Li, Zhenghui (2024) *Radar sensing for ambient assisted living application with artificial intelligence.* PhD thesis.

http://theses.gla.ac.uk/84245/

Radar sensing for Ambient Assisted Living application with Artificial Intelligence

Zhenghui Li

Manuscript submitted in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

UNIVERSITY OF GLASGOW

COLLEGE OF SCIENCE AND ENGINEERING

JAMES WATT SCHOOL OF ENGINEERING

Abstract

In a time characterized by rapid technological advancements and a noticeable trend towards an older average population, the need for automated systems to monitor movements and actions has become increasingly important. This thesis delves into the application of radar, specifically Frequency Modulated Continuous Wave (FMCW) radar, as an emerging and effective sensor in the field of "Activity Recognition." This area involves capturing motion data through sensors and integrating it with machine learning algorithms to autonomously classify human activities. Radar is distinguished by its ability to accurately track complex bodily movements while ensuring privacy compliance.

The research provides an in-depth examination of FMCW radar, detailing its operational principles and exploring radar information domains such as range-time and micro-Doppler signatures. Following this, the thesis presents a state-of-the-art review in activity recognition, discussing key papers and significant works that have shaped the field. The thesis then focuses on research topics where contributions were made. The first topic is human activity recognition (HAR) with different physiology, presenting a comprehensive experimental setup with radar sensors to capture various human activities. The analysis of classification results reveals the effectiveness of different radar representations. Advancing into the domain of resource-constrained system platforms. It introduces adaptive thresholding for efficient data processing and discusses the optimization of these methods using artificial intelligence, particularly focusing on the evolution algorithm such as Self-Adaptive Differential Evolution Algorithm (SADEA). The final chapter discusses the use of Long Short-Term Memory (LSTM) networks for short-range personnel recognition using radar signals. It details the training and testing methodologies and provides an analysis of LSTM networks performance in temporal classification tasks.

Overall, this thesis demonstrates the effectiveness of merging radar technology with machine learning in HAR, particularly in assisted living. It contributes to the field by introducing methods optimized for resource-limited settings and innovative approaches in temporal classification using LSTM networks.

Acknowledgements

I extend my deepest gratitude to my academic supervisor, Dr. Julien Le Kernec, whose expertise, understanding, and patience, added considerably to my graduate experience. Your guidance and mentorship were invaluable in shaping both the direction and success of this research. My sincere appreciation also goes to my co-supervisor team, Dr. Francesco Fioranelli, Dr. Shufan Yang and Dr. Lei Zhang, whose camaraderie, support, constructive suggestions, and exchange of ideas greatly enhanced my research experience.

Heartfelt thanks go to my parents, whose unwavering love and support have been my constant source of strength. Your sacrifices and belief in my abilities have been the driving force behind my every achievement.

I am also immensely grateful to my friends for their endless encouragement, understanding, and companionship. I would particularly like to highlight Dachao, YiXiang, LingYi, Lianfeng, Xinyu, Yang, Wenjing and Peng. The moments we have shared and the support you have provided have been a pillar of strength throughout my academic journey.

My colleagues at Glasgow are also of great support to me, with special mentions for Haobo Li, and Yushi Liu. The collaborative environment, intellectual exchanges, and camaraderie within our research group have significantly enriched my experience. Your insights and support have been invaluable in the completion of this thesis.

This thesis would not have been possible without the contributions and support of these individuals and organizations. I am deeply thankful to each one of them.

Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Zhenghui Li

Table of Contents

Table of Figures

Table of Tables

Acronyms

AAL – Ambient Assisted Living

ADC – Analogue to Digital Converter

AI – Artificial Intelligence

BN – Bayesian Network

CNN – Convolutional Neural Network

CFAR – Constant False Alarm algorithm

CSI – Communication, Sensing and Imaging

CV – Cross-Validation

CVD – Cadence Velocity Diagram

CW – Continuous Wave

1/2/3/4-D – One/ two/ Three /Four dimensions

DAC – Digital to Analogue Converter

DCNN – Deep Convolutional Neural Network

DE – Differential evolution

DF – Decision Fusion

DNN – Deep Neural Network

DRDT – Dynamic Range-Doppler Trajectory

DT – Decision Tree

DL – Deep Learning

DWT – Discrete Wavelet Transform

EA – Evolutionary Algorithm

FFT – Fast Fourier Transform

FNN – Feed-Forward Neural Network

F1– F1-Score

FMCW – Frequency Modulated Continuous Wave

GP – Gaussian Process

GR – Gesture Recognition

GRU – Gated Recurrent Units

HAR– Human Activity Recognition

HD – Human Detection

KNN – K-Nearest Neighbour

LCB – Lower Confidence Bound

LiDAR – Light Detection and Ranging

LNA – Low Noise Amplifier

LHS – Latin Hypercube Sampling

LPF – Low Pass Filter

LSTM – Long Short-Term Memory

mD – Micro Doppler

MDR – Maximum Detectable Range

MDS – Maximum Doppler Shift

MIMO – Multi Input, Multi Output

MOCAP – Motion Capture

MTI – Moving Target Indication

ML – Machine Learning

NB – Naïve Bayes

NCD – Noncommunicable Disease

NLP – Natural Language Processing

OFDM – Orthogonal Frequency-Division Multiplexing

PA – Power Amplifier

PCA – Principal Component Analysis

PPCA – Probabilistic Principal Component Analysis

PR – Personnel Recognition

RC – Recall Combiner

RCS – Radar Cross Section

ReLu – Rectifier Linear Unit

RF – Radio Frequency

RGB – Red Green Blue

RGB-D – Red Green Blue-Depth

RMS – Root Mean Square

RNN – Recurrent Neural Network

SFS – Sequential Forward Selection

SFFS – Sequential Floating Forward Selection

STFT – Short-Time Fourier Transform

SNR – Signal Noise Ratio

SVD – Singular Value Decomposition

SVM – Support Vector Machine

TF – Time-Frequency

TV – Television

UWB – Ultra Wide-Band

WHO – World Health Organization

WT – Wavelet Transform

WVD – Wigner-Ville Distribution

Symbols

B – Bandwidth

b – Bias

C – Penalty parameter

c – Speed of light

c_t – Cell state

E – Loss function

v_r – Radial target velocity

v_s – Radial source velocity

F – Overlapping factor

f_r – Frequency of received signal

f_c – Frequency of transmitted signal

f_D – Doppler shift

f_s – Sampling rate

f_{beat} – Beat Frequency

f_t – Forget gate

H – Hyperplane

h_t – Hidden weight

i_t – Input gate

$I(t)$ – I channel output of FMCW radar

K – Clipping piece of time

o_t – Output gate

$Q(t)$ – Q channel output of FMCW radar

φ_0 – Induced phase shift

m – Length of window

NP – Number of samples

p – Index of Doppler bin

q – Index of range bin

$h(n)$ – Window function

ΔR – Range resolution

R_e – Estimated range

R – Overlap

$s(t)$ – Transmitted Signal

$r(t)$ –Received Signal

T – Chirp period

T_e – Additional thresholding

τ – Time delay

ω – Angular frequency

δ – Aspect Angle

ϵ – Cadence frequency

V – Adaptive thresholding factor

W – Time window length

w_i – Weights

1. Introduction

1.1 Context

The World Health Organization (WHO) projects that the proportion of individuals aged 60 and above will increase from 12% to 22% between 2015 and 2050, which highlights the ongoing demographic shift toward an increasingly ageing population [1]. It is estimated that there will be one person aged over 60 in every six people by 2030 [1]. As this shift in the distribution of the population continues to older ages, the demand for people who have longer life spans do not always enjoy a matching quality of life [2]. Meanwhile, the expanding age group is confronted with several challenges that were previously unforeseen in earlier generations but are now experienced on a much larger scale [3], [4], [5], [6]. The older adults tend to experience chronic diseases more frequently than younger individuals, which often require long-term care and management. Additionally, older adults are at a higher risk of falls and other critical activities, which could lead to severe injuries and disabilities [6], [7]. The noncommunicable disease (NCD) and critical events pose significant public health challenges, which has negative impacts on individual life expectancy. According to the latest report from WHO [8], the global share of deaths attributable to NCDs increased from almost 61% in 2000 to almost 74% in 2019, leading it as the top global cause of death. The critical events, such as fall, often occurs simultaneously with the disease like stroke, which exacerbates the condition of patients.

The UK, like many other developed countries, is experiencing an ageing population trend as shown in Figure 1. According to the Office for National Statistics [9], the proportion of the population aged 65 and over in the UK is increased from 16.4% in 2011 to 18.6% in 2021. This trend is largely due to a combination of factors, including increased life expectancy and a decrease in birth rates. Compared to the rest of the world, the UK's ageing population is relatively high. According to a report by the United Nations, the proportion of the UK population aged 65 and over is projected to be around 25% by 2050, which is much higher than the global average of 16%. However, this trend is not unique to the UK, as many other developed countries such as Japan, Italy, and Germany also have ageing populations. It is a global challenge that requires innovative solutions to address the needs of this growing population.
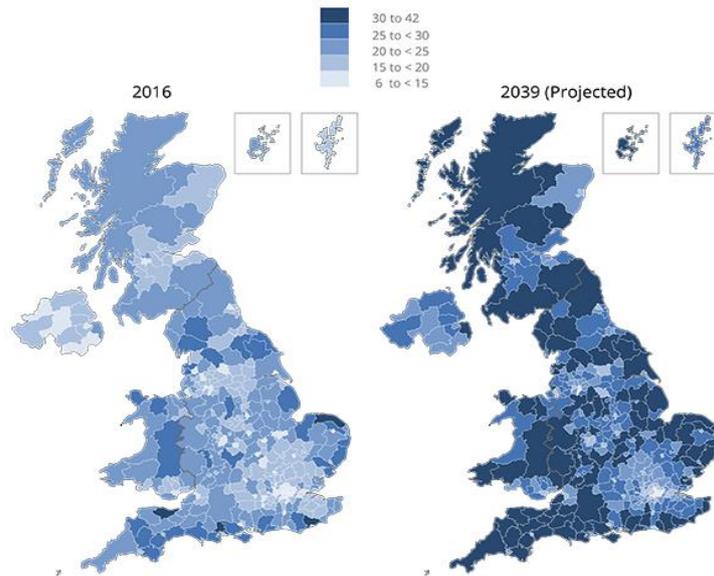
Figure 1 *Proportion of the population aged 65 years and over, 2016 and 2039, UK. 2016 mid-year population estimates for UK, Office for National Statistics, 2014-based subnational population projections for UK, Office for National Statistics, Welsh Government, National Records Scotland and Northern Ireland Statistics and Research Authority, contains OS data © Crown copyright 2018* [2].

To support ageing individuals, ambient assisted living (AAL) is increasingly used to provide low-latency intelligence sensing-based products, services, and systems that enable elderly people to live independently in their homes [10], [11]. By utilizing various technological innovations [12], [13], [14], such as wearable devices, and radar, the elderly can receive support without the need for direct human assistance. The AAL system is designed to provide a higher quality of life and increased independence for seniors while reducing the burden on caregivers and healthcare systems. In addition, these technologies can also provide real-time monitoring and emergency response services. The monitoring of the daily activities of the elderly can help identify possible changes and anomalies that may indicate deteriorating health conditions [15]. For example, the decrease in daily physical activities could be a sign of a potential health issue, such as mobility impairment. Tracking these changes can also contribute to informing diagnosis and prognosis for medical practitioners. Besides, several studies [7], [16] illustrate that the time taken to receive medical aid after a fall event is highly correlated with life expectancy. By providing timely assistance, we can increase the chances of a full recovery and reduce the potential negative impact of falls on older adults.

In recent years, the field of AAL has seen the emergence of several different sensing technologies. These include wearable sensors [17], [18], [19], video-based systems [20], ambient sensors, and radio frequency (RF) sensors [21], [22], [23]. Of these, wearable sensors are designed to measure data related to the movements of an

individual, such as acceleration and angular speed. Wearable sensors are capable of measuring vital signals such as pulse rate and skin temperature, as well as properties and small movements of the attached body parts. Due to their small size and relatively low cost, wearable sensors have become increasingly popular. However, there are some limitations to using these sensors, particularly during motion or sports activities. Wearing the sensors for extended periods can cause discomfort, and they require to be worn constantly to be effective [24]. In order to wearable sensors to be effective, consistent wear is required, which demands a certain level of compliance and cognitive ability from the user to remember to wear the sensors. This is a challenge for ageing people especially those who are suffering from memory loss [24].

The vision-based approach, on the other hand, using cameras to monitor an area and capture images or video for recognition purposes, is one of the most commonly employed methods in recent research works, as it is highly connected with computer vision, providing a range of sophisticated tools derived from the field and a practical framework [25]. This approach provides a contextual representation of activities through pre-processed images or videos, which are then learned to shape a model using various learning methods [25]. However, this approach can lead to privacy invasion and disputes over image rights. People may feel it is quite intrusive as their living conditions are constantly monitored by a camera. Moreover, camera images and recordings provide visual information about the user, which can give hints about their physical size or height, and the environment they live in. Additionally, it is easily affected by lighting conditions [26], and a camera cannot guarantee image quality in both weak and strong light. In adverse lighting conditions or darkness, the vision-based method cannot function correctly without an artificial light source. These variations strongly affect the accuracy of deployed algorithms.

RF sensing has gained popularity due to its flexibility and potential advantages over cameras and wearable sensors. Its contactless capabilities eliminate the need for end-users to wear, carry, or interact with any additional device, which can increase user acceptance and compliance [15], [27]. Furthermore, the absence of plain images or videos to be recorded can help address potential privacy concerns. Radar, in particular, has emerged as one of the most representative RF sensing approaches. It provides rich information in multi-domain, including range, velocity, and angle, and the information can also be presented in various dimensions, which makes it ideal for motion and target detection [28]. Within the various domains of radar information, micro-Doppler has emerged as a commonly utilised approach for capturing the small modulations in the received radar signal that result from the 'micro-motion' of individual body parts,

such as the limbs and legs [29]. Consequently, there has been a significant number of research investigating the utility of radar sensing for tasks such as human activity recognition and personnel recognition.

Research in radar-based human activity recognition for assisted living has risen considerably due to a rapidly increasing of ageing population. Machine learning-based methods [27], [30] are widely employed to classify collected data and features, as a classifier can learn to identify the patterns of different activities through a training process. In recent years, neural network-based [15], [31] methods have demonstrated superior classification accuracy compared to conventional classifiers since they can automatically extract salient features from radar data, albeit at the cost of increased training complexity. Deep learning aims to find the mapping relationship between the training data and the labels through the learning and testing of many samples.

However, classifying human activities based on data from a single domain may not be sufficient to fully characterize them. Additionally, the signal strength of a single radar can be attenuated due to external conditions, such as the aspect angle [32], obstruction, and long distances to the monitored target. To overcome these limitations and improve classification performance, it is possible to use multi-domain radar data or additional radar nodes. This approach allows for the selection of salient features across different domains and the implementation of fusion techniques between multiple domains, other radar nodes, or different sensing modalities that complement radar data.

1.2   Motivation and Object

1.2.1 Motivation

Each radar domain has its advantages and disadvantages. For instance, one of the most commonly used domains is the spectrogram, which shows micro-Doppler signatures based on Short-Time Fourier Transform (STFT). The distinctive motions of different body parts constitute a specific micro-Doppler signature, enabling human activity recognition [33], [34]. However, spectrograms have limitations in time/frequency resolution when activities with low interclass variations are introduced, or high-frequency accuracy is required for data representation. In addition, spectrograms are ill-suited for the separation of multiple close objects, even if the distance between targets is greater than the range resolution as the range information is lost. Moreover, many works still tend to apply the same algorithm, using the same features, to recognise all activities in a multi-class problem. There are few attempts to capitalise

on the diversity of information that can be recorded by various feature combinations and different radar domains.

The results obtained from multiple radar domains can be further improved by implementing selection and information fusion algorithms. By combining the strengths of different radar domains and selecting targeted features for different activities, the classification system can be made more robust in distinguishing similar activities and identifying certain events [28], [35]. However, using multi-domain radar data and additional algorithms, such as feature fusion, can increase the volume of input data or features for machine learning and deep learning. This can lead to complex data processing methods and training models, which in turn can increase computational cost and power consumption. This can be problematic for deployment on embedded platforms, and thus reducing the footprint of the algorithm in terms of energy consumption and silicon is necessary to lower the product price for end-users.

Another challenge faced with radar is that entanglement can occur in the Micro-Doppler signature when considering the multi-target scenarios [28], [35]. These entanglements can hinder the separation of the signature of different targets. If the individuals were moving in the same direction, it would not be possible to disentangle their signatures without prior processing to separate in range and angle the different targets. The multi-domain solution can also provide insight with using radar technology for multi-target activity recognition, which offers a different explanation of data in different domains.

1.2.2 Objectives

The primary objective of this thesis is to develop an adaptive, intelligent, and efficient solution for human activity recognition using radar. Section 1.2.2 provides an overview of the key research topics and motivations that underpin this study. To achieve our goals, this research focused primarily on the following areas:

I. Exploring the potential of different data representation domains of radar and enhance the accuracy of activity recognition through the integration of domain knowledge.
    o It was achieved through three steps.
       1- Exploration of each data representation domain and its corresponding features and comparison of their performances on different activities.

2- Development of a feature fusion and selection algorithm, which provided a comprehensive analysis of the features, domains, and performances to improve the accuracy of activity recognition.

3- Introduction of a hierarchical structure to enhance the performance of human activity recognition. It will leverage a combination of techniques to identify the optimal approach for maximizing the accuracy of activity recognition.

II. Designing a human activity recognition system that achieves satisfactory performance while minimizing computational cost.
   o It was achieved through three steps.

   1- Develop an adaptive data processing system, which can provide a general processing solution for researched domains. This system is also aimed to leverage previous domain knowledge, alongside feature selection and feature fusion, to create a robust, efficient, and general solution for our researched data domains.

   2- Design an ML-based human activity recognition system that is combined with prior data processing techniques, with the goal of achieving high accuracy while reducing the computational cost.

   3- Evaluate the performance of the developed system through experiments to gain valuable insights into its practical applications and potential for implementation in real-world settings.

In addition to the main objectives outlined above, there are several other objectives, or branches of our work that must also be addressed throughout this research. These ancillary objectives will be identified and completed as necessary to ensure that the study is comprehensive, accurate, and robust.

A. Conducting experiments and collect human activity data from real subjects to evaluate the performance of our developed algorithms. The aim is to gain valuable insights into the practical applications of our methods and their potential to be implemented in real-world settings. The data collected will be used to test and validate the accuracy and efficiency of the proposed activity recognition system.

B. Evaluating the performance of various machine learning and deep learning algorithms for activity recognition using radar data. The algorithms will be trained and tested on the same dataset to compare their performance. The goal is to identify the most effective algorithm that achieves high accuracy in

activity recognition using radar data. The results of this study will provide valuable insights into the potential of different machine learning and deep learning algorithms for improving the accuracy of activity recognition using radar data.

