participants laissés en dehors de l'entraînement.

— Temps d'inférence de l'ordre de la milliseconde, qui soutient la réactivité requise pour les interventions d'urgence en milieu clinique.

3. **IdentNet : un modèle de reconnaissance d'identité basé sur un transformer.**

— **Problématique ciblée** : Similarité intra-classe (profils de chute analogues) et nécessité de gérer des dépendances à long terme au sein des signaux CSI.

— **Solution proposée** : Un identificateur d'utilisateur fondé sur un transformer, exploitant l'auto-attention pour distinguer des nuances propres à chaque individu à partir de l'ensemble de la séquence CSI, centré sur l'activité de chute.

— **Avantages et résultats** :

— Dépassement des approches CNN en HIDR, validant la pertinence du transformer pour différencier des utilisateurs aux dynamiques de chute proches.

— Extension possible à d'autres mouvements du quotidien (i.e. la course) moyennant peu de données supplémentaires, idéale pour les scénarios plus critiques de détection de chute.

— Dépassement des modèles avancés (WiTL, CLSTM, SimpleViTFi), renforçant la viabilité du transformer pour la reconnaissance d'identité en contexte de santé personnalisée.

Dans leur ensemble, ces méthodologies répondent directement aux freins liés aux données (petites, déséquilibrées, incomplètes) et aux modèles (architectures complexes, surapprentissage, absence de post-traitement), en simplifiant la mise en œuvre de la HAR par la CSI pour des besoins réels en santé.

5 Aperçu méthodologique

Les contributions de cette thèse mobilisent des stratégies complémentaires, comme suit :

— **HAR-LightCNN** :

— Recourt à l'**augmentation temporelle des données** pour prendre des décisions avec des jeux de données limités ou déséquilibrés en reconnaissance multi-activités.

— Met l'accent sur des **partitions entraînement-validation-test** scrupuleuses pour contrer le « benchmarking inéquitable ».

— Insiste sur un **CNN allégé** autorisant une exécution quasi temps réel.

— **HFD-Adv** :

— Se déploie via une **structure en 4 étapes** : **Prétraitement** (remplissage par seuil, PCA, calibration à l'aide de données sans activité) → **Traitement** (CNN ActiNet avec pondération via un algorithme génétique, LRFinder) → **Traitement amélioré** (apprentissage par transfert sur de nouveaux utilisateurs) → **Post-traitement** (ADM pour des décisions finales robustes).

— Corrige le problème des segments manquants, de la calibration spécifique à l'environnement et des faux négatifs potentiels.

— Déploie une évaluation sur plusieurs utilisateurs complètement nouveaux, validant une forte capacité d'adaptation.

— **IdentNet** :

— S'appuie sur un **transformer** doté d'**embeddings positionnels** et d'**auto-attentions multi-têtes** pour saisir les **dépendances de long terme** dans la CSI.

- Cible principalement la reconnaissance d'identité lors de chutes, mais peut s'adapter à d'autres activités (i.e. la course) avec une calibration minimale.
- Montre des améliorations régulières face aux bases CNN, en captant des indices propres à l'utilisateur à la fois sur le court et le long terme.

Les trois méthodes suivent des **indicateurs unifiés** pour évaluer **l'efficacité** (exactitude, rappel), **l'efficience** (charge de calcul) et **la polyvalence** (robustesse, adaptabilité).

6 Principaux résultats et évaluation

Les résultats obtenus dans les différentes contributions sont comme suit :

- **HAR-LightCNN** :
  - Affiche une forte précision en classification multi-activités, l'augmentation de données rehaussant notablement les F1-scores des classes minoritaires.
  - Démontre une stabilité accrue via de multiples partitions de validation, réglant ainsi les pratiques de test incohérentes antérieures.
  - Se révèle 2 à 3 fois plus efficiente (moins d'opérations en virgule flottante) et présente un gain d'environ 10% d'exactitude par rapport au 2D-CNN cité dans des travaux antérieurs.
  - Manifeste une complexité et une mémoire mobilisées moindres que celles de CNN allégés reconnus (MobileNet, SqueezeNet, ShuffleNet), confirmant son adéquation à la **surveillance de patients** en milieu intérieur en temps réel.
- **HFD-Adv** :
  - Assure une détection de chutes de haut niveau (train-test spécifiques par utilisateur) :
    - Sur un utilisateur laissé à part, atteint jusqu'à 100% de FDR pour un FAR proche de 3,4%.
    - Sur un autre utilisateur, maintient un FDR supérieur à 94%, confirmant une adaptation robuste.
  - Réduit significativement les chutes manquées, grâce à l'ADM, en ajustant les seuils selon des critères démographiques ou liés à l'âge.
  - Combine apprentissage par transfert et augmentation de données pour inclure de nouveaux utilisateurs, avec un temps d'inférence $< 1\,\text{ms}$, garantissant la **réactivité en situations d'urgence**.
- **IdentNet** :
  - Enregistre un gain de 8% par rapport à la HIDR reposant sur un CNN, validant l'apport du transformer pour gérer les longues dépendances et la similarité intra-classe.
  - Dépasse des références de pointe (SimpleViTFi, WiTL, CLSTM), conservant la meilleure exactitude en identification d'utilisateur.
  - Est préentraînable avec une activité simple (i.e. course) puis ajustable avec peu de données de chute, maintenant de bonnes performances sur l'activité initiale tout en améliorant l'identification en cas de chute.
  - Rend possible une **prise en charge personnalisée** en identifiant *qui* a chuté, critère décisif pour allouer rapidement les ressources et veiller à la sécurité du patient.

Au total, ces résultats valident la robustesse de la **CSI** comme **modalité de détection ambiante**, offrant une **excellente précision** pour les activités du quotidien, une **alerte rapide** en cas de chute et une **reconnaissance d'identité** fondée sur l'activité, le tout avec des contraintes de calcul maîtrisables.

7 Impact et perspectives

Grâce à une **surveillance non intrusive et respectueuse de la vie privée**, **HAR-LightCNN**, **HFD-Adv** et **IdentNet** forment un ensemble cohérent pour la **reconnaissance multi-activités**, la **détection d'urgence en cas de chute** et la **reconnaissance d'identité** dans le domaine de la santé :

— **Amélioration de la surveillance patient** : Des solutions légères et privilégiant la confidentialité, qui éliminent les freins liés au port de capteurs.

— **Efficacité des interventions d'urgence** : Une détection de chutes rapide et fiable (avec adaptation à l'utilisateur) diminue les risques de complications graves.

— **Santé personnalisée** : IdentNet favorise des stratégies individuelles, primordiales pour les populations à haut risque ; des tests supplémentaires montrent sa flexibilité sur diverses activités du quotidien.