C. Expanding the feature research to include personnel recognition, alongside the primary focus on human activity recognition. Through this expansion, we aim to determine the effectiveness of the developed algorithms in recognizing and classifying various individuals within radar data. To achieve this objective, we will collect and analyse data from real subjects and employ various machine learning and deep learning algorithms for training and testing purposes. The results of this research will provide valuable insights into the potential of our developed algorithms in personnel recognition applications and multi-target research in future.

1.3   Contributions

1.3.1 Statement of Novelty

The research conducted in this thesis makes several novel contributions to the field of radar-based human activity recognition and personnel recognition.

Contribution linked to Objective I: To investigate the potential of different radar data representation domains for improving the performance of activity recognition, we introduce a novel pre-processing method for radar-based human activity recognition (HAR) utilizing adaptive thresholding. This method automatically generates regions of interest (ROI) from human micro-Doppler signatures, accompanied by a set of specifically designed features tailored for classification across different domains. This contribution links to our objective I.

Contribution linked to Objective II: The thesis investigates various optimization strategies to enhance the performance of the HAR system. These strategies include the fusion of data domains and selection techniques, as well as exploring different parameters of the statistical learning model. Additionally, a hierarchical structure is employed to further optimize the system's performance. These optimization approaches demonstrate significant improvements in the overall performance of the HAR system. This contribution links to our objective II.

Other Contributions: In addition, this study is to conduct experiments and collect human activity data from real subjects, evaluate the performance of various machine

learning and deep learning algorithms for activity recognition using radar data, and expand the feature research to include personnel recognition. We aim to gain valuable insights into the practical applications of our developed algorithms, compare the performance of different algorithms in activity recognition, and explore the effectiveness of our algorithms in recognizing and classifying individuals within radar data.

These contributions provide new insights and techniques for the development of robust and accurate radar-based assisted living systems. Overall, this thesis presents a comprehensive approach to human activity recognition and personnel recognition using radar data, with novel contributions that advance the state of the art in the field.

1.3.2 Elaboration of Contributions

This thesis aims to explore the potential of using different radar domain information for human activity recognition and monitoring. The scope of this research is discussed in the state-of-the-art section, which includes hardware, software, and experimental techniques to leverage radar technology and its applications in healthcare. In this rapidly evolving domain, there are central questions that this thesis attempts to answer. These questions are derived from the literature review and the contributions made by this thesis for improving activity recognition and personnel recognition through the use of radar sensors. The main contributions of this thesis are as follows:

1. **The demonstration of the effects of different domains of radar data on classification accuracy.**
   In this research, we aimed to investigate the potential of different data representation domains of radar for improving the accuracy of activity recognition, as outlined in objective I. Initially, we aligned our work with the state-of-the-art by utilizing existing algorithms and literature focusing on micro-Doppler signatures and machine learning techniques for activity recognition. However, as our research progressed, we discovered the valuable information that other radar domains, such as the range-time map, could offer for human activity classification. Our investigation revealed that the classification rates varied among radar domains, indicating that certain domains exhibited higher accuracy in classifying specific activities.
   Consequently, our early contribution was the investigation of the effect of different radar domains on activity recognition, aligning with the potential of these domains for enhancing activity recognition through the fusion of domain

knowledge. This investigation provided valuable insights into the potential of radar data from various domains, which can be leveraged to improve the accuracy of our activity recognition system.

2. **Proving that utilising features from different domains with feature selection, fusion and different classification structures can increase the performance for human activity recognition.**

   After reviewing the literature and our previous works, we advanced our research by implementing feature selection and fusion frameworks to enable activity recognition. We compared various combinations of radar domains and features using different classification algorithms, validating their performances. To further enhance the accuracy of activity classification, we proposed a hierarchical classification structure. By integrating these approaches, we extracted diverse information from radar data, resulting in improved accuracy for activity recognition. Overall, our investigation focused on the impact of different radar domains on activity classification results and aimed to develop activity recognition methods through feature selection and fusion. These efforts were driven by the objective I of our study, which explores the potential of various radar data representation domains in enhancing activity recognition.

3. **Demonstrating improved recognition of human activity without feature extraction and extending our previous results to personnel recognition.**

   With the development of artificial intelligence methods, deep learning approaches have been increasingly used to exploit radar data for classification problems. Radar data is well-suited for deep learning as it can be represented as either temporal or visual information. We focused on evaluating the performance of various machine learning and deep learning algorithms for activity recognition using radar data. Our aim was to identify the most effective algorithm that achieves high accuracy in activity recognition using radar data. To achieve this, we explored different types of neural networks using spectrograms and compared their performance. Moreover, the objective of this study was to expand our feature research to include personnel recognition, alongside the primary focus on human activity recognition, we proposed a novel approach by combining information from both range-time maps and spectrograms using neural networks, with a specific focus on gait information. This contribution extended the use of artificial intelligence techniques with radar from human activity recognition to personnel recognition, building on our

previous work. Our results provide valuable insights into the potential of our developed algorithms in personnel recognition applications and multi-target research in the future.

4. **Developing a new algorithm for human activity recognition to guarantee accuracy and efficiency.**

In our previous contributions, we proposed solutions for radar-based human activity classification in indoor scenarios. However, these approaches often prioritize accuracy over efficiency, using complex data processing methods or deep neural networks that result in long latencies at the inference/testing stage. Additionally, they typically use the same algorithm to recognize all activities, disregarding the potential of various feature combinations and different radar domains. To address these issues, we present an adaptive thresholding method that leverages the 'phase' domain of radar, which has been seldom investigated in previous studies. This approach achieves high classification accuracy while reducing computational burden compared to most deep learning methods, making it ideal for realistic deployment on embedded platforms. This contribution is linked to the second major objective of this study, which aims to design a human activity recognition system that achieves satisfactory performance while minimizing computational cost. Our approach successfully yields a high gain in classification accuracy while reducing computational burden compared to most deep learning methods, making it suitable for realistic deployment on embedded platforms.

1.3.3 Publications

Patent

1 - Centre National de la Recherche Scientifique (2022), Method and device for human activity classification using radar micro-Doppler and phase, EP21306742.

Journal papers

1 - X. Li, Z. Li, F. Fioranelli, S. Yang, O. Romain, and J. L. Kernec, "Hierarchical Radar Data Analysis for Activity and Personnel Recognition," Remote Sensing, vol. 12, no. 14, p. 2237, Jul. 2020 [29].

2 - Li, Z., Le Kernec, J. , Abbasi, Q. , Fioranelli, F. , Yang, S. and Romain, O. (2023) Radar-based human activity recognition with adaptive thresholding towards resource constrained platforms. Scientific Reports, 13, 3473 [36].

3 - Z. Li, Y. Liu, B. Liu, J. Le Kernec, and S. Yang, "A holistic human activity recognition optimisation using AI techniques," IET Radar, Sonar Navig., Sep. 2023. [37]

Conference papers

1- Z. Li et al., "Multi-domains based human activity classification in radar," IET International Radar Conference (IET IRC 2020), Online Conference, 2020, pp. 1744-1749 [38].

2- Z. Li, J. Le Kernec, F. Fioranelli, O. Romain, L. Zhang and S. Yang, "An LSTM Approach to Short-range personnel recognition using Radar Signals," 2021 IEEE Radar Conference (RadarConf21), Atlanta, GA, USA, 2021, pp. 1-6 [39].

3- Z. Li, F. Fioranelli, S. Yang, J. Le Kernec, Q. Abbasi and O. Romain, "Human Activity Classification with Adaptive Thresholding using Radar Micro-Doppler," 2021 CIE International Conference on Radar (Radar), Haikou, Hainan, China, 2021, pp. 1511-1515 [40].

1.4   Thesis Structure

This thesis is organised as follows:

Chapter 2 serves a necessary background of this thesis. It begins with an overview of the working principle of FMCW radar, including its fundamental parameters and how they affect the radar performances. The signal processing steps of FMCW radar are also described in detail, along with the illustration and discussion of different data domains that contain valuable information for human activity recognition. Then it delves into the machine learning aspect of this research by discussing the algorithm and architecture used in this thesis.

Chapter 3 provides a comprehensive literature review relevant to human healthcare until recently. It examines radar based AAL technology in the aspect of artificial intelligence. We also cover AAL with alternative sensors, such as wearable sensors or camera. Furthermore, the review is also involved the selection and fusion approaches between different types of features. This chapter highlights the gaps in the literature which provided opportunities for the research presented in this thesis to be conducted.

Chapter 4 presents the preliminary results of using handcrafted features for activity recognition with different radar domains. The chapter demonstrates the effectiveness of leveraging different types of information from various domains and compares the

performance of using each domain alone with their fusion results. Moreover, the impact of feature selection as well as different classifiers and fusion methods are thoroughly discussed.

Chapter 5 presents a computational cost-saving model, elaborating the results from Chapter 4, where an adaptive thresholding solution is proposed, along with an introduction to radar representations, specifically phase form, which has been seldom considered in the literature. The performance of the proposed solution is evaluated against the various existing state-of-the-art learning models. Then, a comprehensive optimization strategy is introduced, focusing on the pre-processing aspects of the proposed methods, and the Chapter concludes with a brief summary of results and analysis.

Chapter 6 extends on the work by introducing deep learning models and personnel recognition based on gait analysis. This chapter specifies how automatic feature generation approach using neural networks to classify activities without handcrafted feature extraction algorithm. Besides, the chapter explores the combination of gait analysis with temporal characteristics of radar, using time-dependent neural networks to achieve personnel recognition.

Chapter 7 lastly, we summarize the thesis and draw a vision that suggests possible ways to improve related research.

2. Radar and Machine Learning Background

The detection of human motion with radar involves two fundamental technical components: radar systems and machine learning systems. As the radar plays a crucial role in the overall detection scheme, this chapter first focuses on providing an in-depth overview of radar operation and signal processing required to generate information from human activities. Following the detailed explanation of radar and signal processing, this chapter delves into the machine-learning aspect of the research. The fundamental versions of machine learning algorithms and architectures used for activity recognition are thoroughly explored. This includes discussing various classification algorithms and deep learning models forming the core of the cognitive machine learning systems used in human activity recognition.

By combining the knowledge of radar signal processing with machine learning techniques, this chapter sets the foundation for the subsequent sections of the thesis, where we integrate these components to achieve accurate and efficient human motion detection and activity recognition. An understanding of both radar technology and machine learning principles is vital for developing an effective human activity recognition system with radar-based sensors.

The organisation of the chapter is as follows: In sections 2.1 and 2.2, we will introduce radar system structures, basic radar signal processing, and the micro-Doppler effect, particularly in the case of human motion. In section 2.3, we focus on different radar representations. In sections 2.4 and 2.5, both machine learning and deep learning algorithms are introduced respectively. We will provide an overview and basic concepts of machine learning. Some of the classification algorithms will also be pointed out since they are employed in the next chapters.

2.1 Radar

The word 'radar' is an acronym, RADAR, standing for radio detection and ranging. The transmitter is responsible for emitting the electromagnetic waves. These waves are then directed towards a target through the air using an antenna. In a monostatic radar, the same antenna also receives the waves reflected back from the target. A separate antenna may be used for receiving waves in multistatic or quasi-monostatic configurations to avoid blind distances [41]. Subsequently, a receiver captures the received information, and a processor is utilized to process and display this information [42].

The radar can be categorized in terms of whether it transmits radio frequency (RF) signals actively (active radar) and listens to the backscattered signals or it relies on illuminations of opportunity (passive radar). Active radar systems are the most commonly used form of radar. These systems actively transmit RF signals toward a target and then receive the reflected signals to determine the target's properties, such as its range, velocity, and size. The term "active" refers to the radar's capability to generate its own electromagnetic waves for transmission [42]. Unlike active radar systems, passive radar does not have a transmitter of its own. Instead, it relies on existing electromagnetic signals in the environment, such as broadcast TV and radio signals, to illuminate the target [42].

For active radar, there are many categorization rules. For example, according to the radar configuration, the radar can be categorized as monostatic radar, bistatic radar (Figure 2), multistatic radar and Multiple-Input Multiple-Output radar.
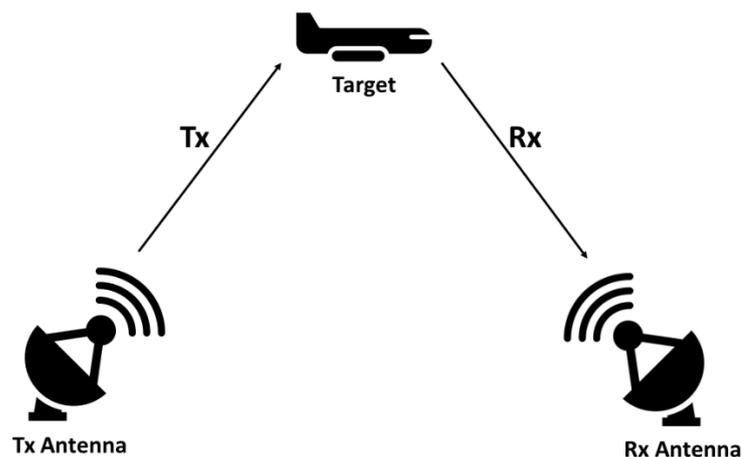
Figure 2 *The basic operation of a bistatic radar detector. The transmitted wave is denoted as Tx and the received wave is Rx.*

- Monostatic radar uses the same antenna for both transmit and receive or co-located two antennas as transmitter and receiver. For single antenna used, duplexer is needed to separate the transmit chain from receive chain and vice versa.
- Bistatic radar uses two spatially separate antennas for the transmitter and the receiver. They are not co-located and are separated by a distance.
- Multistatic radar consists of multiple monostatic radar or bistatic radar that are spatially separated, with a shared area of coverage. Multistatic system can afford spatial diversity to improve target detection and tracking capabilities.

- MIMO radar utilizes multiple antennas for both transmitting and receiving, enhancing radar capabilities beyond traditional monostatic or bistatic systems. This setup, with each transmit antenna emitting a distinct signal and multiple receive antennas capturing reflections, allows for improved spatial resolution and target detection. The multiple antenna paths enable better target localization and robustness against interference, making it effective in complex scenarios.

Table 1 The comparison of different radars Configuration.

Feature	Monostatic Radar	Bistatic Radar	Multistatic Radar	MIMO Radar
Antenna Configuration	Single or Co-located antenna for Tx and Rx	Separate antennas for Tx and Rx	Multiple antennas for TX and/or RX	Multiple antennas for TX and RX
Hardware Complexity	Lower	Higher	Due to the number of units.	Highest
Cost	Generally lower	Generally higher	Higher	Highest
Signal Processing	Simplified	More complex	More complex	Most complex
Synchronization	Easier (co-located Tx and Rx)	More complex	More complex	Most complex
Latency	Generally lower	May vary	May vary	May vary

Radar configuration is a critical part of radar-based systems, particularly in applications like HAR, where the quality and type of data collected can significantly impact the performance of system [43]. The configuration determines various parameters such as the frequency band, pulse repetition frequency and antenna arrangement. It is also related to the aspect angles, which also influences collected signals [43].

Table 1 illustrate a brief comparison between different radar configurations. However, there are more than one categorization methods. In terms of transmitted signal type, the radar can be categorized as CW radar and Pulse radar.

- The CW radar emits a constant RF signal and measures the frequency shift (Doppler shift) in the returned signal to determine the velocity of a target. It has many advantages such as high resolution in velocity measurements and lower power consumption [44], however there is no range information. The constant

RF signal could be modulated over time, usually in a linear or sinusoidal manner, and it will become a frequency modulated CW radar, also known as FMCW radar [45] with which you can measure range and velocity.

- On the other hand [46], pulse radar emits short bursts of radio waves and listens for the echo to measure both distance and velocity according to time delay and Doppler shift analysis. It is commonly used in air traffic control and marine radar since it is effective for long-range detection.

A main comparison between the radars is summarized in Table 2.

Table 2 The comparison of different radars in terms of transmitted signals.

Feature	CW Radar	Pulse Radar	FMCW radar	MIMO radar
Signal Type	Unmodulated Continuous	Pulsed with Modulations	Frequency-modulated continuous	FMCW TDM/DDM* or orthogonal waveforms
Range Measurement	No	Yes	Yes	Yes
Shorter Range Target Detection	Better	Worse	Better	Better
Velocity Measurement	Yes	Yes	Yes	Yes
Complexity	Lower	Higher	Higher	Higher
Power Consumption	Lower	Higher	Higher	Higher

*TDM: Time Division Multiplexing, DDM: Doppler Division Multiplexing

In this thesis, a FMCW radar is implemented to collect human activity data, since it is particularly suitable for HAR research under indoors scenario: FMCW Radar uses frequency modulation to encode the time information into the frequency domain. This allows for rapid signal processing, as the range and velocity information can be extracted simultaneously from the received signal using Fourier Transform techniques [41] resulting in a light computational load. Meanwhile, it can provide high temporal resolution due to its continuous wave nature, and fine range resolution [41]. The high-resolution data from FMCW Radar can be effectively used as input features for machine learning algorithms, enhancing the performance of HAR systems. FMCW radar systems are also low-consumption. The combination of low consumption and low signal processing complexity makes it particularly suitable for in-home

environments. Note however, that multiple FMCW radar operating at the same frequency may interfere with each other. This will not be the case in this research.

Statistical and deep learning models can be trained to recognize complex human activities by learning the intricate patterns in the radar signals and its features [28]. Furthermore, pulse radar, often has a minimum range limitation due to the need for a "listen" period between pulses. It also needs time to switch between the transmit and receive modes. This switching time creates a "blind zone" around the radar where targets may be too close to be detected. This can make it less suitable for short-range applications [41] hence the need for quasi-monostatic radar with one transmit and one receive so there are no blind zones.

2.2 FMCW Radar Basic Principles

In this section, the basic principles behind the operation of radar, particularly in FMCW radar, will be introduced. The signal pre-processing of raw data, involving modulation, range and Doppler extraction. This will provide the required technological background for the exploration of different radar representations.

2.2.1 Doppler effect

The Doppler Effect refers to the change in frequency or wavelength of a wave in relation to an observer moving relative to the source of the wave. In radar systems, the Doppler Effect manifests as a frequency shift in the reflected radar signal when the target is moving toward or away from the radar [47]. The Doppler Effect provides crucial information about the velocity of a target, which is essential for distinguishing between different types of human activities based on their velocity profiles, such as walking, running, or falling [47].

When the target and observer are both moving, there is a frequency shift between the transmitted and received signal due to the Doppler Effect. The frequency of the received echo signal is given in Eq. (2.1) [47].

$$f_r = \left(\frac{c \pm v_r}{c \mp v_s} \right) f_c \qquad (2.1)$$

where f_c is the frequency of the transmitted wave, and c denotes the speed of light. v_r and v_s are the target and source velocities. In radar research, the majority of the

works considered the radar to be stationary for human activity recognition, and therefore the equation can be simplified as shown in Eq. (2.2).

$$f_r = \left(\frac{c \pm v_r}{c}\right) f_c \qquad (2.2)$$

For different configurations of radar, the Doppler effect calculation has different factors to consider, and thus we illustrate both monostatic radar and bistatic radar. Multistatic being a superposition of bistatic radar configuration, the principles are the same as bistatic.