Les perspectives futures incluent les idées suivantes :

— **Extension des activités** : Ajout de nouvelles activités quotidiennes ou cliniques, afin de renforcer la polyvalence du système.

— **Environnements multi-utilisateurs** : Suivi simultané de plusieurs individus dans un même lieu.

— **Exécution en périphérie (edge)** : Compression et optimisation accrues des modèles pour une exécution en temps réel sur des appareils à ressources restreintes.

— **Fusion de capteurs** : Combinaison avec d'autres modalités de mesure afin d'enrichir le contexte et d'augmenter la précision.

En poursuivant ces axes, la **HAR fondée sur la CSI** peut devenir une plateforme **évolutive et respectueuse de la vie privée** pour une **prise en charge intelligente et pilotée par les données**, garantissant un suivi en temps réel, la détection rapide des chutes et la reconnaissance d'identité dans une grande diversité de milieux cliniques et résidentiels.

List of Publications

Peer-Reviewed Journal Articles

H. El Zein, F. Mourad-Chehade, H. Amoud. "*CSI-Based Human Activity Recognition via Lightweight CNN Model and Data Augmentation*". IEEE Sensors Journal, vol. 24, no. 15, pp. 25060–25069, Aug. 1, 2024.
DOI: 10.1109/JSEN.2024.3414168

H. El Zein, F. Mourad-Chehade, H. Amoud. "*Advanced Fall Detection: A Four-Stage Deep Learning Approach Using Wi-Fi CSI*". IEEE Sensors Journal (submitted).

International Conference Articles

H. El Zein, F. Mourad-Chehade, H. Amoud. "*Intelligent Real-time Human Activity Recognition Using Wi-Fi Signals*". 2023 International Conference on Control, Automation and Diagnosis (ICCAD), Rome, Italy, pp. 1–5.
DOI: 10.1109/ICCAD57653.2023.10152429

H. El Zein, F. Mourad-Chehade, H. Amoud. "*Leveraging Wi-Fi CSI Data for Fall Detection: A Deep Learning Approach*". 2023 5th International Conference on Bio-engineering for Smart Technologies (BioSMART), Paris, France, pp. 1–4.
DOI: 10.1109/BioSMART58455.2023.10162090

H. El Zein, F. Mourad-Chehade, H. Amoud. "*Wi-Fi CSI-Based Fall Detection Using CNN with Advanced Decision Mechanism*". International Conference on e-Health and Bioengineering (EHB), Springer Nature Switzerland, pp. 685–693, Nov. 2023.
DOI: 10.1007/978-3-031-62502-2_77

National Conference Article

H. El Zein, F. Mourad-Chehade, H. Amoud. "*Reconnaissance Intelligente de l'activité Humaine en Temps Réel à l'aide de Signaux Wi-Fi*". Colloque en TéléSANté et dispositifs biomédicaux, Université Paris 8, CNRS, Jun 2023, Paris Saint Denis, France. hal-04220662.

Bibliography

[Ahmed, 2020] Hasmath Farhana Thariq Ahmed, Hafisoh Ahmad, Kulasekharan Narasingamurthi, Houda Harkat et Swee King Phang. « DF-WiSLR : device-free Wi-Fi-based sign language recognition ». *Pervasive and Mobile Computing* 69 (2020), p. 101289 (cf. p. 55).

[Alam, 2022] Ekram Alam, Abu Sufian, Paramartha Dutta et Marco Leo. « Vision-based human fall detection systems using deep learning : A review ». *Computers in biology and medicine* 146 (2022), p. 105626 (cf. p. 106, 107).

[Alzaabi, 2024] Aaesha Alzaabi, Tughrul Arslan et Nick Polydorides. *Wi-Fi CSI Sensing for Sleep Disturbances in the Care of Older People*. 2024 (cf. p. 55).

[Alzubaidi, 2021] Laith Alzubaidi, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma et al. « Review of deep learning : concepts, CNN architectures, challenges, applications, future directions ». *Journal of big Data* 8 (2021), p. 1-74 (cf. p. 50, 66, 68-70, 80, 83, 86).

[Arksey, 2005] Hilary Arksey et Lisa O'Malley. « Scoping studies : towards a methodological framework ». *International journal of social research methodology* 8.1 (2005), p. 19-32 (cf. p. 38).

[Arshad, 2022] Muhammad Haseeb Arshad, Muhammad Bilal et Abdullah Gani. « Human Activity Recognition : Review, Taxonomy and Open Challenges ». *Sensors* 22.17 (2022) (cf. p. 10, 11).

[Asghari Oskoei, 2007] Mohammadreza Asghari Oskoei et Huosheng Hu. « Myoelectric control systems—A survey ». *Biomedical Signal Processing and Control* 2.4 (2007), p. 275-294 (cf. p. 51, 129).

[Ashfaq, 2024] Nazish Ashfaq, Muhammad Hassan Khan et Muhammad Adeel Nisar. « Identification of Optimal Data Augmentation Techniques for Multimodal Time-Series Sensory Data : A Framework ». *Information* 15.6 (2024) (cf. p. 66).

[Australian Institute of Health and Welfare AIHW, 2023] Australian Institute of Health and Welfare (AIHW). *Falls*. https://www.aihw.gov.au/reports/injury/falls. 2023 (cf. p. 104).

[Ba, 2016] JL Ba. « Layer normalization ». *arXiv preprint arXiv :1607.06450* (2016) (cf. p. 154).

[Bai, 2018] Shaojie Bai, J Zico Kolter et Vladlen Koltun. « An empirical evaluation of generic convolutional and recurrent networks for sequence modeling ». *arXiv preprint arXiv :1803.01271* (2018) (cf. p. 51, 144, 145, 150, 153).

[Beyer, 2019] Dirk Beyer, Stefan Löwe et Philipp Wendler. « Reliable benchmarking : requirements and solutions ». *International Journal on Software Tools for Technology Transfer* 21.1 (2019), p. 1-29 (cf. p. 89, 90).

[Bian, 2024] Jichen Bian, Chong Tan, Peiyao Tang et Min Zheng. « Wi-Fi-Based Personnel Identity Recognition : Addressing Dataset Imbalance With C-DDPMs ». *IEEE Signal Processing Letters* 31 (2024), p. 1339-1343 (cf. p. 145, 146, 160, 161).

[Bibbò, 2023] Luigi Bibbò et Marley M. B. R. Vellasco. « Human Activity Recognition (HAR) in Healthcare ». *Applied Sciences* 13.24 (2023) (cf. p. 65).

[Brownlee, 2018] Jason Brownlee. *Deep learning for time series forecasting : predict the future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery, 2018 (cf. p. 51, 76, 78, 83, 84, 86).