For monostatic configuration radar [47], since the transmitter and receiver are at the same location, and the electromagnetic wave travels the distance twice between the radar and the target for the transmit and receive process, therefore the relationship becomes:

$$f_D = 2(f_r - f_c) = -\frac{2v_r}{c} f_c \qquad (2.3)$$

In bistatic radar system [47], the Doppler shift would become much more complex because the angles and distances between radars are considered. As shown in Figure 3, the transmitter and receiver are separated by a baseline of distance L. The phase between the transmitted signal and the received signal changes along with the target moving, and the Doppler frequency shift is measured by the phase change rate. To simplify the process, the bisector is introduced because it provides a reference line that is equidistant from both the transmitter and the receiver. This simplification is particularly useful when the transmitter and receiver are separated by a significant distance [47], and the target is also at a considerable distance from both. The equation, in a nutshell, can be represented as Eq. (2.4):

$$f_D = \frac{2f_c}{c} |V| \cos\left(\frac{\varphi}{2}\right) \cos\delta \qquad (2.4)$$

Where f_c is the frequency of transmitted signal, c denotes the speed of light, V is the velocity vector of moving target, φ is the bistatic angle between transmitted wave and received wave, and δ is the aspect angle between the direction target and the bisector of target.
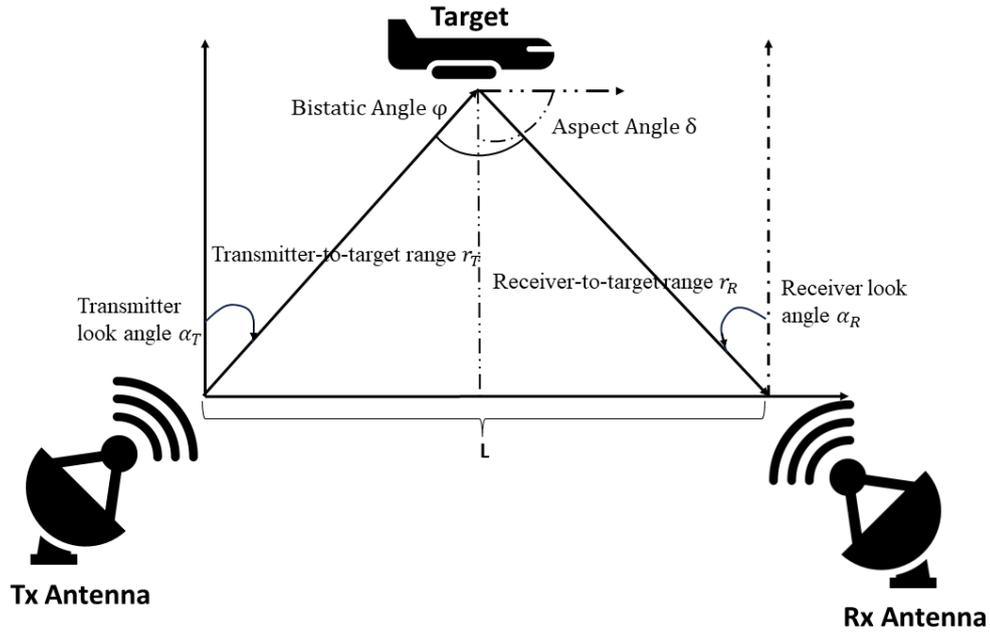
Figure 3 *The two-dimensional bistatic radar system configuration.*

2.2.2 FMCW radar

With advancements in computing and processors, new modulation methods, which can encode both range and velocity simultaneously, are now employed. Frequency Modulated Continuous Wave (FMCW) radar is a type of radar system that continuously transmits a frequency-modulated signal. The transmitted signal is reflected off a target, and the radar receiver detects the returned signal. The frequency difference between the transmitted and received signals is used to determine the range (distance) to the target, while the frequency shift due to the target movement provides information about its velocity [42], [47].

Transmitted signals could be modulated in different patterns [48]. Figure 4 and Figure 5 illustrate the linear sawtooth pattern of modulation. As is shown in Figure 4, the frequency of linearly modulated signal, also known as chirp signal, is varied along with time in a linear fashion. Besides the sawtooth pattern, other types are also used such as triangular and square wave [48].

In the signal processing stage [46], [47], [48], the difference in frequency between transmit and receive waveform is known as beat frequency. This frequency difference allows to the measurement of the distance to the reflecting object. Since the frequency difference is proportional to the time delay τ between the radar and the target, the range information can easily be extracted by the beat frequency and extracting the frequency change, as shown in Eq. (2.5) and Eq. (2.6)

$$\frac{\tau}{T} = \frac{f_{beat}}{B} \tag{2.5}$$

$$range = \frac{\tau c}{2} \tag{2.6}$$

where f_{beat} is the beat frequency, T is chirp length and B is the bandwidth of chirp signal.
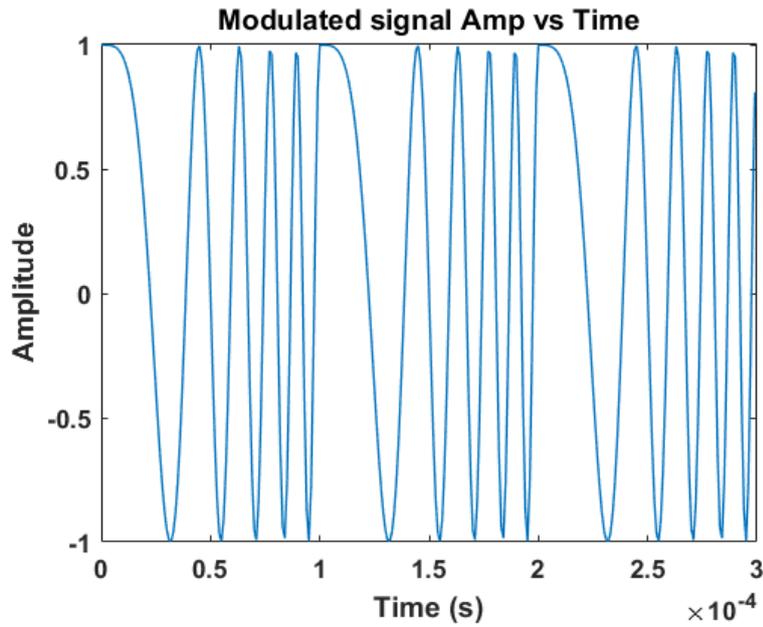
Figure 4 *FMCW radar chirp signal in the time domain.*
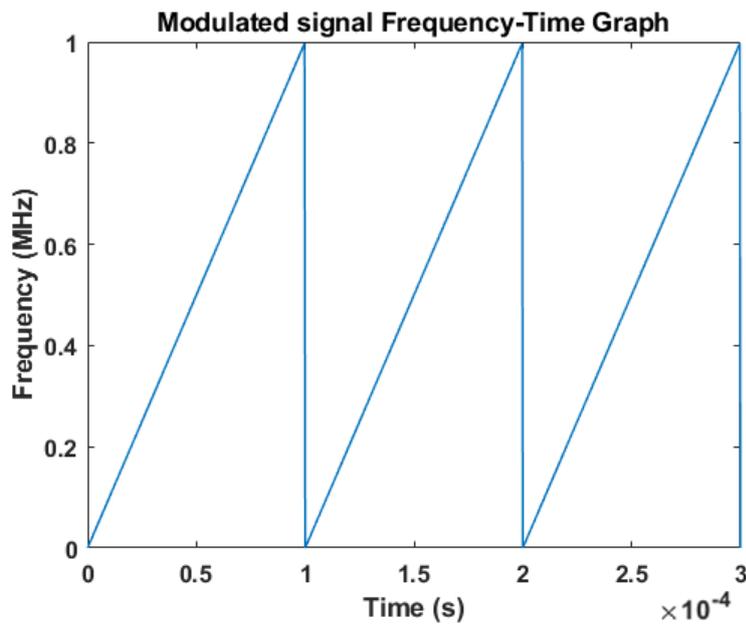
Figure 5 *FMCW radar chirp signal time-frequency representation.*

However, the range resolution ΔR, referring to the size of a range bin, is linked with the radar bandwidth B, as depicted in Eq. (2.7):

$$\Delta R = \frac{c}{2B} \qquad (2.7)$$

Hence, the bandwidth is significant to FMCW radar. Without enough bandwidth, FMCW radars cannot distinguish close targets as a separation greater than $2\Delta R$ is needed to separate targets and not result in an extended target.

2.2.3 FMCW radar system structure

Figure 6 illustrates a synoptic of an FMCW radar [49]. The transmitter part consists of a digital-analogue converter (DAC) for converting a digital controlled order into analogue voltage. A voltage-controlled oscillator (VCO) is also used for generating a corresponding RF signal according to the input voltage. The signal is then upconverted to the required operating band with a superheterodyne architecture. The signal is then split between the receiver path for 'dechirping' and the other path feeds into a power amplifier (PA). The signal after amplification is transmitted into free space with an antenna. The backscattered signals are captured by the receiving antenna and amplified through a low noise amplifier (LNA). The received signal is fed to the RF input of two mixers and the replica of the chirp from the transmitter path goes through a second splitter. One path goes in the local oscillator input of the in-phase (I) channel mixer and the other is 90º phase shifted before the local oscillator input of the mixer in the quadrature (Q) channel. The I/Q components at the intermediate frequency output of the mixers are the result of dechirping i.e., removing the frequency modulation by mixing the replica from the transmitter with the received signals yielding low-frequency components (beat frequencies) and high frequency components. The latter are low-pass filtered to only keep the beat frequencies. The I/Q IF components are then digitized with Analogue to Digital Converters (ADC).

I/Q components are commonly used in radar system for several important advantages: I/Q components allow the received signal to be represented as a complex signal, which is essential for capturing both amplitude and phase information [50] as well as further relaxing the requirements on the ADCs and simplifying signal processing by removing the step to reconstruct the imaginary part of the received signal. This presents however some drawbacks in cost with added hardware with a second IF channel for the Q channel meaning also that imbalances in phase and amplitude between I and Q channels will appear affecting the signal quality.

Complex signals are beneficial to accurately determine the velocity direction of targets [51]. Besides, I/Q components can provide valuable information, which can be used for feature extraction [52]. Moreover, they can improve the signal to noise ratio (SNR), especially in low SNR environments, thereby enhancing detection capabilities [53].
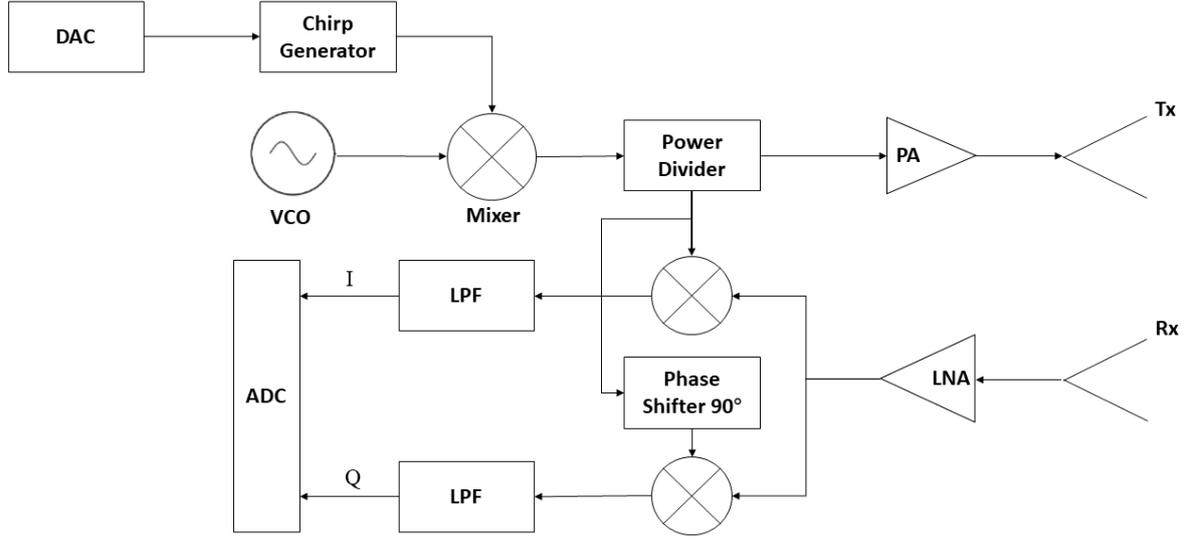
Figure 6 *Block Diagram of a typical FMCW radar system. VCO: Voltage-Controlled Oscillator, DAC: Digital to Analogue Converter, ADC, Analogue to Digital Converter, LNA: Low Noise Amplifier, PA: Power Amplifier, LPF: low-pass filter, I: In-phase, Q: Quadrature, Tx: Transmitter, Rx: Receiver.*

2.2.4 Frequency Modulated Continuous Wave

The FMCW radar can provide both range and radial velocity of target. In this thesis, the linear frequency modulation is discussed only since it has been applied to the experiments conducted for this work. The chirp signal, as is shown in Figure 5, is composed of a linear frequency sweep with a fixed duration [42], [48], [54]. If the chirp ramp sequence is swept from a carrier frequency f_c, then frequency at any given time instant $f(t)$ can be expressed as Eq. (2.8) [41], [48], [54]:

$$f(t) = f_c + \frac{B}{T}t = f_c + \alpha t, \tag{2.8}$$

Where T is the sweep duration, B is the bandwidth of chirp, and α denotes the chirp rate, which is also illustrated in Figure 5. The corresponding instantaneous phase can be obtained from the Eq. (2.9):

$$\omega(t) = \frac{d\varphi(t)}{dt} = 2\pi f(t) \tag{2.9}$$

Thus, we can acquire the equation of $\varphi(t)$ shown in Eq. (2.10).

$$\varphi(t) = \int_0^t 2\pi f(t)dt = 2\pi\left(f_c t + \frac{\alpha}{2}t^2\right) + \varphi_0 \tag{2.10}$$

Where φ_0 denotes the initial phase.

The transmitted signal now can be expressed as Eq. (2.11):

$$s(t) = A\cos\left(2\pi\left(f_c t + \frac{\alpha}{2}t^2\right) + \varphi_0\right) \tag{2.11}$$

Where A denotes the amplitude of transmitted signal. It is a constant and is related to signal power.

For a target at a range R, the time delay τ of received signal can be characterised as Eq. (2.12) [41], [48], [54]:

$$\tau = \frac{2R}{c} \tag{2.12}$$

Considering Eq. (2.11), the received signal $r(t)$ can be expressed by:

$$r(t) = B\cos\left(2\pi\left(f_c(t-\tau) + \frac{\alpha}{2}(t-\tau)^2\right) + \varphi_0\right) \tag{2.13}$$

Where B denotes the amplitude of received signal. It is also a constant and is related to signal power.

Assuming normalized amplitude for both transmitted and received chirp signals, and without taking the initial phase into consideration, we get two general expressions for the transmitted signal $s(t)$ and the received signal $r(t)$ [41], [48], [54]:

$$s(t) = \cos\left(2\pi\left(f_c t + \frac{\alpha}{2}t^2\right)\right) \tag{2.14}$$

$$r(t) = \cos\left(2\pi\left(f_c(t-\tau) + \frac{\alpha}{2}(t-\tau)^2\right)\right) \tag{2.15}$$

The transmitted and received signals are mixed at the demodulator generating the I/Q components, and hence the in-phase mixer output, also known as I component, is the product of two functions, as shown in Eq. (2.8):

$$I(t) = s(t)r(t) = \cos\left(2\pi\left(f_c t + \frac{\alpha}{2}t^2\right)\right)\cos\left(2\pi\left(f_c(t-\tau) + \frac{\alpha}{2}(t-\tau)^2\right)\right) \quad (2.16)$$

Using trigonometric identity equation $\cos(x)\cos(y) = \frac{1}{2}(\cos(x+y) + \cos(x-y))$, the equation can be rewritten as Eq. (2.17):

$$I(t) = \frac{1}{2}(\cos(2\pi t(2f_c - \alpha\tau) + 2\pi\alpha t^2 - 2\pi f_c\tau) + \cos(\pi\alpha\tau^2 - 2\pi\alpha t\tau - 2\pi f_c\tau))(2.17)$$

Note that an LPF is used after the mixing stage, which eliminates the components with high-frequency, and the final equation of the output I signal is written as Eq. (2.18).

$$I(t) = \frac{1}{2}\cos(\pi\alpha\tau^2 - 2\pi\alpha t\tau - 2\pi f_c\tau) \quad (2.18)$$

The quadratic component Q, is given by mixing the received signal with a 90°-shifted replica of the transmitted signal and is shown in Eq. (2.19):

$$Q(t) = \frac{1}{2}(\sin(2\pi t(2f_c - \alpha\tau) + 2\pi\alpha t^2 - 2\pi f_c\tau) + \sin(\pi\alpha\tau^2 - 2\pi\alpha t\tau - 2\pi f_c\tau))(2.19)$$

Also, after filtering with the LPF, the quadratic component can be rewritten as Eq. (2.20):

$$Q(t) = -\frac{1}{2}\sin(\pi\alpha\tau^2 - 2\pi\alpha t\tau - 2\pi f_c\tau) \quad (2.20)$$

One of the most important advantages of I/Q components is that it can represent radar signals in complex form. Thus, the general signal $s(t)$ of FMCW radar can be denoted as Eq. (2.21), and phase $\varphi(t)$ can be denoted as Eq. (2.22):

$$s(t) = I(t) + jQ(t) = e^{-\pi\alpha\tau^2 + 2\pi\alpha t\tau + 2\pi f_c\tau} \quad (2.21)$$

$$\varphi(t) = -\pi\alpha\tau^2 + 2\pi\alpha t\tau + 2\pi f_c\tau \quad (2.22)$$

The FMCW radar utilizes the beat frequency to obtain the time delay τ, and the beat frequency could be calculated through the derivative of the phase of the digitised signal. The distance between the radar and the target can be expressed as a function of τ, and hence, we can acquire the function of angular frequency ω_{beat}, beat frequency f_{beat} and estimated range R_e, as is shown in Eq. (2.23) to Eq. (2.25).

$$\omega_{beat} = \frac{d(-\pi\alpha\tau^2 + 2\pi\alpha t\tau + 2\pi f_c\tau)}{dt} = 2\pi\alpha\tau \quad (2.23)$$

$$f_{beat} = \frac{\omega_{beat}}{2\pi} = \alpha\tau \tag{2.24}$$

$$R_e = \frac{cf_{beat}}{2\alpha} \tag{2.25}$$

In the previous equations, we focus on an ideal and simple circumstance: single chirp with a still target. In real life, the FMCW radar has successive chirps, and the target, at most of times, is moving towards or away from the radar with radial velocity v. Thus, the time delay τ_r could be rewritten as

$$\tau_r = \frac{2\big(R + v(t_s + nT)\big)}{c} = \frac{2R + 2v(t_s + nT)}{c} \tag{2.26}$$

Where n is the number of chirps that we analysed. t_s is the time from the start to the nth chirp. By substituting the τ in Eq. (2.23) with τ_r. Similarly, the beat frequency f'_{beat} can be obtained as Eq. (2.27):

$$f'_{beat} = \frac{(2R\alpha + 2vf_c + 2nBv)}{c} = \frac{2R\alpha}{c} + \frac{2vf_c}{c} + \frac{2nBv}{c} \tag{2.27}$$

From Eq. (2.26), it is observed that a Doppler shift and a frequency component generated during sweeps could now add on the frequency of entire beat frequency. Generally, the frequency component generated during sweeps could be negligible since $c \gg 2nBv$. However, in case of large numbers of chirps, which means an extreme long observation time, this term is possibly not negligible.