[Cao, 2021] Yangjie Cao, Zhiyi Zhou, Chenxi Zhu, Pengsong Duan, Xianfu Chen et Jie Li. « A Lightweight Deep Learning Algorithm for WiFi-Based Identity Recognition ». *IEEE Internet of Things Journal* 8.24 (2021), p. 17449-17459 (cf. p. 146, 160).

[Cardenas, 2023] Jorge D. Cardenas, Carlos A. Gutierrez et Ruth Aguilar-Ponce. « Deep Learning Multi-Class Approach for Human Fall Detection Based on Doppler Signatures ». *International Journal of Environmental Research and Public Health* 20.2 (2023) (cf. p. 49, 133).

[Centers for Disease Control and Prevention, 2021] Centers for Disease Control and Prevention. *Injuries Among Children and Teens*. https://www.cdc.gov/injury/features/child-injury/index.html. 2021 (cf. p. 104, 130).

[Centers for Disease Control and Prevention, 2023a] Centers for Disease Control and Prevention. *Facts About Falls*. https://www.cdc.gov/falls/facts.html. 2023 (cf. p. 105, 130, 144).

[Centers for Disease Control and Prevention, 2023b] Centers for Disease Control and Prevention. *Keep on Your Feet—Preventing Older Adult Falls*. https://www.cdc.gov/injury/features/older-adult-falls/index.html. 2023 (cf. p. 104, 105).

[Centers for Disease Control and Prevention, 2023c] Centers for Disease Control and Prevention. *Older Adult Falls Data*. https://www.cdc.gov/falls/data/index.html. 2023 (cf. p. 104-106, 135, 162).

[Chai, 2021] Xiaoqing Chai, Renjie Wu, Matthew Pike, Hangchao Jin, Wan-Young Chung et Boon-Giin Lee. « Smart wearables with sensor fusion for fall detection in firefighting ». *Sensors* 21.20 (2021), p. 6770 (cf. p. 49).

[Chakraborty, 2024] Shalini Chakraborty et Grischa Liebel. « Modelling guidance in software engineering : a systematic literature review ». *Software and Systems Modeling* 23.1 (2024), p. 249-265 (cf. p. 38).

[Chaudhary, 2018] S. Chaudhary, J. Figueroa, S. Shaikh, E.W. Mays, R. Bayakly, M. Javed et al. « Pediatric falls ages 0–4 : understanding demographics, mechanisms, and injury severities ». *Injury Epidemiology* 5.7 (2018) (cf. p. 130).

[Chawla, 2013] Nitesh V Chawla et Darcy A Davis. « Bringing big data to personalized healthcare : a patient-centered framework ». *Journal of general internal medicine* 28 (2013), p. 660-665 (cf. p. 143).

[Chen, 2023a] Chen Chen, Gang Zhou et Youfang Lin. « Cross-Domain WiFi Sensing with Channel State Information : A Survey ». *ACM Comput. Surv.* 55.11 (2023) (cf. p. 14-16, 19-24, 31, 35, 50, 54, 72, 127).

[Chen, 2023b] Hui-Hsin Chen, Chi-Lun Lin et Chun-Hsiang Chang. « WiFi-based detection of human subtle motion for health applications ». *Bioengineering* 10.2 (2023), p. 228 (cf. p. 55).

[Chen, 2023c] Sheng Chen, Wei Yang, Yang Xu, Yangyang Geng, Bangzhou Xin et Liusheng Huang. « AFall : Wi-Fi-Based Device-Free Fall Detection System Using Spatial Angle of Arrival ». *IEEE Transactions on Mobile Computing* 22.8 (2023), p. 4471-4484 (cf. p. 49).

[Chen, 2022] Shih-Yuan Chen et Chi-Lun Lin. « Subtle Motion Detection Using Wi-Fi for Hand Rest Tremor in Parkinson's Disease ». *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2022, p. 1774-1777 (cf. p. 55).

[Chen, 2019] Zhenghua Chen, Le Zhang, Chaoyang Jiang, Zhiguang Cao et Wei Cui. « WiFi CSI Based Passive Human Activity Recognition Using Attention Based BLSTM ». *IEEE Transactions on Mobile Computing* 18.11 (2019), p. 2714-2724 (cf. p. 47, 97).

[Chu, 2023] Yi Chu, Kanapathippillai Cumanan, Sathish K. Sankarpandi, Stephen Smith et Octavia A. Dobre. « Deep Learning-Based Fall Detection Using WiFi Channel State Information ». *IEEE Access* 11 (2023), p. 83763-83780 (cf. p. 49).

[Chung, 2024] Jane Chung, Ingrid Pretzer-Aboff, Pamela Parsons, Katherine Falls, Eyuphan Bulut et al. « Using a Device-Free Wi-Fi Sensing System to Assess Daily Activities and Mobility in Low-Income Older Adults : Protocol for a Feasibility Study ». *JMIR Research Protocols* 13.1 (2024), e53447 (cf. p. 55).

[Custance, 2023] Oliver Custance, Saad Khan et Simon Parkinson. « Classifying Participant Standing and Sitting Postures Using Channel State Information ». *Electronics* 12.21 (2023) (cf. p. 143).

[Demrozi, 2020] Florenc Demrozi, Graziano Pravadelli, Azra Bihorac et Parisa Rashidi. « Human Activity Recognition Using Inertial, Physiological and Environmental Sensors : A Comprehensive Survey ». *IEEE Access* 8 (2020), p. 210816-210836 (cf. p. 17, 65).

[Demrozi, 2023] Florenc Demrozi, Cristian Turetta, Alejandro Masrur, Martin Schmidhammer, Christian Gentner, Samarjit Chakraborty et al. « A Dataset on CSI-based Activity Recognition in Real-World Environments ». *Authorea Preprints* (2023) (cf. p. 25).

[Deng, 2023] Lang Deng, Jianfei Yang, Shenghai Yuan, Han Zou, Chris Xiaoxuan Lu et Lihua Xie. « GaitFi : Robust Device-Free Human Identification via WiFi and Vision Multimodal Learning ». *IEEE Internet of Things Journal* 10.1 (2023), p. 625-636 (cf. p. 143).

[Diaz, 2023] Guillermo Diaz, Iker Sobron, Iñaki Eizmendi, Iratxe Landa, Johana Coyote et Manuel Velez. « Channel phase processing in wireless networks for human activity recognition ». *Internet of Things* 24 (2023), p. 100960 (cf. p. 15, 16).

[Disease Control, 2023] Centers for Disease Control et Prevention. *Information on Safety in the Home & Community for Parents with Infants & Toddlers (Ages 0-3)*. https://www.cdc.gov/parents/infants/safety.html. 2023 (cf. p. 130).

[Dong, 2018] Zehua Dong, Fangmin Li, Julang Ying et Kaveh Pahlavan. « Indoor motion detection using Wi-Fi channel state information in flat floor environments versus in staircase environments ». *Sensors (Switzerland)* 18.7 (2018) (cf. p. 48).

[Duan, 2023] Pengsong Duan, Xianguang Diao, Yangjie Cao, Dalong Zhang, Bo Zhang et Jinsheng Kong. « A Comprehensive Survey on Wi-Fi Sensing for Human Identity Recognition ». *Electronics* 12.23 (2023), p. 4858 (cf. p. 99, 100, 143, 144, 149, 163, 164).

[Fard Moshiri, 2021] Parisa Fard Moshiri, Reza Shahbazian, Mohammad Nabati et Seyed Ali Ghorashi. « A CSI-Based Human Activity Recognition Using Deep Learning ». *Sensors* 21.21 (2021) (cf. p. 26, 27, 47, 65, 72-74, 76, 78, 80, 90, 93, 94, 97, 115, 131, 133).

[Fernández, 2018] Alberto Fernández, Salvador García, Mikel Galar, Ronaldo C Prati, Bartosz Krawczyk et Francisco Herrera. *Learning from imbalanced data sets*. T. 10. 2018. Springer, 2018 (cf. p. 50, 66, 76, 81, 83).

[Ferrari, 2015] Rossella Ferrari. « Writing narrative style literature reviews ». *Medical writing* 24.4 (2015), p. 230-235 (cf. p. 38).

[Ganaie, 2022] M.A. Ganaie, Minghui Hu, A.K. Malik, M. Tanveer et P.N. Suganthan. « Ensemble deep learning : A review ». *Engineering Applications of Artificial Intelligence* 115 (2022), p. 105151 (cf. p. 51, 129).

[García-Laencina, 2010] Pedro J García-Laencina, José-Luis Sancho-Gómez et Aníbal R Figueiras-Vidal. « Pattern classification with missing data : a review ». *Neural Computing and Applications* 19 (2010), p. 263-282 (cf. p. 118).

[Ge, 2023] Yao Ge, Ahmad Taha, Syed Aziz Shah, Kia Dashtipour, Shuyuan Zhu, Jonathan Cooper et al. « Contactless WiFi Sensing and Monitoring for Future Healthcare - Emerging Trends, Challenges, and Opportunities ». *IEEE Reviews in Biomedical Engineering* 16 (2023), p. 171-191 (cf. p. 10, 11, 17, 31, 33, 35, 55, 72, 98, 99, 109, 110, 135).

[Goodfellow, 2016] Ian Goodfellow, Yoshua Bengio et Aaron Courville. *Deep learning*. MIT press, 2016 (cf. p. 35, 51, 68-70, 76, 78, 80, 81, 83, 84, 86, 91, 121, 122, 124, 128, 144, 145, 150, 153).

[Grant, 2009] Maria J Grant et Andrew Booth. « A typology of reviews : an analysis of 14 review types and associated methodologies ». *Health information & libraries journal* 26.2 (2009), p. 91-108 (cf. p. 38, 39).

[Gringoli, 2019] Francesco Gringoli, Matthias Schulz, Jakob Link et Matthias Hollick. « Free your CSI : A channel state information extraction platform for modern Wi-Fi chipsets ». *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*. 2019, p. 21-28 (cf. p. 25, 74).

[Gu, 2021] Yu Gu, Huan Yan, Mianxiong Dong, Meng Wang, Xiang Zhang, Zhi Liu et al. « WiONE : One-Shot Learning for Environment-Robust Device-Free User Authentication via Commodity Wi-Fi in Man–Machine System ». *IEEE Transactions on Computational Social Systems* 8.3 (2021), p. 630-642 (cf. p. 146).

[Guo, 2023] Runhao Guo, Heng Li, Dongliang Han et Runze Liu. « Feasibility Analysis of Using Channel State Information (CSI) Acquired from Wi-Fi Routers for Construction Worker Fall Detection ». *International Journal of Environmental Research and Public Health* 20.6 (2023) (cf. p. 49, 133).

[Guo, 2018] Xiaonan Guo, Jian Liu, Cong Shi, Hongbo Liu, Yingying Chen et Mooi Choo Chuah. « Device-free personalized fitness assistant using WiFi ». *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.4 (2018), p. 1-23 (cf. p. 55).

[Gupta, 2024] Chhaya Gupta, Nasib Singh Gill, Preeti Gulia, Sangeeta Yadav, Giovanni Pau, Mohammad Alibakhshikenari et al. « A Real-Time 3-Dimensional Object Detection Based Human Action Recognition Model ». *IEEE Open Journal of the Computer Society* 5 (2024), p. 14-26 (cf. p. 17, 65, 98, 99).

[Halperin, 2011] Daniel Halperin, Wenjun Hu, Anmol Sheth et David Wetherall. « Tool release : Gathering 802.11 n traces with channel state information ». *ACM SIGCOMM computer communication review* 41.1 (2011), p. 53-53 (cf. p. 25, 28).

[Han, 2014] Chunmei Han, Kaishun Wu, Yuxi Wang et Lionel M. Ni. « WiFall : Device-free fall detection by wireless networks ». *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*. 2014, p. 271-279 (cf. p. 47).

[Hastie, 2009] Trevor Hastie, Robert Tibshirani, Jerome H Friedman et Jerome H Friedman. *The elements of statistical learning : data mining, inference, and prediction*. T. 2. Springer, 2009 (cf. p. 51).

[He, 2009] Haibo He et Edwardo A. Garcia. « Learning from Imbalanced Data ». *IEEE Transactions on Knowledge and Data Engineering* 21.9 (2009), p. 1263-1284 (cf. p. 122).

[He, 2020] Ying He, Yan Chen, Yang Hu et Bing Zeng. « WiFi Vision : Sensing, Recognition, and Detection With Commodity MIMO-OFDM WiFi ». *IEEE Internet of Things Journal* 7.9 (2020), p. 8296-8317 (cf. p. 11, 15, 17, 33, 35, 106).

[Howard, 2019] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan et al. « Searching for mobilenetv3 ». *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, p. 1314-1324 (cf. p. 88).

[Hu, 2022] Yuqian Hu, Feng Zhang, Chenshu Wu, Beibei Wang et K. J. Ray Liu. « DeFall : Environment-Independent Passive Fall Detection Using WiFi ». *IEEE Internet of Things Journal* 9.11 (2022), p. 8515-8530 (cf. p. 48, 49).

[Iandola, 2016] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally et Kurt Keutzer. « SqueezeNet : AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size ». *arXiv preprint arXiv :1602.07360* (2016) (cf. p. 88).

[IEEE, 2021] IEEE. « IEEE Standard for Information Technology–Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks–Specific Requirements - Part 11 : Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications ». *IEEE Std 802.11-2020 (Revision of IEEE Std 802.11-2016)* (2021), p. 1-4379 (cf. p. 13, 20-24).