2.3 Radar Information Domains
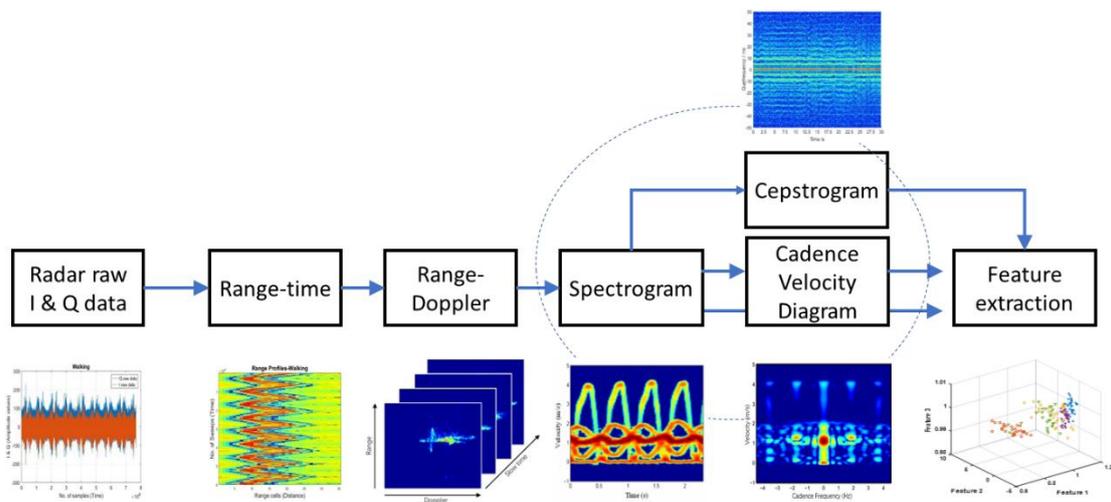
Figure 7 *Information domain of radar data used for human activity classification* [28].

With the advancement of radar technology, radar has been used more and more into civilian applications. One of the most promising avenues is Human Activity Recognition (HAR) [35], [55]. In the previous sections, we introduced beat frequency, which is one of the basic principles of FMCW radar. By employing advanced signal processing techniques, subtle human movements and micro-motions can be extracted, thereby opening up possibilities for applications like fall detection and gait analysis [56]. Figure 7 demonstrates the current mainstreams of radar domains, which are implemented in HAR. Namely, they are raw data, range-time, range Doppler, Cepstrogram, Doppler-time a.k.a. Spectrogram, and Cadence Velocity Diagrams.

2.3.1 Range-time domain

The generation of a range profile is a critical component in FMCW radar systems, especially for applications in HAR. The range profile provides a snapshot of the environment by indicating the presence of targets at various distances from the radar. Similar to the previous formulations, range information of beat-note signal could be extracted through transferring from time to frequency domain, which means employ Fourier Transform.

By combining Eq. (2.22) and Eq. (2.27). The mathematical formulations of radar signal are shown in the Eq. (2.28) and Eq. (2.29). Note that the component $\frac{2nBv}{c}$ in Eq. (2.27) is negligible as we have discussed before, so it is not considered.

$$\varphi_R(t) = \frac{4\pi f_c v}{c} nT + 2\pi \left(\frac{2vf_c}{c} + \alpha\tau \right) \tag{2.28}$$

$$s(t) = e^{-\varphi_R(t)} \tag{2.29}$$

Note that the Fourier Transform is traditionally defined through continuous integrals over an infinite time span. However, the computation of continuous integrals is impractical, and the requirement for infinite duration is unattainable in real-world applications. To address these limitations, the use of the Discrete Fourier Transform is essential [53].
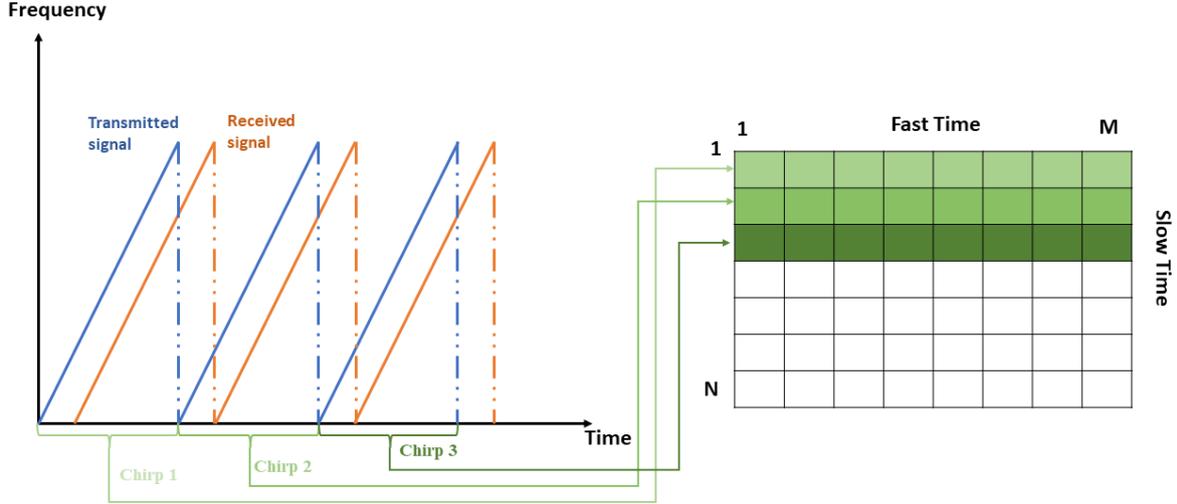
Figure 8 *FMCW radar information storage.*

Figure 8 demonstrates a $N \times M$ matrix. N is the number of successive chirps, also known as slow time, and M represents the number of bins for each chirp, also known as fast time. Because the beat frequency signal is discretized by sampling, the $M = f_s T$, where f_s is the sampling rate.

Applying fast Fourier Transform (FFT) to the signal Eq. (2.29) as is illustrated in Figure 8 meaning in the fast time direction for every received chirp. It gives:

$$S(p, q) = \sum_{q=1}^{M} s(p, q) e^{-j\frac{2\pi}{M}qt} \tag{2.30}$$

Where $S(q)$ is frequency domain signal of the target and q is the new index of range bins.

Figure 9 illustrates range-time maps obtained by the FFT process for six different human activities, where the colour level of these heatmaps indicates the received signal strength. (a) represents walking with the swinging of arms as its periodic range pattern, (b) and (c) are symmetric so they denote sitting down and standing up, respectively. (d) and (e) are drinking water and picking up an object, whereas (f) is the simulation of a fall. Additionally, the smearing across the range is caused by the radar device itself, by standing waves and an imbalance between the I and Q channels in amplitude and phase, which could be ignored due to the low signal amplitude.
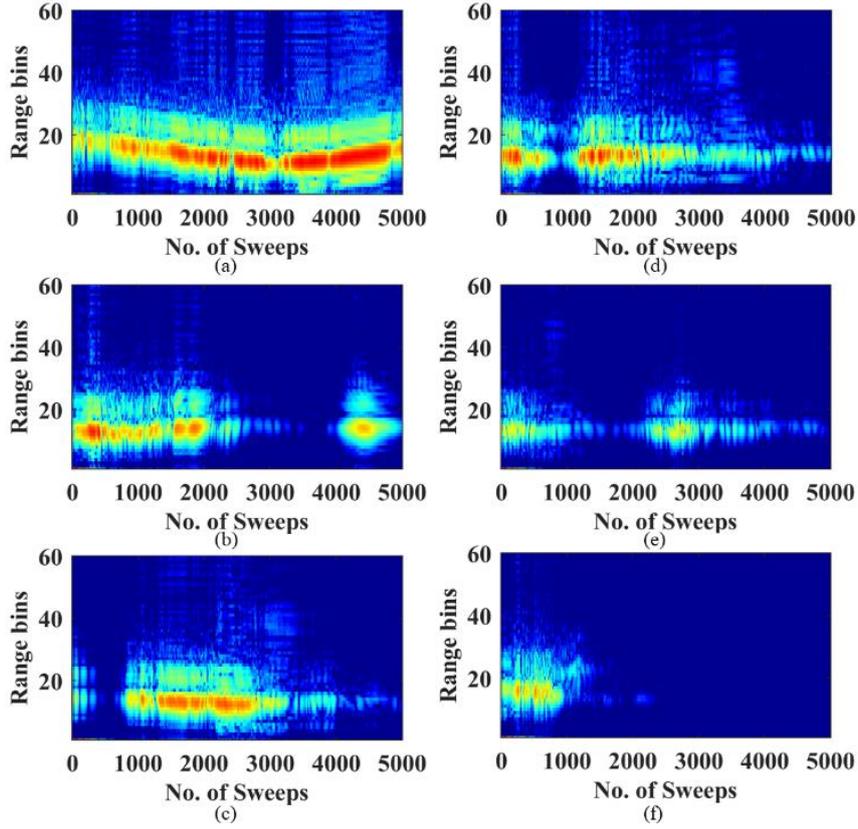
Figure 9 *The time range information of a young adult performing different activities. (a) walking, (b) sitting down, (c) standing up, (d) picking up an object, (e) drinking, (f) fall.*

2.3.2 Range-Doppler domain.

Range-Doppler mapping is a pivotal concept in the application of FMCW radar systems, particularly in the field of HAR [57]. This technique combines the range information obtained from the frequency modulation of the signal with the Doppler shift information to create a two-dimensional representation of the target's range and velocity. The Range-Doppler map is particularly useful in HAR as it provides a comprehensive view of the spatial and velocity characteristics of human activities. For instance, it can differentiate between a person who is stationary and one who is moving, or between different types of movements like walking and running [58].

In Range-time map, the range information is obtained along with sweep time. The FFT is performed on fast time to extract range information in different chirps. To further get velocity information, now it is necessary to perform the second FFT on slow time to generate range-Doppler matrix, that is, the 2D-FFT, are used to generate the Range-Doppler map from the received radar signals [59]. It can be expressed as Eq. (2.31):

$$S(p,q) = \sum_{p=1}^{N} \left[\sum_{q=1}^{M} s(p,q) e^{-j\frac{2\pi}{M}qt} \right] e^{-j\frac{2\pi}{N}pt} \qquad (2.31)$$

Eq. (2.31) adds another FFT based on Eq. (2.30). $S(p,q)$ is the frequency domain signal of target, and p, q are the new index of Doppler bins and range bins.

Like Figure 9, the range Doppler information of different activities from a young participant is also illustrated in Figure 10.
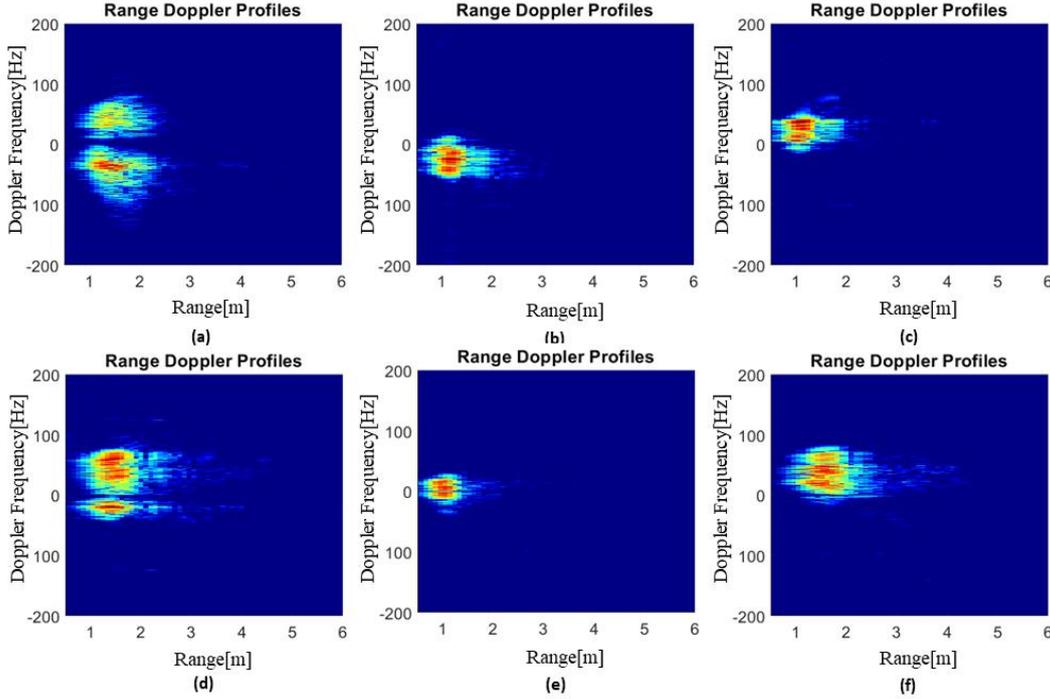
Figure 10 *The range Doppler information of a young adult performing different activities. (a) walking, (b) sitting down, (c) standing up, (d) picking up an object, (e) drinking, (f) fall.*

2.3.3 Doppler-time domain – Micro-Doppler Signature - Spectrogram

Human motion can be broadly categorized into two types: macro-motion and micro-motion. Macro-motion refers to activities that result in a change in the overall position of the human body, such as running or walking. In contrast, micro-motion involves activities where the overall position of body remains relatively constant but includes smaller, localized movements [47]. These could be limb or head movements, as well as physiological activities like breathing and heartbeat when the body is stationary. Due to the low Signal-to-Clutter and Noise Ratio (SCNR) associated with micro-motion, especially in the case of hand movements, feature extraction from micro-motion has become a critical aspect of target classification [47].

Micro-Doppler (mD) plays a crucial role in radar systems, especially in detecting small oscillatory movements. The ability of a radar system to detect mD signatures depends on its carrier frequency. Higher frequency systems can detect smaller mD signatures, making the centre frequency a critical property for mD applications. These mD shifts are time-varying frequency shifts that can be extracted from the complex output of a radar system.

The micro-movements of the human body produce radar echoes with distinct characteristics. Traditional Fourier Transform methods are effective for analysing stationary signals but fall short when applied to non-stationary signals commonly encountered in HAR [60], since they are not well-suited for capturing these time-varying frequency characteristics [61]. These signals often exhibit variations in Doppler frequency content over time, making it challenging to capture their intricacies using only time-domain or frequency-domain analysis. Instead, Joint Time-Frequency Analysis (JTFA) is commonly employed to analyse micro-Doppler features and extract micro-motion characteristics, since JTFA encompasses both linear and quadratic time-frequency distributions [62], which offers a two-dimensional representation that encapsulates both temporal and spectral characteristics of a signal [63]. By analysing a signal in both the time and frequency domains simultaneously, TFA offers invaluable insights into the complex dynamics of human activities [63].

The utilization of JTFA In radar signal processing serves as a powerful mechanism for generating micro-Doppler signatures. These signatures capture the small-scale Doppler shifts caused by micro-motions [47]. These could be small oscillatory movements in a mechanical system, including biological systems like humans. These tiny movements can generate detectable Doppler shifts in high-frequency systems, and the shifts appear as sidebands in radar signatures. These sidebands valuable insights that can be leveraged to identify and characterize specific movements, such as the rotation of mechanical rotors and the swing of human limbs [47].

Various JTFA techniques exist, each with its own set of advantages and limitations. Some of the most commonly used methods in radar-based HAR with radar are included as follows:

Short Time Fourier Transform (STFT) [64] is an extension of the Fourier Transform designed to analyse non-stationary signals. It works by dividing the signal into overlapping or non-overlapping segments and applying the Fourier Transform to each segment. This provides a time-frequency representation that captures the local frequency content of the signal at different time intervals. It is easy to implement and

provides a straightforward interpretation of time and Doppler frequency localization. For signal $s(n)$, the STFT is given as Eq. (2.32).

$$STFT(n, \omega) = \sum_{m=-\infty}^{\infty} s(n)h(n - mR)e^{-j\omega n} \qquad (2.32)$$

for $n = 0, \dots, N - 1$, where N is the total number of time samples. $s(n)$ is the input signal, $h(n)$ is called window function. R is the hop or sliding size in samples, which is also the overlap between two nearest windows, and m is the length of window.

STFT has the highest calculation efficiency. It is the most commonly used time-domain spectral calculation method. However, this method also has some limitations, that is, because the time resolution and Doppler frequency resolution will be limited by the width of window function, it cannot achieve the best effect at the same time.

STFT performs the Fourier transform on a short-time window basis rather than taking the Fourier transform of the entire signal using one long-time window. With the time-limited window function, the resolution of the STFT is determined by the window size. There is a trade-off between the time resolution and the Doppler frequency resolution [65]. A larger window has a higher-frequency resolution but a poorer time resolution. Overlapping windows can help with edge discontinuities, to generate a smooth signature at the cost of increased computational load [65].

In addition to STFT, Wavelet Transform (WT) [66] employs wavelets, which are localized wave functions, to analyse signals at multiple scales. Unlike STFT, WT allows for variable time-frequency resolution, making it more flexible for analysing signals with non-uniform frequency content. WT has multi-resolution capability. A 'mother' wavelet is used and its scaled and translated versions, detect frequencies at multiple resolutions. There is a variety of choices including, but not limited to: Haar, Mexican Hat, and Morlet etc.

Moreover, Wigner-Ville Distribution (WVD) [67] is a quadratic TFA method that offers high-resolution time-frequency transform to characterize the spectral and temporal behaviour of the signal. However, it suffers from the problem of cross-term interference (i.e., the WVD of the sum of two signals is not the sum of their individual WVDs). If a signal contains more than one component in the joint time-frequency domain, its WVD will contain cross terms that may complicate the interpretation of the signal.