[Imanzadeh, 2024] Soodabeh Imanzadeh, Jafar Tanha et Mahdi Jalili. « Ensemble of deep learning techniques to human activity recognition using smart phone signals ». *Multimedia Tools and Applications* (2024), p. 1-30 (cf. p. 150).

[Insights, 2024] Fortune Business Insights. *Gesture Recognition Market Size, Share & Forecast*. 2024 (cf. p. 10).

[Institute, 2016] ECRI Institute. « Patient Identification : Executive Summary ». *HealthManagement* 16.4 (2016) (cf. p. 143).

[Iwana, 2021a] Brian Kenji Iwana et Seiichi Uchida. « An empirical survey of data augmentation for time series classification with neural networks ». *Plos one* 16.7 (2021), e0254841 (cf. p. 66-68, 80, 83).

[Iwana, 2021b] Brian Kenji Iwana et Seiichi Uchida. « Time Series Data Augmentation for Neural Networks by Time Warping with a Discriminative Teacher ». *2020 25th International Conference on Pattern Recognition (ICPR)*. 2021, p. 3558-3565 (cf. p. 67, 68).

[Kahn, 2020] Nicole F Kahn et Robert Graham. *Promoting positive adolescent health behaviors and outcomes : Thriving in the 21st century*. National Academies Press, 2020 (cf. p. 130).

[Khan, 2021] Muhammad Imran Khan, Mian Ahmad Jan, Yar Muhammad, Dinh-Thuan Do, Ateeq ur Rehman, Constandinos X Mavromoustakis et al. « Tracking vital signs of a patient using channel state information and machine learning for a smart healthcare system ». *Neural Computing and Applications* (2021), p. 1-15 (cf. p. 143).

[Kingma, 2014] Diederik P Kingma et Jimmy Ba. « Adam : A method for stochastic optimization ». *arXiv preprint arXiv :1412.6980* (2014) (cf. p. 124).

[Klin, 2018] R. Klin. *At What Age Do We Peak ?* https://www.wmchealth.org/blogs/at-what-age-do-we-peak-98. 2018 (cf. p. 130).

[LeCun, 2015] Yann LeCun, Yoshua Bengio et Geoffrey Hinton. « Deep learning ». *nature* 521.7553 (2015), p. 436-444 (cf. p. 51).

[Lee, 2022] Seunghee Lee, Bummo Koo, Sumin Yang, Jongman Kim, Yejin Nam et Youngho Kim. « Fall-from-height detection using deep learning based on IMU sensor data for accident prevention at construction sites ». *Sensors* 22.16 (2022), p. 6107 (cf. p. 49).

[Limpanadusadee, 2018] Jarinya Limpanadusadee, Panyarat Kesawattana, Thitiwud Wongsawat et Damras Wongsawang. « EldTec : Improvement on Wearable Sensor for Elderly Fall Detection ». *2018 Seventh ICT International Student Project Conference (ICT-ISPC)*. 2018, p. 1-6 (cf. p. 49).

[Lin, 2023] Chi Lin, Pengfei Wang, Chuanying Ji, Mohammad S Obaidat, Lei Wang, Guowei Wu et al. « A contactless authentication system based on WiFi CSI ». *ACM Transactions on Sensor Networks* 19.2 (2023), p. 1-20 (cf. p. 146).

[Liu, 2019] Meiguang Liu, Lei Zhang, Panlong Yang, Liangfu Lu et Liangyi Gong. « Wi-Run : Device-free step estimation system with commodity Wi-Fi ». *Journal of Network and Computer Applications* 143 (2019), p. 77-88 (cf. p. 55).

[Loshchilov, 2017] I Loshchilov. « Decoupled weight decay regularization ». *arXiv preprint arXiv :1711.05101* (2017) (cf. p. 156).

[Lv, 2020] Tianqi Lv, Xiaojuan Wang, Lei Jin, Yabo Xiao et Mei Song. « Margin-Based Deep Learning Networks for Human Activity Recognition ». *Sensors* 20.7 (2020) (cf. p. 150).

[Ma, 2019] Yongsen Ma, Gang Zhou et Shuangquan Wang. « WiFi sensing with channel state information : A survey ». *ACM Computing Surveys (CSUR)* 52.3 (2019), p. 1-36 (cf. p. 14-16, 19-25, 31, 84, 106, 110, 117, 136).

[Minvielle, 2021] Etienne Minvielle, Aude Fourcade, Thomas Ricketts et Mathias Waelli. « Current developments in delivering customized care : a scoping review ». *BMC Health Services Research* 21 (2021), p. 1-29 (cf. p. 143, 164).

[Mo, 2021] Hyunggeun Mo et Seungku Kim. « A Deep Learning-Based Human Identification System With Wi-Fi CSI Data Augmentation ». *IEEE Access* 9 (2021), p. 91913-91920 (cf. p. 146, 160).

[Mobasheri, 2022] Bahareh Mobasheri, Seyed Reza Kamel Tabbakh et Yahya Forghani. « An approach for fall prediction based on kinematics of body key points using LSTM ». *International journal of environmental research and public health* 19.21 (2022), p. 13762 (cf. p. 49).

[Montavon, 2012] Grégoire Montavon, Geneviève Orr et Klaus-Robert Müller. *Neural networks : tricks of the trade*. T. 7700. springer, 2012 (cf. p. 53, 54, 86, 91).

[Munn, 2018] Zachary Munn, Micah DJ Peters, Cindy Stern, Catalin Tufanaru, Alexa McArthur et Edoardo Aromataris. « Systematic review or scoping review ? Guidance for authors when choosing between a systematic or scoping review approach ». *BMC medical research methodology* 18 (2018), p. 1-7 (cf. p. 38, 39).

[Nakamura, 2020] Takashi Nakamura, Mondher Bouazizi, Kohei Yamamoto et Tomoaki Ohtsuki. « Wi-Fi-CSI-based Fall Detection by Spectrogram Analysis with CNN ». *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*. 2020, p. 1-6 (cf. p. 49).

[Nakamura, 2022] Takashi Nakamura, Mondher Bouazizi, Kohei Yamamoto et Tomoaki Ohtsuki. « Wi-Fi-based Fall Detection Using Spectrogram Image of Channel State Information ». *IEEE Internet of Things Journal* (2022) (cf. p. 48).

[National Institute on Aging, 2022] National Institute on Aging. *Falls and Fractures in Older Adults : Causes and Prevention*. https://www.nia.nih.gov/health/falls-and-falls-prevention/falls-and-fractures-older-adults-causes-and-prevention. 2022 (cf. p. 105, 106, 109, 135, 136).

[Ng, 2017] Andrew Ng. « Machine learning yearning ». *URL : http ://www. mlyearning. org/(96)* 139 (2017), p. 30 (cf. p. 53, 91).

[NHS England, 2023a] NHS England. *Delivery Plan for Recovering Urgent and Emergency Care Services*. 2023 (cf. p. 105, 106, 135-137).

[NHS England, 2023b] NHS England. *Major plan to recover urgent and emergency care services*. https://www.england.nhs.uk/2023/01/major-plan-to-recover-urgent-and-emergency-care-services/. 2023 (cf. p. 105).

[Occupational Safety, 2022] National Institute for Occupational Safety, Centers for Disease Control Health (NIOSH) et Prevention (CDC). *Falls in the Workplace*. https://www.cdc.gov/niosh/topics/falls/default.html. 2022 (cf. p. 105, 135).

[Palipana, 2018] Sameera Palipana, David Rojas, Piyush Agrawal et Dirk Pesch. « FallDeFi : Ubiquitous fall detection using commodity Wi-Fi devices ». *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1.4 (2018), p. 1-25 (cf. p. 48, 49).

[Petersen, 2015] Kai Petersen, Sairam Vakkalanka et Ludwik Kuzniarz. « Guidelines for conducting systematic mapping studies in software engineering : An update ». *Information and Software Technology* 64 (2015), p. 1-18 (cf. p. 38-40).

[Proakis, 2007] John G Proakis. *Digital signal processing : principles, algorithms, and applications, 4/E*. Pearson Education India, 2007 (cf. p. 33).

[Proakis, 2008] John G Proakis et Masoud Salehi. *Digital communications*. McGraw-hill, 2008 (cf. p. 13-16, 18-24).

[Ramezani, 2018] Ramin Ramezani, Yubin Xiao et Arash Naeim. « Sensing-Fi : Wi-Fi CSI and accelerometer fusion system for fall detection ». *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE. 2018, p. 402-405 (cf. p. 49).

[Rappaport, 2024] Theodore S Rappaport. *Wireless communications : principles and practice*. Cambridge University Press, 2024 (cf. p. 13-16, 18-24).

[Sadouk, 2018] Lamyaa Sadouk. « CNN Approaches for Time Series Classification ». *Time Series Analysis*. Sous la dir. de Chun-Kit Ngan. Rijeka : IntechOpen, 2018. Chap. 4 (cf. p. 68-70, 83, 84, 86).

[Sandler, 2018] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov et Liang-Chieh Chen. « Mobilenetv2 : Inverted residuals and linear bottlenecks ». *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, p. 4510-4520 (cf. p. 88).

[Schäfer, 2021a] Jörg Schäfer, Baldev Raj Barrsiwal, Muyassar Kokhkharova, Hannan Adil et Jens Liebehenschel. « Human Activity Recognition Using CSI Information with Nexmon ». *Applied Sciences* 11.19 (2021) (cf. p. 26, 28, 30, 116).

[Schäfer, 2021b] Jörg Schäfer, Baldev Raj Barrsiwal, Muyassar Kokhkharova, Hannan Adil et Jens Liebehenschel. « Human activity recognition using csi information with nexmon ». *Applied Sciences (Switzerland)* 11.19 (2021) (cf. p. 48, 133).

[Shi, 2022] Zhenguo Shi, J. Andrew Zhang, Richard Yida Xu et Qingqing Cheng. « Environment-Robust Device-Free Human Activity Recognition With Channel-State-Information Enhancement and One-Shot Learning ». *IEEE Transactions on Mobile Computing* 21.2 (2022), p. 540-554 (cf. p. 47).

[Singh, 2020] Anuradha Singh, Saeed Ur Rehman, Sira Yongchareon et Peter Han Joo Chong. « Sensor Technologies for Fall Detection Systems : A Review ». *IEEE Sensors Journal* 20.13 (2020), p. 6889-6919 (cf. p. 106, 107).

[Snyder, 2019] Hannah Snyder. « Literature review as a research methodology : An overview and guidelines ». *Journal of business research* 104 (2019), p. 333-339 (cf. p. 38).

[Sokolova, 2009] Marina Sokolova et Guy Lapalme. « A systematic analysis of performance measures for classification tasks ». *Information processing & management* 45.4 (2009), p. 427-437 (cf. p. 156).

[Strohmayer, 2024] Julian Strohmayer et Martin Kampel. « Data Augmentation Techniques for Cross-Domain WiFi CSI-Based Human Activity Recognition ». *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer. 2024, p. 42-56 (cf. p. 65).

[Sultana, 2021] Arifa Sultana, Kaushik Deb, Pranab Kumar Dhar et Takeshi Koshiba. « Classification of indoor human fall events using deep learning ». *Entropy* 23.3 (2021), p. 328 (cf. p. 49).

[Tan, 2018] Bo Tan, Qingchao Chen, Kevin Chetty, Karl Woodbridge, Wenda Li et Robert Piechocki. « Exploiting WiFi Channel State Information for Residential Healthcare Informatics ». *IEEE Communications Magazine* 56.5 (2018), p. 130-137 (cf. p. 10, 11, 17, 55, 65, 72, 98, 99, 143, 149).

[Thanh Noi, 2018] Phan Thanh Noi et Martin Kappas. « Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery ». *Sensors* 18.1 (2018) (cf. p. 51).

[Touhiduzzaman, 2025] Md Touhiduzzaman, Steven M Hernandez, Peter E Pidcoe et Eyuphan Bulut. « Wi-PT-Hand : Wireless Sensing based Low-cost Physical Rehabilitation Tracking for Hand Movements ». *ACM Transactions on Computing for Healthcare* 6.1 (2025), p. 1-25 (cf. p. 55).

[Trusts, 2022] The Pew Charitable Trusts. *Biometrics Can Help Match Patients to Their Electronic Health Records*. The Pew Charitable Trusts, 2022 (cf. p. 143).

[Varga, 2024] Domonkos Varga. « Critical Analysis of Data Leakage in WiFi CSI-Based Human Action Recognition Using CNNs ». *Sensors* 24.10 (2024) (cf. p. 52, 53, 76, 78, 90, 91).

[Vaswani, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez et al. « Attention is all you need. Advances in neural information processing systems ». *Advances in neural information processing systems* 30.2017 (2017) (cf. p. 144, 149, 150, 152, 154).

[Vetterli, 2014] Martin Vetterli, Jelena Kovačević et Vivek K Goyal. *Foundations of signal processing*. Cambridge University Press, 2014 (cf. p. 33).

[Vicente, 2020] Astrid M Vicente, Wolfgang Ballensiefen et Jan-Ingvar Jönsson. « How personalised medicine will transform healthcare by 2030 : the ICPerMed vision ». *Journal of Translational Medicine* 18 (2020), p. 1-4 (cf. p. 143, 149, 163, 164).