Spectrogram is a widely used method to display time-varying spectral density of a time-varying signal. It is a spectro-temporal representation and provides the actual change of Doppler frequency contents of a signal over time. The spectrogram is calculated by using the short-time Fourier transform (STFT) and represented by the squared magnitude of the STFT without keeping phase information of the signal:

$$Spectrogram(n, \omega) = |STFT(n, \omega)|^2 \qquad (2.33)$$
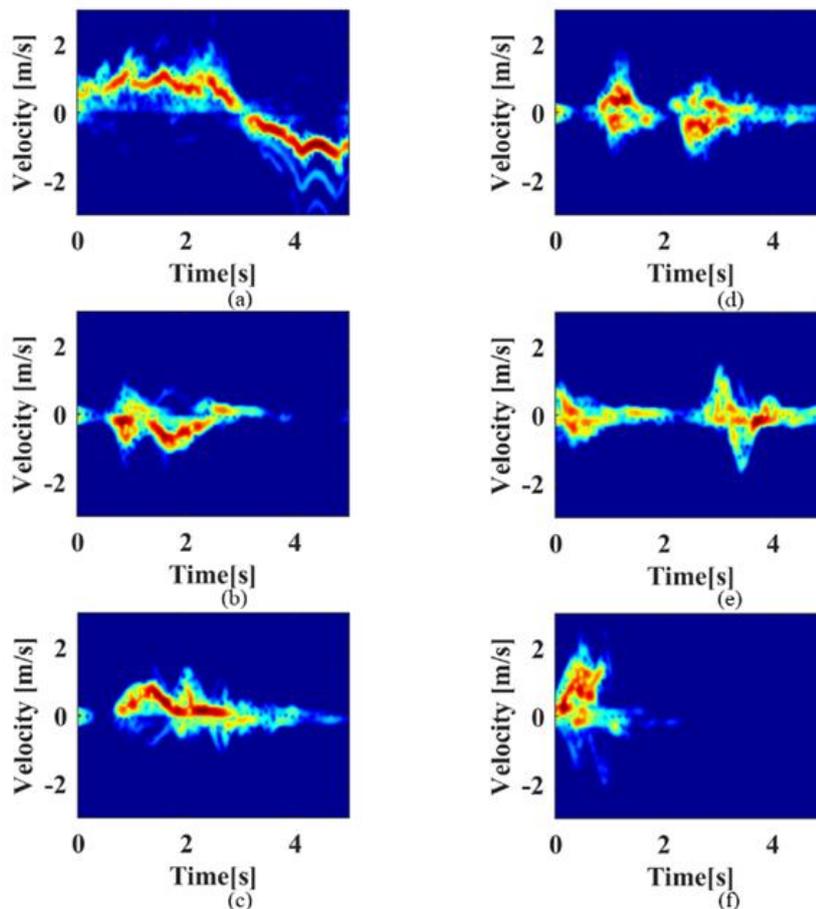
Figure 11 *The mD of different activities of a young adult. (a) walking, (b) sitting down, (c) standing up, (d) picking up an object, (e) drinking, (f) fall.*

Figure 11 illustrates the spectrograms for the same activities depicted in Figure 9 and Figure 10. The micro-Doppler effect reveals distinct patterns associated with the movements of different body parts. For example, in (a), a typical micro-Doppler signature, the central mass represents the torso's movement, while spikes or protrusions on this central mass can be attributed to limb swings and head rotations. Features extracted from these micro-Doppler signatures are particularly effective in distinguishing between similar activities. This is because these features are closely correlated with the micro-motion dynamics exhibited by the subject under observation.

In summary, the combination of TF analysis and the understanding of mD provides a powerful tool in radar signal processing. Specifically, the use of STFT enables the efficient computation of TF analysis while ensuring the effective detection and characterization of micro motions. This could have wide applications, from detecting small parts moving in mechanical systems to identifying and tracking human movements in security systems.

2.3.4 Cadence Velocity Diagram

Cadence-Velocity Diagrams (CVDs) have emerged as a critical tool in radar-based HAR, offering a two-dimensional representation that encapsulates both the cadence and velocity of a moving target [65]. Cadence refers to the frequency of repetitive motion, such as steps per minute. Different activities have distinct cadence patterns. In time-domain analyses, activities like walking and running may produce similar waveforms, making it challenging to distinguish between them based solely on amplitude and time. However, when these activities are represented in a CVD, their unique cadence-velocity patterns become evident [68]. These diagrams are particularly effective due to their ability to capture the periodic properties of cyclic human activities, such as walking, in a sparser and power-efficient representation within the TF domain [69].

As STFT employs FFT within a sliding window framework, the periodicity of the backscattered signal is preserved in its micro-Doppler signature. This results in each Doppler bin exhibiting the same periodicity. Consequently, when compared to the signal representation in the time domain, the cyclic characteristics of human activities like walking persist within the time-frequency domain. This representation is sparser yet possesses a higher power content. By performing an FFT along the time axis of the mD signature, as shown in Eq. (2.33), the cadence frequencies can be extracted for each Doppler bin.

$$CVD(\epsilon, \omega) = \left| \sum_{n=0}^{N-1} spetrogram(n, \omega)e^{-j\frac{2\pi n\epsilon}{N}} \right| \qquad (2.34)$$

Where ϵ is the cadence frequency. In HAR, CVDs are powerful for classifying activities that may appear similar in time-domain analyses but exhibit distinct cadence-velocity characteristics. The Figure 12 demonstrates the CVDs for different daily activities.
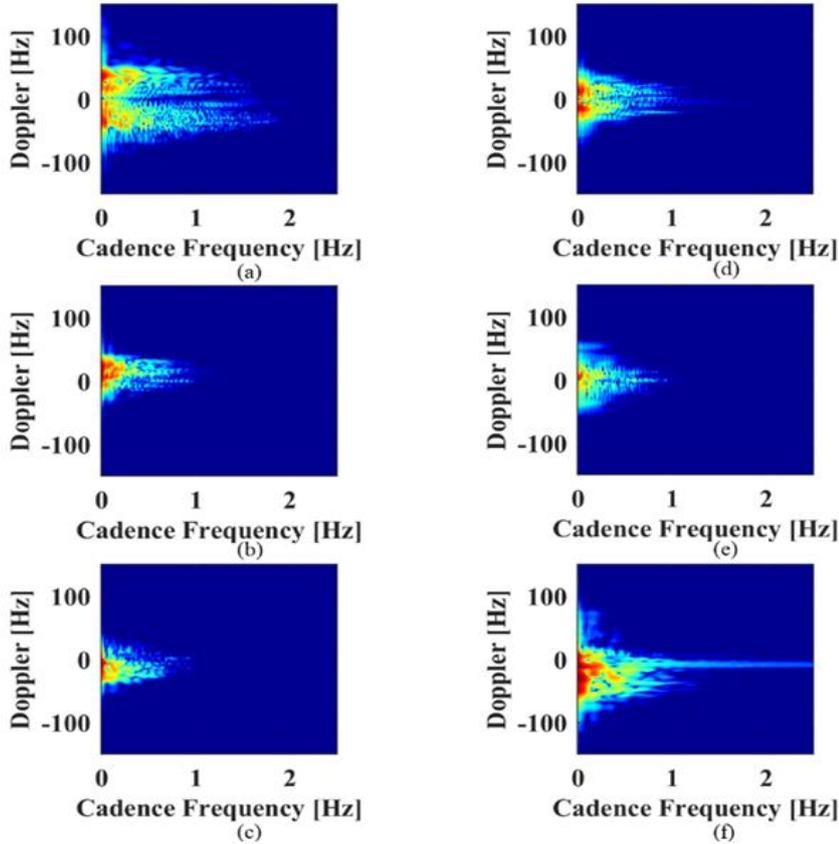
Figure 12 *The CVDs of different activities of a young adult. (a) walking, (b) sitting down, (c) standing up, (d) picking up an object, (e) drinking, (f) fall.*

2.3.5 Other domains

Apart from the previous domains that we introduced, there are still various radar representations, which provide rich information. It is crucial to broaden our scope to explore other domains in radar signal processing that offer complementary advantages. Each domain provides a unique lens through which we can examine and interpret radar data, thereby enriching the feature set available for machine learning algorithms or neural networks.

Similar to CVD, the cepstrum or cepstrogram serves as a transformative tool for radar signals, particularly for achieving linear separation of convolved signals. The radar cepstrum [15], [69] is derived by applying the Inverse Discrete Fourier Transform (IDFT) to the logarithm of the absolute energy within the spectrogram. This process effectively maps the data into a Quefrequency-time domain, where Quefrequency is the inverse of frequency. The information content of the radar cepstrum is closely tied to the energy distribution within the spectrogram, offering another dimension for signal analysis.

Furthermore, cyclostationarity [70] is also introduced to radar signal processing. Cyclostationarity is a statistical property of signals that exhibit periodic variations in their statistical characteristics over time. Unlike stationary signals, where statistical properties like mean and variance are constant, cyclostationary signals have statistical features that change in a periodic manner [71]. In the context of radar applications, when a periodic pulse train is transmitted and then reflected off a moving individual, and such a reflection induces a random modulation in the received signal, thereby exhibiting cyclostationary properties. Likewise, the cross-correlation between the transmitted and received signals also shows periodic behaviour in time.

In addition to the previous two-dimensional domains, three-dimensional (3D) representations in terms of time, range, and Doppler frequency, are also considered in HAR with radar system. One of the most common approaches is the 3D radar data cube, which forms a sequence of frames as a function of time that resemble video files. The range-Doppler cube has gained popularity in classification of human activity and gesture [72], [73]. As is shown in Figure 13, the range-Doppler-time cube is a series of range-Doppler map connected along with slow time. Furthermore, the radar point cloud [73], which may be derived from range-Doppler cube [74], has been gaining traction in the field of machine learning, where they serve as a robust input for algorithms designed to recognize or classify objects and activities.
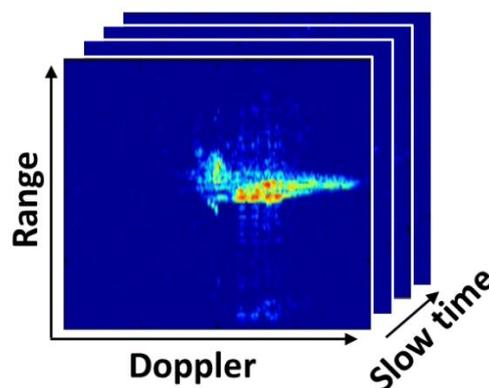
Figure 13 A illustration of 3D range-Doppler Cube

The exploration of 3D representations for radar data extends beyond the conventional range-Doppler cube, incorporating advanced techniques for more nuanced motion detection and analysis. Ronny [75] applied the Radon transform on range-maps generated by radar data. The Radon transform is effective in identifying linear features within an image, making it particularly suitable for detecting translational motions. By applying it to range-maps, which represent the distance of targets from the radar

sensor over time, the method can effectively discern movements that involve a change in position, such as walking or moving objects.

In MIMO radar systems, the use of orthogonal waveforms from a single channel is just one aspect of their capability. MIMO radar offers a variety of other data representations that leverage the spatial diversity of multiple channels, such as Range-Angle Point Clouds [76] and Range-Angle-Time Maps [77]. The multiple channels in MIMO radar systems contribute to improved signal processing capabilities. This includes enhanced SNR, where multiple channels can be used to aggregate signal data, effectively boosting the SNR and improving target detectability. It also refines Constant False Alarm Rate (CFAR) Detection: The spatial diversity of MIMO radar allows for more robust CFAR algorithms, reducing the likelihood of false alarms while maintaining sensitivity to actual targets. These advanced representations and processing capabilities significantly improve upon the limitations of single-channel radar systems, offering more detailed and reliable target detection and tracking.

In the realm of HAR using radar, the exploration of various radar domains such as time-frequency, range-Doppler, and micro-Doppler signatures provides a rich set of features that capture the intricate dynamics of human movements. These domains offer unique perspectives on the data, each revealing different aspects of the underlying activities. However, the true potential of these radar domains is fully realized when integrated with machine learning algorithms. Machine learning serves as a powerful tool to automatically learn and generalize from these high-dimensional feature spaces, thereby enhancing the ability to accurately classify and recognize complex human activities. This synergy between radar domains and machine learning not only elevates the performance of HAR systems but also opens new avenues for research, pushing the boundaries of what is achievable in activity recognition and radar signal processing. An overview of radar HAR with machine learning algorithms is illustrated in Figure 14.
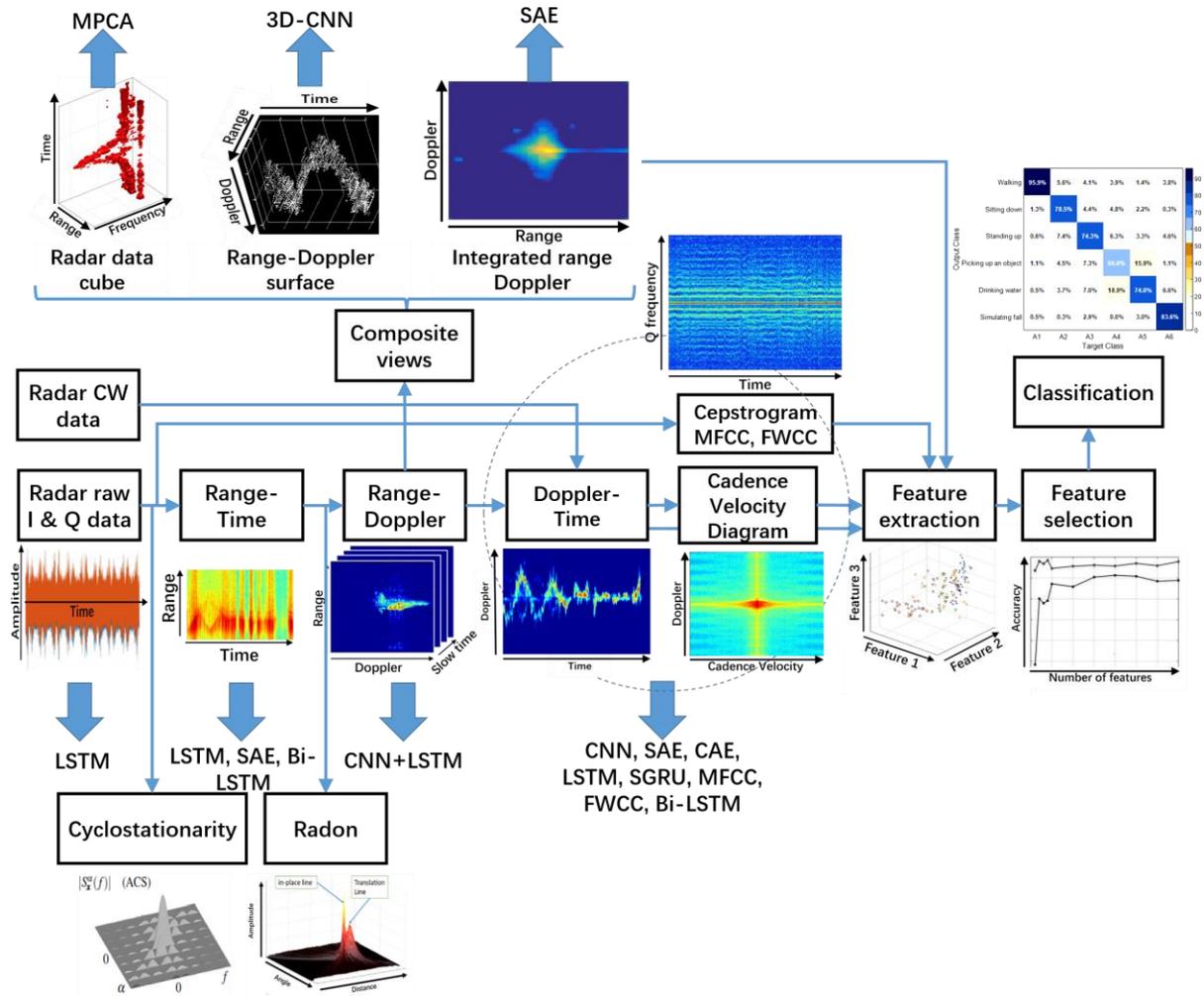
Figure 14 *An overview of the human activity classification with machine learning algorithms with different radar information domains adapted from* [43]*.*

2.4 Machine Learning

Radar systems can provide rich information about the target under observation, and through JTFA, they can be transformed not only into digital information, but also visual representations, motivating the investigation of contemporary solutions using Machine Learning (ML). Machine Learning is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to perform tasks without being explicitly programmed [78]. The primary aim is to allow machines to learn from data so that they can give accurate predictions or decisions. There are three main types of machine learning: supervised learning, unsupervised learning, and reinforcement learning [78].

In the context of radar for humans, ML techniques have increasingly been integrated, offering a robust and automated approach to interpreting complex radar signals. Traditional radar signal processing methods often fall short in capturing the intricate patterns associated with human activities, such as walking, running, or even more subtle movements like breathing [79]. Machine learning algorithms, particularly deep learning models like Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) [28], have shown promise in capturing these nuances. These models are trained to recognize complex human activities by learning intricate patterns in radar signals, thereby enhancing the performance of HAR systems.

In this section, we firstly overview the ML techniques, then we focus on aspects of ML algorithms and that are specific to HAR problems based on radar. By doing so, it demonstrates a basic workflow for employing ML techniques in the analysis of radar data related to human activity.

2.4.1 Supervised Learning

Supervised learning algorithms are a specialized subset of machine learning techniques that operate on a set of labelled data, comprising input-output pairs usually provided through human annotation [78]. These algorithms aim to establish a mathematical model that captures the underlying relationships between the input and output variables. The model is trained using an objective function to guide the optimization process. In the context of radar-based HAR, supervised learning methods such as Linear Regression [80], Decision Trees [81], Support Vector Machines (SVM) [82], and K-nearest Neighbours (KNN) [28] are commonly employed. These algorithms are trained on datasets where radar signals are annotated with corresponding human activities like walking, running, or sitting. The quality of the training data has a significant impact on the model's performance, highlighting the importance of meticulous pre-processing. Supervised learning tasks are generally categorized into regression, which predicts numerical outputs, and classification, which outputs category variable [78]. For example, in HAR systems, classification algorithms are often used to categorize human activities based on different radar domains, producing highly accurate results.

2.4.2 Unsupervised Learning

Unsupervised learning [78] is a category of machine learning algorithms that operate on unlabelled data, aiming to discover hidden patterns or structures within the data. Unlike supervised learning, which relies on labelled examples to learn the mapping

between inputs and outputs, unsupervised learning focuses on finding intrinsic relationships in the data. This makes it particularly useful for tasks like clustering, dimensionality reduction, and anomaly detection. Common algorithms in this category include K-Means clustering [83], hierarchical clustering [84], and Principal Component Analysis (PCA) [28]. The absence of labels means that the quality of the results is often evaluated using different criteria, such as the compactness of clusters or the explained variance in the data. Unsupervised learning is widely used in various domains, including natural language processing, computer vision, and bioinformatics, where labelled data may be scarce or expensive to obtain. In radar-based scenarios, such as human activity recognition and radar systems, it is frequently employed for denoising signals [85], classifying human activities [86], and identifying abnormal events [87].