[Wahid, 2020] Md Ferdous Wahid, Reza Tafreshi et Reza Langari. « A Multi-Window Majority Voting Strategy to Improve Hand Gesture Recognition Accuracies Using Electromyography Signal ». *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28.2 (2020), p. 427-436 (cf. p. 51, 129).

[Wang, 2024] Chao Wang, Yinfan Ding, Meng Zhou et Lin Tang. « Wi-Fi sensing gesture control algorithm based on semi-supervised generative adversarial network ». *PeerJ Computer Science* 10 (2024), e2402 (cf. p. 55).

[Wang, 2022] Chao Wang, Lin Tang, Meng Zhou, Yinfan Ding, Xueyong Zhuang et Jie Wu. « Indoor Human Fall Detection Algorithm Based on Wireless Sensing ». *Tsinghua Science and Technology* 27.6 (2022), p. 1002-1015 (cf. p. 48).

[Wang, 2019] Fei Wang, Jianwei Feng, Yinliang Zhao, Xiaobin Zhang, Shiyuan Zhang et Jinsong Han. « Joint Activity Recognition and Indoor Localization With WiFi Fingerprints ». *IEEE Access* 7 (2019), p. 80058-80068 (cf. p. 47, 57).

[Wang, 2017a] Hao Wang, Daqing Zhang, Yasha Wang, Junyi Ma, Yuxiang Wang et Shengjie Li. « RT-Fall : A Real-Time and Contactless Fall Detection System with Commodity WiFi Devices ». *IEEE Transactions on Mobile Computing* 16.2 (2017), p. 511-526 (cf. p. 48, 49).

[Wang, 2021a] Yongchuan Wang, Song Yang, Fan Li, Yue Wu et Yu Wang. « FallViewer : A Fine-Grained Indoor Fall Detection System with Ubiquitous Wi-Fi Devices ». *IEEE Internet of Things Journal* 8.15 (2021), p. 12455-12466 (cf. p. 48).

[Wang, 2017b] Yuxi Wang, Kaishun Wu et Lionel M. Ni. « WiFall : Device-Free Fall Detection by Wireless Networks ». *IEEE Transactions on Mobile Computing* 16.2 (2017), p. 581-594 (cf. p. 47-49).

[Wang, 2021b] Zhengjie Wang, Wenwen Dou, Mingjing Ma, Xiaoxue Feng, Zehua Huang, Chengming Zhang et al. « A survey of user authentication based on channel state information ». *Wireless Communications and Mobile Computing* 2021.1 (2021), p. 6636665 (cf. p. 144).

[Wang, 2014] Zhi Wang, Jinsong Han, Wei Xi et Jizhong Zhao. « Efficient and secure key extraction using channel state information ». *The Journal of Supercomputing* 70 (2014), p. 1537-1554 (cf. p. 143).

[Wang, 2018] Zhu Wang, Bin Guo, Zhiwen Yu et Xingshe Zhou. « Wi-Fi CSI-Based Behavior Recognition : From Signals and Actions to Activities ». *IEEE Communications Magazine* 56.5 (2018), p. 109-115 (cf. p. 10, 11, 15, 55, 65, 72, 99).

[Ware, 2018] Shweta Ware, Chaoqun Yue, Reynaldo Morillo, Jin Lu, Chao Shang, Jayesh Kamath et al. « Large-scale automatic depression screening using meta-data from wifi infrastructure ». *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.4 (2018), p. 1-27 (cf. p. 55).

[Weber, 2019] Lukas M Weber, Wouter Saelens, Robrecht Cannoodt, Charlotte Soneson, Alexander Hapfelmeier, Paul P Gardner et al. « Essential guidelines for computational method benchmarking ». *Genome biology* 20 (2019), p. 1-12 (cf. p. 52, 89, 90).

[Wilson, 2023] Bryan R. Wilson et Ashley Woodrow. « Patient Safety in Emergency Medical Services ». *Contemporary Topics in Patient Safety*. Sous la dir. de Philip N. Salen et Stanislaw P. Stawicki. Rijeka : IntechOpen, 2023. Chap. 4 (cf. p. 143).

[World Health Organization, 2021] World Health Organization. *Falls*. https://www.who.int/news-room/fact-sheets/detail/falls. 2021 (cf. p. 104, 105, 130, 135, 137, 144, 162).

[Xia, 2023] Zhengxin Xia et Su Chong. « WiFi-based indoor passive fall detection for medical Internet of Things ». *Computers and Electrical Engineering* 109 (2023) (cf. p. 49).

[Xiao, 2019] Chunjing Xiao, Daojun Han, Yongsen Ma et Zhiguang Qin. « CsiGAN : Robust Channel State Information-Based Activity Recognition With GANs ». *IEEE Internet of Things Journal* 6.6 (2019), p. 10191-10204 (cf. p. 47, 72).

[Xie, 2015] Yaxiong Xie, Zhenjiang Li et Mo Li. « Precise power delay profiling with commodity WiFi ». *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. 2015, p. 53-64 (cf. p. 25).

[Xu, 2020] Yuhui Xu, Lingxi Xie, Xiaopeng Zhang, Xin Chen, Bowen Shi, Qi Tian et al. « Latency-aware differentiable neural architecture search ». *arXiv preprint arXiv :2001.06392* (2020) (cf. p. 53).

[Yang, 2023a] Jianfei Yang, Xinyan Chen, Han Zou, Chris Xiaoxuan Lu, Dazhuo Wang, Sumei Sun et al. « SenseFi : A library and benchmark on deep-learning-empowered WiFi human sensing ». *Patterns* 4.3 (2023) (cf. p. 52-54, 68, 83, 86, 90, 91).

[Yang, 2020] Li Yang et Abdallah Shami. « On hyperparameter optimization of machine learning algorithms : Theory and practice ». *Neurocomputing* 415 (2020), p. 295-316 (cf. p. 123).

[Yang, 2023b] Zheng Yang, Yi Zhang et Qian Zhang. « Rethinking Fall Detection With Wi-Fi ». *IEEE Transactions on Mobile Computing* 22.10 (2023), p. 6126-6143 (cf. p. 49).

[Yi, 2017] Subin Yi, Janghoon Ju, Man-Ki Yoon et Jaesik Choi. « Grouped convolutional neural networks for multivariate time series ». *arXiv preprint arXiv :1703.09938* (2017) (cf. p. 68-70, 83, 84).

[Yosinski, 2014] Jason Yosinski, Jeff Clune, Yoshua Bengio et Hod Lipson. « How transferable are features in deep neural networks ? » *Advances in neural information processing systems* 27 (2014) (cf. p. 128).