2.4.3 Other Learning Paradigm

Semi-supervised learning [88] is a machine learning paradigm that combines a small amount of labelled data with a large amount of unlabelled data during training. This approach is particularly useful when acquiring a fully labelled dataset is expensive or time-consuming. Semi-supervised learning algorithms aim to make effective use of both labelled and unlabelled data to improve the performance of model, often leveraging the structure in the unlabelled data to enhance the learning process. Techniques such as label propagation, self-training, and co-training are commonly employed in semi-supervised learning [88]. This method is advantageous in HAR scenarios [89] where labelled radar data may be scarce or costly to obtain, allowing the model to generalize better to new, unseen activities.

Reinforcement Learning (RL) [90] is a type of machine learning paradigm where an agent learns to make decisions by interacting with an environment. The agent receives rewards or penalties based on the actions it takes, aiming to maximize the cumulative reward over time. Although relatively little attention has been paid to RL in radar based HAR, it is still particularly effective in problems where the decision-making process is sequential and the goal is long-term, such as in vehicular networking [91], robotics [92], and certain types of optimisation problems.

2.4.4 Legacy Classification Algorithms

In radar based HAR, the typical machine learning methodology is classification. This process generally commences with the extraction of a predefined set of features from a training dataset. These features are then applied to test data to categorize each

sample into specific classes. In the following section, we will delve into machine learning algorithms that have gained prominence in this specialized research domain.

2.4.4.1 K Nearest Neighbour

KNN [79], [93] is a simple yet effective classification techniques with relatively low computational cost. Figure 15 illustrates the basic principles of KNN where the value of K equals to three, five and seven. The algorithm operates by classifying new data based on the K closest points. To make a prediction for an unknown sample, a training set is formed, consisting of the K closest points. The distance between data points is often calculated using Euclidean distance or Manhattan distance [94]. The majority class among these neighbours is then used to determine the class label of the unknown samples. The choice of K is crucial and varies depending on the specific problem at hand. To mitigate decision clashes, it is generally advisable to select an odd value for K, especially when dealing with binary classification tasks. This ensures that there is a clear majority class, thereby avoiding ties.
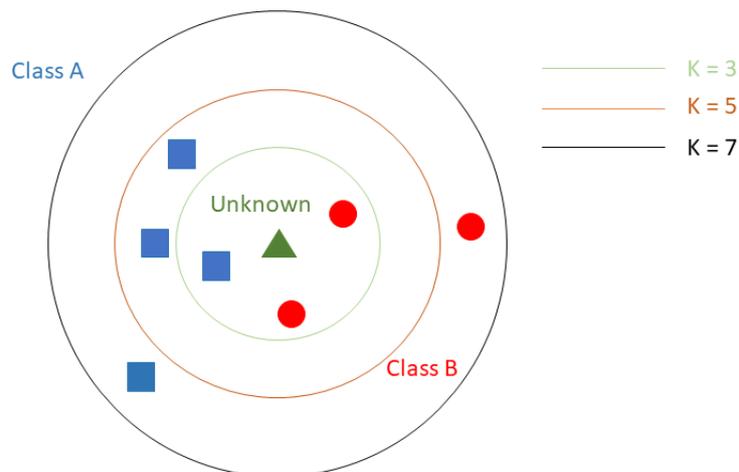
Figure 15 *Scheme of K nearest Neighbour classifiers* [78].

2.4.4.2 Support Vector Machine

The Support Vector Machine (SVM) [28], [79] is a widely recognized and robust classifier, particularly effective in the domain of indoor HAR. The algorithm aims to construct a hyperplane that best separates feature points of different classes, based on their distribution in the feature space. The so-called 'support vectors' are the feature points that lie closest to this decision boundary. These vectors play a crucial role in determining the position and orientation of the separating hyperplane. The objective of

the SVM algorithm is to maximize the margin between the positive and negative hyperplanes, which is achieved by optimizing the position of these support vectors, as illustrated in Figure 16 [95]. The mathematical representation of a linear SVM is given as shown in Eq. (2.34) and (2.35).

$$H = w^T x_i + b = 0 \tag{2.35}$$

$$min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \delta_i \tag{2.36}$$

Eq. (2.34) and Eq. (2.35) demonstrate the hyperplane and max margin of SVM, respectively, where w is the vector of hyperplane. x_i is the input vector and b is bias. δ_i is the slack variable, allowing the input to be closer to the hyperplane. C is the penalty parameter and if C is large, the SVM becomes strict and tries to get all points to be on the right side of the hyperplane, which makes the constraints harder.

SVM is a maximum margin classifier, that aims to find hyperplane that yields the largest possible margin of separation between clusters of data from each respective class. SVM is particularly effective in cases where the data is linearly separable or nearly so, but it can also handle non-linear data by employing kernel methods. A 'kernel trick' can be utilised to map features to a higher dimension. Choices for the kernel are based on the desired hyperplane: gaussian, polynomial (quadratic and cubic) and Radial Basis Functions (RBF) are commonly employed [69].
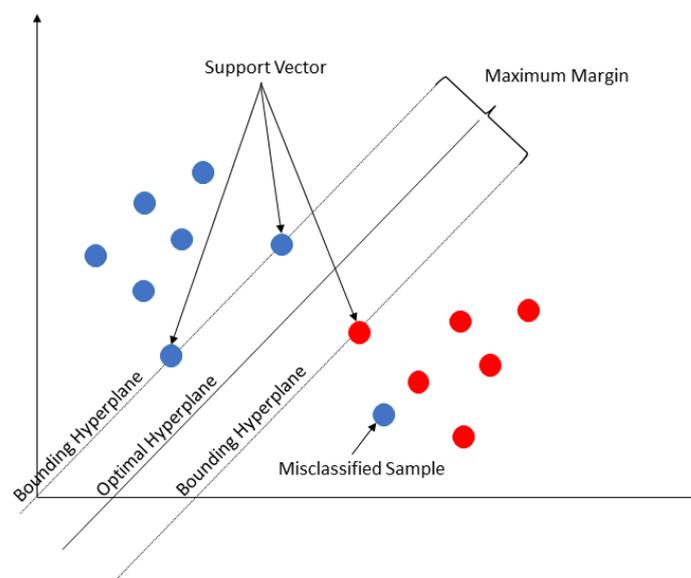
Figure 16 Scheme of Support Vector Machine classifiers.

2.4.4.3 K Means Clustering

The K-Means [96] algorithm is a widely used clustering technique in machine learning for unlabelled data. It aims to partition a set of data points into K non-overlapping subsets (or clusters) based on their feature vectors. Similar to KNN, it is also used Euclidean distance to obtain the clusters in terms of nearest centroid. Algorithm 1 illustrated how K means clustering works.

Algorithm 1 K-means clustering algorithm

Initialization: K-the number of clusters; N-the number of samples in datasets.

set the initial centre $c_1 \dots c_K$ of clusters

    for $i \leftarrow N_1, N_2 \dots, N$ **do**

        find the nearest centre c and assign N_i to the corresponding cluster.

        Update centre of clusters c using mean of the members

    end for

Output clustering result.

The algorithm converges when the assignments no longer change or change minimally between iterations [96]. K-Means is computationally efficient but sensitive to the initial placement of centroids. Various methods, such as K-Means++ [96], have been proposed to provide better initialization to improve the quality of the final clusters.

2.4.4.4 Other algorithms

Apart from the aforementioned algorithms, there exists a series of algorithms that have been successfully applied in HAR research. The Naive Bayes algorithm [78] is a probabilistic classification technique based on Bayes' Theorem [97], which is widely used for classification tasks in machine learning. It assumes conditional independence among the features given the class label, an assumption that is considered 'naïve', hence the name, as is shown in Eq. (2.36):

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)} \tag{2.37}$$

where $p(C_k|x)$ is the probability of class C_k given predictor x. $p(C_k)$ is the probability of class C_k and $p(x)$ is the probability of predictor x. $p(x|C_k)$ denotes the likelihood of samples given the class.

Decision Tree [78] is a type of supervised learning algorithm commonly used for classification tasks. The algorithm constructs a tree-like model to map feature vectors to class labels. It consists of several nodes: internal node and leaf node. Each internal node of the tree corresponds to a feature in the input space, each branch represents a decision rule, and each leaf node signifies a class label. The decision-making process is transparent and can be visualized, making it easier to understand the predictions of the model [78].

In advancement to the Decision tree, Random Forest [78] builds multiple decision trees during the training phase and outputs the mode of the classes of the individual trees for a given input in classification tasks. A common random forest model is shown as Figure 17. Random Forest improves upon the performance of a single decision tree by reducing overfitting and increasing the generalization ability [78]. It achieves this by introducing randomness in two ways: first, by bootstrapping the training data for each tree, and second, by randomly selecting a subset of features for each split in the decision tree. Specifically, a bootstrap sample is a random sample of the data drawn with replacement, usually of the same size as the original dataset. This means that some data points may appear more than once in the sample, while others may not appear at all. This algorithm will construct a decision tree for every training data, and most voted prediction result as the final prediction result [78].
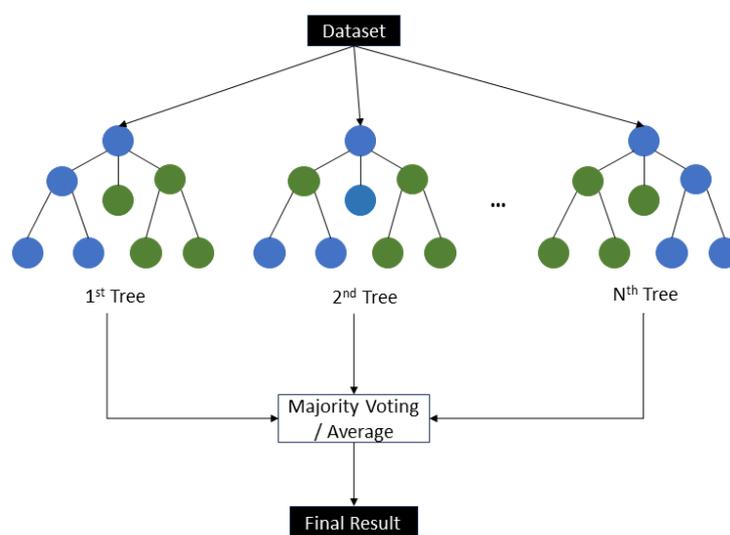
Figure 17 *Scheme of Random Forest classifiers.*

While this discussion has covered several key machine learning algorithms commonly used for classification tasks in radar based HAR, it is worth noting that the field of machine learning is vast and continuously evolving. There are numerous other techniques and variations that have been developed to address specific challenges in classification and other types of machine learning tasks. These include but are not limited to ensemble methods, neural networks, and hybrid learning architectures. For a comprehensive understanding and exploration of these and other advanced methods, readers are encouraged to consult specialized literature and research articles [28], [74], [78], [79], [88].

2.5   Neural Networks

While traditional ML algorithms like SVMs and KNN have been foundational in radar-based HAR, the advent of deep learning has significantly expanded the scope and capabilities of activity recognition systems [79]. Deep learning (DL) models, particularly neural networks, offer the advantage of automatic feature extraction, eliminating the need for manual feature engineering that is often required in traditional machine learning approaches [98]. This has led to more robust and accurate models that can handle the complexities and variabilities inherent in human activities. The transition from ML to DL in radar based HAR represents a paradigm shift, moving from handcrafted features to end-to-end learning, thereby offering a more holistic approach to understanding human behaviours [28], [35].

Neural networks, particularly Convolutional Neural Networks (CNNs) [28], [79] and Recurrent Neural Networks (RNNs) [99], [100], have emerged as powerful tools for Human Activity Recognition (HAR) in radar-based systems. CNNs excel in capturing spatial hierarchies and are highly effective in processing radar data for HAR [74]. RNNs, on the other hand, are adept at handling sequential data, making them well-suited for time-series analysis in radar based HAR [101]. More recently, the application of hybrid CNN-RNNs in conjunction with multi radar representations has opened new avenues for HAR [102], [103]. These neural network architectures offer robust and accurate models for classifying and predicting human activities based on radar data. Consequently, DL has not only become the prevailing approach in the ongoing work in ML research, but also gained substantial traction in radar-based classification studies. This shift has expanded the horizons of what is achievable in the field, making deep learning a focal point for cutting-edge research in radar-based Human Activity Recognition [28], [74].

2.5.1 Fundamentals

2.5.1.1 The neurons

A neuron is the fundamental block of a neural network. It takes a set of inputs x_1, x_2, …, x_i, and applies a set of weights w_1, w_2, …, w_i, with a bias b. The combination would be passed through an activation function f and obtain the result. It can be defined as Eq. (2.37).

$$Output = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{2.38}$$

Weights and biases [104] are pivotal components that shape the transformation and processing of input data. Weights are parameters that adjust the connection strength between neurons, determining the influence one neuron exerts on another. Initially set to small random values, these weights are crucial for breaking symmetry and kickstarting the learning process. Biases, another set of parameters, work alongside weights to adjust the neuron output. They shift the activation function, aiding in better data fitting, and are also typically initialized to small values or zeros.

2.5.1.2 Weight and Bias

The process of updating weights and biases is integral to learning procedures. It begins with forward propagation, where inputs are processed through the network using current weights and biases to generate predictions. The network performance is then evaluated by comparing these predictions against actual target values, calculating the loss or error. This is where backpropagation [105] comes into play, a critical learning stage where the network adjusts its parameters based on the error. The gradient of the loss function is calculated with respect to each weight and bias, using the chain rule of calculus.

The gradient [104], [106] represents the direction and rate of the steepest increase of a function. In the context of neural networks, this function is typically the loss function, which measures the difference between the predictions and the actual data. By employing gradient descent, the network iteratively adjusts its parameters, taking larger steps when far from the optimal solution and smaller, more precise steps as it nears the minimum. This adaptive approach, facilitated by backpropagation, ensures efficient and effective learning, crucial for deep neural network architectures.

Back-propagation [105] is the algorithm used for minimizing the error in predictions of the neuron network. It adjusts the weights and biases in the direction that minimally decreases the error. The weight and bias update can be represented as:

$$w_{i-new} = w_{i-old} - \alpha \frac{\partial E}{\partial w_i} \tag{2.39}$$

$$b_{new} = b_{old} - \alpha \frac{\partial E}{\partial b} \tag{2.40}$$

where α is the learning rate and E is the loss function. The learning rate [107] decides how much the weights should change during training. High rates might lead to miss the best solution, but low rates extend training time. The loss function [107] measures how far off the predictions of model are from the actual results. It helps to know if the model is getting better or not.

2.5.1.3 Activation Function

Activation functions introduce non-linearity into the network, allowing it to learn from the error and make adjustments. Common activation functions include sigmoid, tanh and rectified linear unit (ReLU). ReLU outperforms the other two methods, since sigmoid is defined between [0,1] saturating at the bounds and tanh is between [-1,1]. They both encounter the vanishing gradient problem, where the gradients become too small for the network to learn effectively. ReLU helps mitigate this problem because its gradient is either zero (for negative inputs) or one (for positive inputs). The activation functions are defined as below:

$$sigmoid: \sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.41}$$

$$Tanh: tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2.42}$$

$$ReLU: ReLU(x) = \max(0, x) \tag{2.43}$$

The non-linear activation functions are pivotal, allowing networks to capture and represent complex patterns within the input data, enabling them to model complex relationships with high efficacy.

2.5.2 Convolutional Neural Networks

2.5.2.1 Feed-Forward Neural Networks

Before we introduce CNN, it is essential to understand Feed-Forward Neural Networks (FNN). It [104] serves as one of the simplest types of artificial neural networks. In a Feed-Forward Neural Network, the connections between the nodes, also known as neurons, do not form any cycles. This is crucial for differentiating them from recurrent neural networks. The network consists of an input layer, one or more hidden layers, and an output layer [104]. Each neuron in a layer is connected to every neuron in the subsequent layer, with each connection having an associated weight. The layers are fully connected, meaning each neuron in a layer is connected to every neuron in the adjacent layers. The neurons use an activation function to transform the weighted sum of their inputs [104].

While Feed-Forward Neural Networks offer a general-purpose architecture suitable for a wide array of tasks, they often fall short in capturing spatial and temporal dependencies in data. This limitation is particularly evident in image and video recognition tasks, where the spatial arrangement of pixels is crucial for accurate classification [104]. To address these challenges, Convolutional Neural Networks (CNNs) have been developed as a specialized kind of neural network for processing data with a grid-like topology, such as an image [104].

CNNs are a specialized type of neural network architecture that has proven highly effective in tasks related to image recognition, object detection and various visual tasks [104]. Inspired by biological processes, namely the organization of the animal visual cortex, CNNs were first introduced by LeCun et al. in their seminal work on LeNet-5, a pioneering model for digit recognition [108].

2.5.2.2 The structure and applications

The architecture of a CNN is designed to process data with a grid-like topology, such as an image, which is a grid of pixels. The network employs three fundamental types of layers: convolutional, pooling (or subsampling), and fully connected [109].

- Convolutional Layers: These are the core building blocks of a CNN. The layer parameters consist of a set of learnable filters (or kernels), each of which has a small receptive field, but extends through the full depth of the input volume. During the forward pass, each filter convolves across the width and height of the input volume and computes dot products between the entries of the filter

and the input, producing a 2-dimensional activation map for each filter. All activation maps are stacked along the depth dimension to produce the output volume.

- Pooling Layers: Following each convolutional layer in a CNN, it is common to have a pooling layer for down sampling. Pooling layers reduce the dimensions of the data, which decreases the computational complexity for upcoming layers. Pooling can be of different types, such as Max Pooling, Average Pooling, etc. Max pooling takes the largest element from the rectified feature map, while average pooling takes the average.

- Fully Connected Layers: After several convolutional and pooling layers, the architecture typically has one or more fully connected layers. These layers flatten the high-level features learned by the preceding convolutional layers to vector form and perform classification on the image.
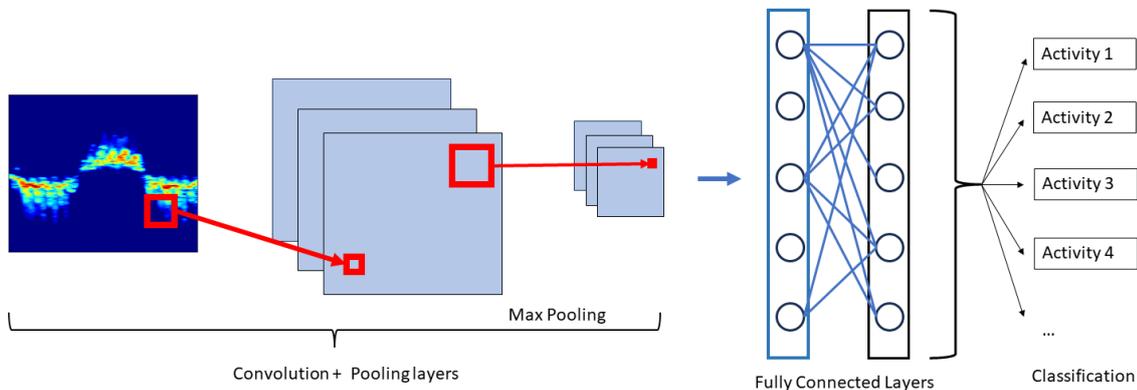
Figure 18 *The common structure of the CNN application in HAR with radar.*

As is shown in Figure 18, in the context of radar based HAR, CNNs offer significant advantages for feature extraction and classification. The spatial hierarchies captured by CNNs are particularly useful for recognizing complex human activities that generate intricate radar signatures [74]. For instance, CNNs can be employed to automatically learn and identify the micro-Doppler signatures associated with different human activities, such as walking, running, or falling [28]. This eliminates the need for manual feature engineering [31], which is often a cumbersome and error-prone process. The application of CNNs in radar-based HAR is an emerging area that promises to push the boundaries of what is achievable in terms of both accuracy and computational efficiency [110].

2.5.3 Recurrent Neural Networks and Long Short-Term Memory

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are classes of artificial neural networks that are designed to recognise patterns in sequences of data, such as text, genomes, handwriting, or the spoken word [104]. They are particularly well-suited for tasks that require the consideration of context or the memory of prior inputs.

RNNs, introduced by Hopfield in 1982 [111], possess a unique feature that distinguishes them from other neural networks - loops in the network of nodes. This means that information can circulate in the network for an arbitrary amount of time, enabling RNNs to work with sequential data by maintaining a 'memory' of prior inputs. This design addresses the necessity to consider context in tasks like language modelling, where the meaning of a word depends on the preceding words.

The basic RNN structure includes input, hidden, and output layers. However, RNNs face the challenge of long-term dependencies due to the so-called vanishing and exploding gradient problem [112], which makes it difficult for the network to learn and tune the parameters when dealing with long sequences.

To overcome this problem, Long Short-Term Memory (LSTM) networks, a special kind of RNN, were introduced by Hochreiter & Schmidhuber in 1997 [113]. The LSTM is designed to have more prolonged memory, and hence, it can handle the issues related to learning long-term dependencies.

The fundamental difference between LSTMs and traditional RNNs lies in the hidden layer. Instead of having a single neural network layer, LSTM has a cell state, which can be thought of as a 'conveyor belt' running through the entire chain with some minor linear interactions. LSTM manipulates the cell state through structures called gates (input gate, forget gate, and output gate) which can learn to regulate the flow of information into and out of the cell state. Each gate is a feed-forward neural network with sigmoid activation, outputting a value between 0 and 1, representing how much of each component should be let through.

LSTM uses a gate structure to achieve its function, which contains three types of gates: input gate i, output gate o, and forget gate f [113]. By controlling the gates, the cell can determine the storing, writing, and reading operation of information. For each time step t, x_t is the input to the memory cell layer, and the updated states of each parameter are shown in the following Eq. (2.22) to (2.23).

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \qquad (2.44)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \qquad (2.45)$$

$$c_t = f_t c_{t-1} + i_t tanh(W_{xc}h_{t-1} + W_{xc}x_t + b_c) \qquad (2.46)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \qquad (2.47)$$

$$h_t = o_t tanh(c_t) \qquad (2.48)$$

where $\sigma(x)$, W and b represents the sigmoid function, weight, and bias factor, respectively. For the input x_t, the input gate i_t can accumulate new value flows into the memory cell. The forget gate f_t determines what needs to be discarded from the memory of a cell, which means it can force the memory cell to forget things that are not significant. c_t demonstrates how the memory of cell updates in terms of the new input and the previous value. The output gate o_t determines what should be output to the next cell from the current memory cell and h_t is the hidden output of the current cell.

In radar-based HAR, RNNs and LSTM networks offer unique capabilities for modelling temporal dependencies in radar signals [101], [114]. These architectures are particularly well-suited for capturing the sequential nature of human movements as reflected in radar data. RNNs can be applied to model the time-series data of radar signals, making them effective for recognizing activities that involve sequential or periodic movements, such as walking or cycling [39]. However, traditional RNNs suffer from the vanishing gradient problem, which makes them less effective for learning long-term dependencies.

LSTMs, an advanced form of RNN, overcome this limitation by using memory cells that allow them to capture long-term temporal dynamics effectively. This makes LSTMs highly suitable for complex activities that involve a series of actions in a specific order, such as picking up an object followed by walking [39]. For example, LSTMs have been used to classify complex activities by analysing the micro-Doppler signatures generated by human movements in radar data. Furthermore, hybrid models combining CNNs, and LSTMs have also been explored to leverage both spatial and temporal features for more robust and accurate HAR [98], [102].

2.5.4 Attention Mechanism and Transformer

The attention mechanism [115] and the transformer [116] architecture represent significant advancements in the field of deep learning, particularly in natural language

processing (NLP) and beyond. These concepts have revolutionized how models handle sequential data, offering improvements in efficiency, scalability, and performance over traditional recurrent neural network (RNN) approaches [115].

Introduced as a solution to the limitations of RNNs and their variants (like LSTMs and GRUs) in processing long sequences, the attention mechanism allows models to focus on different parts of the input sequence when predicting each part of the output sequence [115], [117]. This is akin to how humans pay attention to specific portions of information while ignoring others when comprehending text or images. The key idea is to dynamically weigh the significance of input elements based on their relevance to the task at hand, enabling the model to make more contextually informed decisions. Attention mechanisms have been particularly effective in tasks such as machine translation, where the model needs to consider the entire input sequence to accurately translate a sentence.

Building on the success of attention mechanisms, the transformer architecture, introduced in the seminal paper "Attention is All You Need" by Vaswani et al. in 2017 [116], fully utilizes attention mechanisms to process sequences, eschewing recurrence entirely. Transformers consist of two main components: an encoder that processes the input sequence and a decoder that generates the output sequence. The core of the transformer is the self-attention mechanism, which allows the model to weigh the importance of different words in the input sequence when processing each word [115], [118]. This enables the transformer to capture complex dependencies and relationships within the data.

Transformers have several advantages over their predecessors. They can process entire sequences simultaneously rather than sequentially, which significantly improves training speed and efficiency. This parallel processing capability, combined with their ability to capture long-range dependencies in data, has made transformers the architecture of choice for a wide range of applications, including but not limited to text summarization, sentiment analysis, and even areas outside NLP like computer vision [116], [117].

The introduction of transformers has led to the development of large-scale pre-trained models like BERT [119], GPT [120], and others, which have set new benchmarks across numerous NLP tasks. These models leverage vast amounts of data to learn rich representations of language, which can then be fine-tuned on specific tasks to achieve state-of-the-art performance.

In summary, the attention mechanism and transformer architecture have marked a paradigm shift in how sequential data is processed, offering a more flexible and powerful framework for a variety of complex tasks across different domains.

2.6 Classification and Performance

2.6.1 Fitting and Validation

The concept of fitting in machine learning is fundamentally linked to the adjustment of weights and biases within a classifier [104], [107], [121]. Fitting can be visualized as the process of defining a margin that separates different classes in a dataset. The fitting process can lead to three distinct outcomes: underfitting, good fit, or overfitting.

- Underfitting [122] occurs when the margin of classifier is too broad, failing to encompass all relevant observations within the primary cluster of a class. This results in a high rate of false positives and a general misclassification of data points, often due to an overly relaxed bias parameter. Conversely, a good fit is achieved when the classifier's margin is appropriately tight around the primary cluster, accurately identifying the class of interest while minimizing false positives. This balance ensures that the classifier is sensitive enough to detect the class without being overly inclusive.
- Overfitting [121], on the other hand, is a condition where the classifier overly extends its margin in an attempt to include outlier points. This leads to a deceptive initial accuracy, as the classifier seems to recognize the class of interest effectively. However, this comes at the cost of increased false positives due to the overlap of margins with other class clusters. Overfitting models may not necessarily account for all outliers, leading to a skewed understanding of the data.

In the training stage of machine learning applications, classifiers often oscillate between overfitting and underfitting. Achieving a good fit requires numerous training iterations, where the classifier gradually converges to an optimal balance through an iterative process of fitting adjustments, guided by gradient updates. This intricate process of tuning and training the classifier is critical for developing a model that accurately represents and predicts the underlying patterns in the data.

Validation [92] is a critical process in machine learning, designed to evaluate the performance on a dataset separate from the training data. Its primary aim is to gauge the ability to generalize to new, unseen data. The process begins with splitting the

available dataset into subsets: typically, a training set and a validation set, and occasionally a third subset, the test set. The performance is evaluated using metrics appropriate to the problem, such as accuracy, precision, recall, or mean squared error. Based on this evaluation, the model may be tuned by adjusting hyperparameters, altering its architecture, or revising the feature selection process. The goal is to refine the model to perform well not just on the training data but also on unseen data. This iterative process of training, validating, and tuning continues until the model achieves satisfactory performance on the validation set. Validation is vital for several reasons [123]:

- It helps detect overfitting, where a model is too tailored to the training data and fails to generalize.
- It allows for the comparison of different models or configurations to select the best performer.
- It is essential for hyperparameter tuning, providing a feedback loop independent of the training data.
- Lastly, it offers an estimate of the model's expected real-world performance.

There are various types of validation methods [123]. Holdout validation is the simplest, involving a straightforward split of the dataset into training and validation sets. K-Fold Cross-Validation divides the data into 'K' subsets, with the model trained and validated 'K' times, each time using a different subset for validation. Leave-One-Out Cross-Validation is a specific case of K-fold cross-validation where 'K' equals the number of data points, with each data point used once as a single validation set.

In summary, validation is an indispensable step in the machine learning workflow, ensuring the development of robust, generalizable models suitable for real-world applications.

2.6.2 Measures of Success

This section delves deeper into the performance metrics previously introduced. Confusion metrics [124], related to the classification efficacy of a specific class, are delineated in Equations (2.49) to (2.51). These metrics are further simplified in the context of a binary classification problem, as depicted in Table 3, where rows and columns represent output and target classes, respectively. Sensitivity, also known as Recall, quantifies the rate at which the class of interest (Class A in this instance) is correctly classified. In contrast, Specificity is associated with the accurate

identification of classes other than the class of interest. Additionally, Precision, or Positive Predictive Value (PPV), is calculated as the proportion of true positives in relation to the total count of samples classified as belonging to the class of interest.

Table 3 A Binary Confusion Matrix

Output	A	B
A	True Positive	False Positive
B	False Negative	True Negative

$$Sensitivity = \frac{TP}{TP + FN} \qquad (2.49)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (2.50)$$

$$Precision = \frac{TP}{TP + FP} \qquad (2.51)$$

In addition to the previously discussed metrics, this research also introduces two common measures to assess the overall performance of classifiers. The F-measure, as referenced in [125], is employed to provide a balanced view of the performance by combining sensitivity and precision through their harmonic mean. This metric effectively highlights the balance between 'missing positives' and 'false alarms' in the classification process. Besides, classification accuracy [124] is utilized to gauge the rate of correctly classified instances across all classes. This metric is particularly valued for its ability to provide a balanced overview of classifier performance. Notably, when applied to a single class, classification accuracy transforms into a measure of sensitivity, offering insights into the classifier's ability to correctly identify positive instances within that class.

$$F1\ score = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \qquad (2.53)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (2.54)$$

There are also several metrics that specialized employed to some filed. For example, the Jaccard Index [126], standing out for its effectiveness in quantifying the similarity between two sets, is a widely used metric for quantifying the similarity and diversity between two sets. It is particularly effective in the field of image segmentation, where it calculates the ratio of the intersection to the union of the predicted and actual segmentation areas. This index ranges from 0 to 1, where a value of 1 denotes perfect overlap and 0 indicates no overlap at all. A high Jaccard Index signifies that the segmentation algorithm has accurately delineated the region of interest, closely matching the ground truth. Similarly, the F1 score, or Dice coefficient [123], offers another perspective on measuring success in segmentation tasks. While the Jaccard Index focuses on the proportion of overlap, the Dice coefficient emphasizes the balance between precision (the proportion of true positive results) and sensitivity (the ability to correctly identify positives). Both metrics are integral in evaluating the efficacy of segmentation algorithms, providing insights into their accuracy and reliability in differentiating and identifying regions within an image.

2.7 Summary

This chapter provides an overview covering the fundamental principles of radar and signal processing procedures, with a focus on Frequency Modulated Continuous Wave (FMCW) radar, a type employed in this project. The study delves into the representation of radar, particularly in micro-Doppler signatures that hold essential information about human activities. It further explores the utilization of machine learning and deep learning techniques for activity recognition using radar data, laying the groundwork for understanding their application. This includes an introduction to the basics of artificial intelligence, various algorithms, and the significant role of data selection and fusion to boost the accuracy and resilience of learning models. The significance of model fitting, validation techniques, and performance evaluation metrics is also emphasized, serving as a foundation for the subsequent chapters on machine learning and deep learning techniques in human activity recognition tasks.

3. Machine Learning Applications in Assisted Living with Radar

In this Chapter, a literature survey is presented covering the state-of-the-art solutions that apply different ML techniques in assisted living with radar, especially in HAR, with a summary of the reviewed papers presented in Table 5. The survey considers sensing technologies and classification algorithms, providing a structured lens through which we scrutinize the evolution of radar-based human recognition—from its previous stages [79], [127], [128] reliant on traditional methods to its current state, augmented by advanced machine learning algorithms. By doing so, we aim to shed light on the pivotal role these techniques have played in enhancing both the accuracy and computational efficiency of such systems. Moreover, this chapter compels us to delve into the existing challenges in this rapidly evolving research field. In this chapter, the question also prompts the exploration of emerging trends and potential future directions for this interdisciplinary field. Thus, this overarching question not only frames the scope of this literature review but also serves as the motivation behind each section, ensuring a cohesive, thorough, and forward-looking analysis.

This chapter is structured as follows: In Section 3.1 it begins with a comparative analysis of various sensors used in HAR, highlighting their capabilities and limitations. Following this, the chapter delves into existing research in the radar domain, offering insights into the state-of-the-art methodologies and developments in Section 3.2. Section 3.3 places a significant focus on gait analysis, examining its critical role in HAR and the nuances of its interpretation through radar technology. Then Section 3.4 explores methods to enhance accuracy in HAR, specifically through the lens of feature selection and fusion techniques. Section 3.5 is a summary that not only synthesizes the key discussions but also presents emergent research questions and identifies gaps in current studies, paving the way for future research directions in radar based HAR.

3.1 Why Radar in assisted living?

Activity recognition sensors generally falls into two categories: wearable and non-wearable sensors [129], [130]. Wearable sensors, depending on their placement on the body, in pockets, or around the neck, gather data at varying resolutions from human movements. The quality and type of data collected can differ significantly based on the sensor's location relative to the torso and limbs. These sensors, equipped with accelerometers, magnetometers, and/ or gyroscopes, measure acceleration, magnetic

fields, and angular movements, respectively. They can also directly measure the skin temperature and arterial movements for vital signs monitoring.
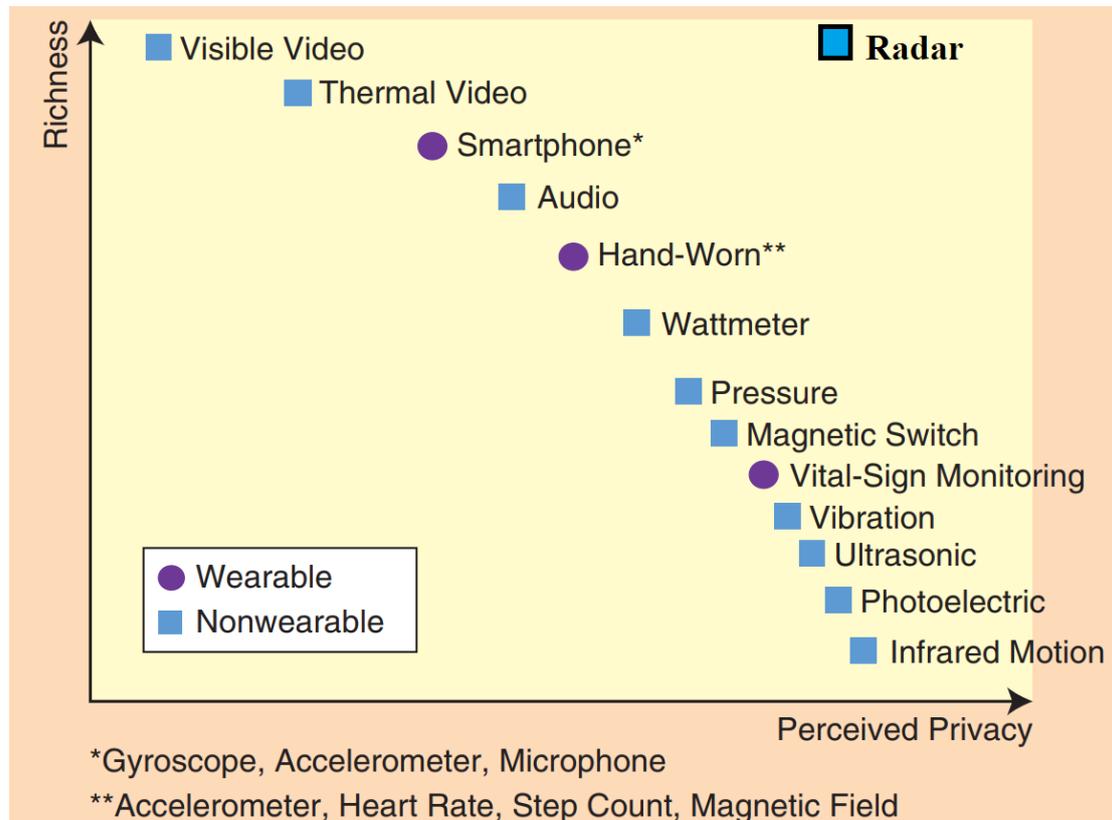
Figure 19 Perceived user privacy and richness in information far different types of sensors used for indoor monitoring and assisted living adapted from [131]

Non-wearable sensors, on the other hand, maintain a distance from the subjects, making them less invasive and requiring less management by end-users, who often are older individuals with impaired cognition in the context of assisted living [132], [133]. These sensors also rely less on temporary power sources like batteries, an added benefit from a management perspective. Figure 19 illustrates the richness of the sensors versus the perceived privacy of a person using these sensors. It is clear from the figure that sensors that provide rich information about a person are usually not perceived as privacy-preserving. For example, a video camera that allows for the recognition of almost any human activity in its field of view cannot be used in most rooms due to the perceived invasion of privacy by the users. Radar stands out as it provides rich information and good perceived privacy because it does not take plain images. Also, most users have Wi-Fi routers at home working on similar principles. Hence, we assume that radar will be better accepted and seen as less invasive. Furthermore, with the fusion of communication and sensing in 6G. People will have access points deployed in their home environment which will rely on radar sensing

techniques. However, radar systems at present require the installation of a bespoke system complicating the deployment process.