[Yousefi, 2017] Siamak Yousefi, Hirokazu Narui, Sankalp Dayal, Stefano Ermon et Shahrokh Valaee. « A Survey on Behavior Recognition Using WiFi Channel State Information ». *IEEE Communications Magazine* 55.10 (2017), p. 98-104 (cf. p. 15, 16, 26, 28, 29, 75, 76, 79, 81, 90, 93, 95-97).

[Zeng, 2016] Yunze Zeng, Parth H. Pathak et Prasant Mohapatra. « WiWho : WiFi-Based Person Identification in Smart Spaces ». *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 2016, p. 1-12 (cf. p. 145).

[Zerveas, 2021] George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty et Carsten Eickhoff. « A transformer-based framework for multivariate time series representation learning ». *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*. 2021, p. 2114-2124 (cf. p. 144, 149, 150).

[Zhang, 2018a] Bowen Zhang, Limin Wang, Zhe Wang, Yu Qiao et Hanli Wang. « Real-Time Action Recognition With Deeply Transferred Motion Vector CNNs ». *IEEE Transactions on Image Processing* 27.5 (2018), p. 2326-2339 (cf. p. 17, 65, 98, 99).

[Zhang, 2015] Daqing Zhang, Hao Wang, Yasha Wang et Junyi Ma. « Anti-fall : A non-intrusive and real-time fall detector leveraging CSI from commodity WiFi devices ». *Inclusive Smart Cities and e-Health : 13th International Conference on Smart Homes and Health Telematics, ICOST 2015, Geneva, Switzerland, June 10-12, 2015, Proceedings 13*. Springer. 2015, p. 181-193 (cf. p. 49).

[Zhang, 2018b] Feng Zhang, Chen Chen, Beibei Wang et K. J. Ray Liu. « WiSpeed : A Statistical Electromagnetic Approach for Device-Free Indoor Speed Estimation ». *IEEE Internet of Things Journal* 5.3 (2018), p. 2163-2177 (cf. p. 48).

[Zhang, 2004] Harry Zhang. « The optimality of naive Bayes ». *Aa* 1.2 (2004), p. 3 (cf. p. 51).

[Zhang, 2022] Jie Zhang, Zhongmin Wang et QingLi Yan. « Intelligent user identity authentication in vehicle security system based on wireless signals ». *Complex & Intelligent Systems* 8.2 (2022), p. 1243-1257 (cf. p. 146).

[Zhang, 2016] Jin Zhang, Bo Wei, Wen Hu et Salil S. Kanhere. « WiFi-ID : Human Identification Using WiFi Signal ». *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)*. 2016, p. 75-82 (cf. p. 145).

[Zhang, 2018c] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin et Jian Sun. « Shufflenet : An extremely efficient convolutional neural network for mobile devices ». *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, p. 6848-6856 (cf. p. 88).

[Zhao, 2017] Bendong Zhao, Huanzhang Lu, Shangfeng Chen, Junliang Liu et Dongya Wu. « Convolutional neural networks for time series classification ». *Journal of Systems Engineering and Electronics* 28.1 (2017), p. 162-169 (cf. p. 68-70, 83, 84).

[Zhuravchak, 2022] Andrii Zhuravchak, Oleg Kapshii et Evangelos Pournaras. « Human activity recognition based on wi-fi csi data-a deep neural network approach ». *Procedia Computer Science* 198 (2022), p. 59-66 (cf. p. 25).

[Zou, 2017] Yongpan Zou, Weifeng Liu, Kaishun Wu et Lionel M. Ni. « Wi-Fi Radar : Recognizing Human Behavior with Commodity Wi-Fi ». *IEEE Communications Magazine* 55.10 (2017), p. 105-111 (cf. p. 10, 13, 17, 65, 72).

Hadi EL ZEIN

Doctorat : Optimisation et Sûreté des Systèmes

Année 2025

Apprentissage profond pour la classification multidimensionnelle en reconnaissance d'activités humaines

La reconnaissance d'activités humaines (HAR) via la *Channel State Information* (CSI) s'impose comme une solution non-intrusive dans le secteur de la santé. Cette thèse aborde deux problématiques majeures liées aux données et aux modèles, à travers trois objectifs : (1) développer une méthodologie de HAR efficace pour la reconnaissance d'activités en temps réel, (2) améliorer la précision de la détection de chutes dans un contexte d'interventions médicales urgentes, et (3) permettre l'identification d'utilisateurs à partir de schémas comportementaux subtils. Pour y parvenir, des solutions d'apprentissage profond intègrent l'augmentation de données, des architectures convolutionnelles ou basées sur des transformers, ainsi qu'un mécanisme de decision avancé. Trois contributions principales en découlent. D'abord, HAR-LightCNN traite les jeux de données limités et déséquilibrés grâce à un réseau de neurons convolutionnel allégé. Ensuite, HFD-Adv introduit une méthodologie en quatre étapes (fonction de coût optimisée par algorithmes génétiques, transfert d'apprentissage, post-traitement démographique) pour minimiser les faux négatifs. Enfin, IdentNet exploite l'auto-attention d'un transformer afin de distinguer des utilisateurs ayant des profils d'activité similaires. En rendant ces systèmes plus fiables et adaptatifs, ce travail contribue à des applications concrètes dans le domaine de la santé, depuis la simple reconnaissance d'activités jusqu'à l'intervention personnalisée en cas de chute.

Mots clés : CSI (télécommunications) – reconnaissance de l'activité humaine (informatique) – apprentissage profond – chutes (accidents) – suivi médical.

Deep Learning for the Classification of Multidimensional Signals Applied to Human Activity Recognition

Human Activity Recognition (HAR) using Wi-Fi Channel State Information (CSI) enables privacy-preserving, device-free monitoring in healthcare settings. This thesis tackles data- and model-related concerns through three objectives: (1) developing an efficient real-time HAR methodology, (2) enhancing fall detection accuracy for urgent healthcare interventions, and (3) enabling user identification based on subtle behavioral patterns. To achieve this, novel deep learning solutions combine data augmentation, convolutional or transformer-based architectures, and advanced decision mechanisms. Three core contributions follow. First, HAR-LightCNN addresses small, imbalanced datasets with a lightweight CNN. Second, HFD-Adv integrates genetic algorithm-optimized loss functions, transfer learning, and demographic-sensitive post-processing to enhance fall detection. Finally, IdentNet leverages transformer self-attention to distinguish users with similar activity patterns. By increasing reliability and adaptability, these methodologies support real-world healthcare applications, from basic activity recognition to personalized emergency responses for fall incidents.

Keywords: CSI (telecommunications) – human activity recognition – deep learning (machine learning) – falls (accidents) – medical monitoring.

Thèse réalisée en partenariat entre :

الجامعة اللبنانية
UNIVERSITE LIBANAISE

utt
UNIVERSITÉ DE TECHNOLOGIE
TROYES

Ecole Doctorale "Sciences pour l'Ingénieur"