Looking at the most prevalent solutions for assisted living that offer a transformative potential to improve the quality of life for the elderly and chronic patients, we will look at the challenges going forward for indoor monitoring. Table 4 depicts the challenge comparison between camera and radars.

Table 4 Current main challenges for camera and radar technologies.

Wearable Sensors [134] - Challenges	Camera [135] - Challenges	Radar [28] - Challenges
• **Data Accuracy:** Ensuring the accuracy of data collected from wearable sensors is crucial. Factors like sensor placement, calibration, and interference can affect data quality. • **Battery Life:** Wearable sensors often have limited battery life, requiring frequent recharging or battery replacement, which may not be convenient for elderly users or those with limited mobility. • **User Compliance:** The effectiveness of wearable sensors depends on consistent usage. Forgetting to wear the device or discomfort caused by the device can lead to data gaps. • **Environmental Factors:** Conditions like humidity, temperature, and physical obstructions can affect sensor performance, leading to inaccurate data.	• **Coverage area and depth-sensing range**. Differently from wearable devices, vision-based sensors have a limited coverage area, and many sensors may be required to monitor the whole apartment, leading to higher costs of installation. • **Occlusions:** Vision-based sensors suffer from occlusions, for example, from pieces of furniture. The coverage area may also be limited by the presence of some occluding objects, which are temporarily interposed between the subject to be monitored and the sensor. • **Skeleton data reliability**. Many algorithms based on Kinect sensors rely on skeleton data, which can be used to extract the position and posture of the human. However, for the skeleton information to be correctly estimated, the person should be facing the sensor. Moreover, the estimation algorithm can detect some spurious skeletons that are objects. • **Privacy:** The people surveyed fill a sense of privacy invasion	• **Presence of strong scatterers and clutter** in indoor environments which may generate multipath and ghost targets, or obscure the person to be monitored from the sensor, which can also be a problem for RGB-D sensors; • **Multi-occupancy:** the possibility of having pets or other people (e.g., visitors, multiple elderly) moving inside the monitored area, thus complicating the signature, and generating false alarms. Again, this could potentially be a problem for RGB-D sensors as well; • **Emission regulations:** the compliance of the selected radar waveforms with directives from the telecommunication regulatory bodies, with potential constraints in terms of the achievable bandwidth and transmitted power, hence limiting the range resolution and the Signal to Noise Ratio (SNR); • **Aspect angle dependence:** The dependence of the micro-Doppler signature on the cosine

	with cameras. • **Law**: The data/image rights that apply to the cameras and ethics linked to its operation are not mature yet and could quickly become a legal minefield in the future. Cameras would also be able to tackle multi-occupancy	of the aspect angle between the velocity vector of the movement and the line-of-sight of the radar, which in some cases can significantly attenuate the signatures and make them unsuitable for feature extraction aimed at fall detection; • **Fine-grained activity classification in a continuum:** the possibility to reliably detect a fall, irrespective of the type of movement or activity performed before, and of the dynamics of the fall itself (falling forward or backward, tripping rather than losing balance or consciousness, falling while sitting or standing up from chairs or sofas). This would imply developing fall detection procedures that can take into account the actual dynamics of elderly people moving, for instance, the effects on the radar signatures of using walking assistive devices.

Among non-wearable sensors, radar has received increased attention [79] as a potential alternative to established sensors like video cameras. Its appeal lies in its insensitivity to lighting conditions and its easy integration into homes [114]. Modern radar systems can blend inconspicuously into smart home environments, functioning much like a common Wi-Fi router [136]. In addition, radar tends to pose fewer privacy concerns than cameras, as it does not collect explicit images or videos of users and their private environments [137], [138], [139].

3.2 Review of Radar data domains in HAR

The ability to exploit different radar data domains in current HAR systems can be attributed to several key advancements.

1. modern radar systems have undergone significant improvements in resolution, range, and sensitivity, enabling the collection of more granular and detailed data. This is supported by research in ambient intelligence technologies, which focus on constant monitoring through networks of sensors and actuators, thereby enhancing the capabilities of radar systems in HAR applications [64], [140], [141].

2. The advances in computational power have made the processing of complex radar data in real-time possible, opening up new possibilities for HAR applications [28], [56], [142].

3. Advanced machine learning algorithms have played a crucial role in handling the intricacies of different radar data domains and automatically extracting relevant features for activity recognition [79], [143].

4. The availability of large, labelled datasets has further facilitated the development of robust models capable of navigating these diverse data domains [114].

5. Interdisciplinary research, bringing together expertise from fields like signal processing, machine learning, and human-computer interaction, has led to more comprehensive approaches to HAR [101], [114].

The democratization of access to these technologies through open-source software libraries has also allowed a broader range of researchers to contribute to ongoing advancements in the field.

3.2.1 Spectrogram with Machine Learning

In 2009, Y. Kim and H. Ling [144] pioneered the use of micro-Doppler signatures from continuous wave (CW) radar. Since then, classification efforts have primarily focused on spectrograms, which are Doppler-time representations of received signals derived using the Short-Time Fourier Transform (STFT). This method provides a straightforward effective means of data representation, particularly useful in scenarios with significant interclass variations. In their study [144], they combined micro-Doppler signatures with an SVM classifier to categorize seven different activities. They extracted six critical physical features from the radar data, achieving an accuracy of around 90% when these features were used in combination. This foundational research guided subsequent studies by highlighting the significance of classifying activities, addressing challenges related to radar aspect angles, and pioneering through-the-wall [145] measurement techniques.
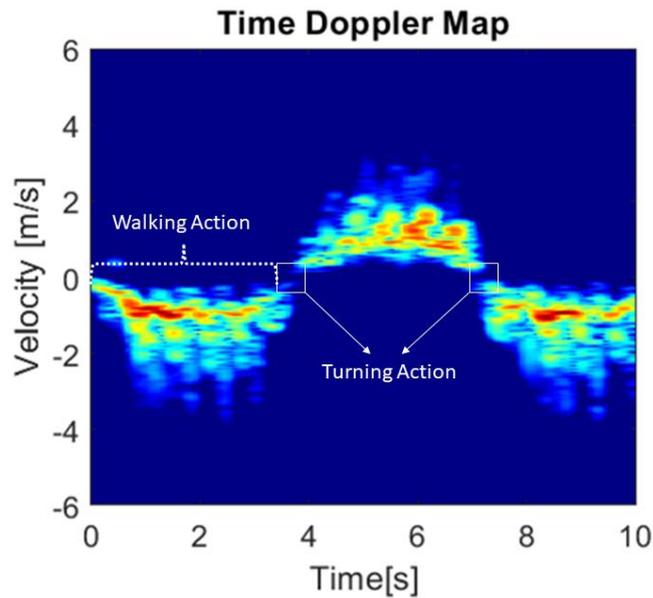
Figure 20 *An illustration of walking spectrogram.*

Human activity recognition using traditional classifiers typically begins with the creation of detailed handcrafted features. These features are derived from the physical attributes of human activities or from pattern features based on statistical calculations and specific micro-Doppler signature images. Fioranelli and colleagues [146] introduced a novel approach that utilizes singular value decomposition (SVD) of the spectrogram. This method simplifies the spectrogram's components to highlight key features. Additionally, Principal Component Analysis (PCA) can be used as an alternative feature extraction technique. PCA conducts a linear transformation of the data to identify the direction with the greatest variance. For single-channel data, the original PCA or 1-D PCA is applied to determine the principal components. In contrast, 2-D PCA and generalized 2-D PCA extend the concept of 1-D PCA to multi-channel data, such as images. These methods represent significant advancements in extracting valuable information from complex radar data, underscoring the impact of feature extraction on classification accuracy.

Conventional classifiers like KNN (K Nearest Neighbour) and SVM (Support Vector Machine) are favoured for their straightforward implementation and low computational demands. KNN, a non-parametric method, predicts labels based on the majority class among the K nearest training samples. Studies [147] and [148] have assessed KNN's effectiveness in radar datasets for human activities, particularly in fall detection scenarios. KNN also acts as a benchmark for more complex algorithms like deep neural networks. In contrast, SVM, which is more intricate, constructs a

61

hyperplane for data separation, utilizing chosen support vectors. Various kernel functions, such as RBF, Quadratic, and Cubic, can enhance SVM classifiers by mapping data into higher dimensions, aiding in diverse classification challenges. SVM stands out in human motion classification among conventional classifiers, often outperforming neural networks, especially when training data is limited. References [147], [144], [127] and [149] have successfully employed SVM with different kernel functions for detecting human targets, classifying human activities, and recognizing micro-gestures. Unlike previous experiment, the work in [150] takes CNN as feature extraction instead of classification. The spectrogram data as input and passed through an auto-correlation function, followed by a CNN to extract action-related features. Finally, a Random Forest classifier was used to predict the action class.

3.2.2 Spectrogram with Deep Learning

The advancement of deep learning models has significantly enhanced system performance in various domains, including visual object classification and speech recognition, even from high baseline levels [104]. These deep models mimic the human brain data processing mechanisms, typically featuring multiple hidden layers that learn complex, high-level characteristics of input data [104]. Training a deep learning-based classifier requires a substantial dataset, leading to a demand for considerable computational resources, such as extensive parallel computing units and storage capacities. Nonetheless, one of the key advantages of deep learning models is their ability to bypass the need for manually crafting features, streamlining the model development process.

Spectrograms can be interpreted as temporal sequences of Doppler frequency distributions, aligning well with sequence models like Recurrent Neural Networks (RNN) and their advanced variant, Long Short-Term Memory (LSTM) Units. These models are adept at capturing the temporal dynamics within spectrograms, enabling them to effectively model time-variant features in radar signals. This capability is particularly beneficial for tasks that require the prediction or detection of temporal patterns in radar data, such as tracking moving targets over time. For instance, in [79] and [151], LSTM algorithms are employed to classify human activities using radar micro-Doppler spectrograms. Additionally, [152] explores the use of Gated Recurrent Units (GRU), a streamlined version of LSTM, achieving performance comparable to the LSTM results in [151]. Bidirectional LSTM models, as discussed in [100], further enhance this approach by analysing temporal sequences in both forward and reverse directions.

On the other hand, spectrograms, as visual representations of the distribution of Doppler frequencies in radar signals over time, are also highly adaptable for computer vision applications. In these spectrograms, each 'pixel' represents the amplitude of a specific Doppler frequency at a given time, akin to a 2D image. This format is particularly conducive to analysis by Convolutional Neural Networks (CNNs), a deep learning model adept at image processing. CNNs excel in automatically identifying spatial patterns within spectrograms, such as frequency modulations or characteristic energy distributions, crucial for tasks like radar signal analysis and target recognition. A notable advancement from of [144] occurred six years later in [153], where the researchers transitioned to a Deep Convolutional Neural Network (DCNN), maintaining high HAR performance, and improving the differentiation between humans and other objects to approximately 97.6% accuracy. Further, studies [154], [155] and [79] have successfully employed CNN algorithms for processing various spectrogram types.

The evolution of machine learning methods has led to the exploration of hybrid approaches that leverage both spatial and temporal features of spectrograms. Convolutional LSTM networks, which merge the spatial feature extraction process of CNNs with the temporal modelling capabilities of LSTMs, exemplify this trend. In [156], a novel combination of 1-D CNN with LSTM is proposed, achieving a remarkable 98.28% accuracy, surpassing traditional 2-D CNN models. Further advancing HAR methodologies, [157] introduced a multi-domain feature attention fusion network (MFAFN). This model overcomes the limitations of single-range or velocity features in human activity description. Additionally, [158] proposed an efficient network based on a lightweight hybrid Vision Transformer (LH-ViT), aiming to simultaneously enhance HAR accuracy and reduce network complexity.

In these varied scenarios, converting raw radar signals into spectrograms enables deep learning models to effectively utilize the frequency-time composition of the signal, a critical aspect for proficient radar signal processing. The choice of approach is heavily dependent on the specific problem and the characteristics of the radar application. The literature is rich with examples where deep neural network-based architectures are employed to interpret radar data, treating it either as image data or as temporal sequences.

3.2.3 Other Domains

The advancement of high-precision radar, driven by the automotive industry anti-collision systems and the evolution of software-defined radio platforms, along with a

surge in computational capabilities, has opened up access to more data-rich domains for feature extraction. This encompasses a diverse array of data types, from 1-D information such as raw data [101], [102], [159] , to 2-D information such as range-time [159], [160], range-Doppler [160], [161], [162], Cadence Velocity Diagrams [163], and cyclostationarity [70], extending 3-D formats, [143], [164], like the range-time-Doppler cube and cloud point.

Short-Time Fourier Transform (STFT), while beneficial, faces the challenge of time-frequency (TF) resolution trade-off, meaning we cannot optimize time and Doppler frequency resolutions simultaneously. More advanced TF distributions, as discussed in [165], were proposed to tackle this problem. However, the spectrogram still has limitations. One significant limitation is the loss of phase information, which can be crucial for understanding intricate details of human movement. Phase information can provide insights into the relative timing of different body parts' movements, which is often critical in gait analysis for applications like fall detection or diagnosing movement disorders. Another limitation is neglecting the spatial or range dimension. This is a critical drawback, especially in applications like assisted living or healthcare monitoring, where understanding the spatial context of human movement is essential.

Given limitations, there is a need to explore other radar domains for classification tasks. One of such domains is the Cadence Velocity Diagram (CVD), easily derived from spectrograms through a simple Fourier Transform (FT). CVDs can reveal crucial activity information – cadence frequency, which denotes average walking speed and stride rate, providing valuable insights particularly in the realm of radar-based gait analysis. In [166] extracts 13 numerical features from Doppler spectrogram and CVD profile [140] to compare the performance between sonar system and radar system. In [167], A. Seifert investigates more features including physical features from CVD profile such as cadence frequencies. Moreover, a sum-of-harmonics (SOH)-based model is proposed to estimate the fundamental Doppler frequency of the gait and the number of harmonics. The final accuracy of the combined feature set achieves 93.8%, which outperforms CVD-based feature set by 7.2%.

The range-time domain [159], [160], also known as the range profile, has recently gained significant attention due to its more efficient processing requirements. This domain bypasses the need for time-frequency analysis, essential for generating Doppler spectrograms. Range profiles provide snapshots of target reflections at specific times, offering crucial data on the distance between the radar and its targets. They are particularly useful for observing dynamic changes in moving targets over

time. The simplicity of range profiles, compared to other domains, makes them well-suited for real-time activity recognition applications where swift and efficient processing is paramount.

The range-Doppler domain [160], [161], [162] presents a comprehensive approach for activity recognition by combining two crucial radar measurements. The range-Doppler domain enriches feature extraction by detailing a target speed and motion alongside its distance, offering a deeper insight into the activities in progress. Merging these measurements can potentially boost the efficacy of activity recognition models, ensuring more precise and dependable identification of intricate activities. This domain produces a range-Doppler map that not only shows the target distance but also captures its relative velocity (Doppler effect), providing a more detailed perspective of the movements and dynamics of the target.

In [70], the author proposes a cyclostationarity-based approach for feature extraction, with activity classification as it detects micro-Doppler is made starting from complex-valued cyclostationary statistical functions of the reflected radar signal. The human activity can be recognized with up to 95.4% by the combination of real and imaginary part. The diversity of data types enhances our ability to differentiate activities based on distance, power spectrum, Doppler, Cadence etc., providing greater discernment, particularly in scenarios with minimal differences between classes. Utilizing these varied data types enables the creation of more refined and precise activity recognition models, adept at distinguishing between closely related activities.

3.3 Gait Analysis

Gait analysis has emerged as a critical tool in the realm of assisted living, offering valuable insights into an individual mobility, stability, and overall health [168]. It serves as a multifaceted tool in assisted living, offering both healthcare and identification solutions. Its ability to provide detailed, actionable data makes it indispensable for ensuring the well-being and safety of residents in these settings. This analytical method involves the systematic study of human motion, employing various techniques ranging from simple observational assessments to more complex radar and sensor-based evaluations. The primary measures associated with gait analysis include spatiotemporal parameters such as stride length, step width, and walking speed, as well as kinematic and kinetic variables like joint angles and ground reaction forces. These measures provide a comprehensive understanding of an individual's walking pattern, thereby serving as reliable indicators of their physical condition [168].

In the context of healthcare, gait analysis plays a pivotal role, particularly among older adults and individuals with mobility impairments [168], [169]. It is instrumental in assessing functional mobility limitations and evaluating the risk of falls, which are significant concerns in assisted living settings [169]. For instance, spatiotemporal gait measures have been found to correlate strongly with established physical performance assessments, making them reliable predictors of fall risks. This is especially crucial for residents with cognitive impairments such as dementia, where traditional assessment methods may not be as effective [168]. For identification purposes [170], gait analysis offers a unique, non-intrusive means of recognizing individuals. Advanced radar systems and wearable sensors can capture intricate details of human gait, which can then be analysed using machine learning algorithms for accurate identification. This is particularly useful in assisted living environments where personalized care is essential, and accurate identification can aid in customizing rehabilitation plans and monitoring routines [170].

Gait analysis is the study of human motion, often focused on locomotion, particularly walking, and running. In the context of radar, gait analysis can provide a non-invasive method for identifying and assessing human activities, health status, and even individual identity, based on the unique micro-Doppler signatures [171] produced by their movements. The use of radar technology in gait analysis has gained considerable attention due to several advantages. Firstly, radar sensors can operate effectively in various environmental conditions, such as darkness or poor lighting, and can even penetrate certain materials, allowing through-wall monitoring. Secondly, radar systems are less intrusive, not requiring wearable sensors or capturing visually identifiable data, thereby preserving privacy.

A few key studies in this area illustrate the potential of radar for gait analysis. In 2014, Wang et. al. in [172], one of the most comprehensive experiments regarding gait monitoring using radar for the elderly. In this work, they highlighted the current lab-based methods and expressed the difference between these locations and the home environment. In this experiment, they generated gait parameters from radar spectrogram signatures and classified four types of gaits from 13 participants.

More recent studies have integrated machine learning techniques to enhance the analysis of radar-derived gait data. For example, an LSTM [173] model was used to classify the walking gait of small groups of people vs individual persons in an outdoor scenario. As for experimental improvements, [174] who validated mD gait signatures as precursors of cognitive ability using data from 74 people. Similarly, in 2020, Le et

al. used an autoencoder model to classify gait patterns from radar signals, with spectrograms and scalograms, demonstrating recognition of different walking conditions [175].

Despite these advancements, similar to HAR, radar-based gait analysis also faces a set of challenges. One of these challenges lies in the heavy reliance on the spectrogram representation of radar data. The spectrogram, which visualizes the frequencies present in a signal as it changes over time, is incredibly useful for identifying periodic patterns, such as the cyclical movement of limbs during walking or running. However, the spectrogram as a standalone representation may not fully exploit the richness of radar data. It is also worth noting that while spectrograms are effective in capturing the frequency-time characteristics of a signal, they may not adequately represent other potentially useful features. For instance, phase information, which can be crucial for understanding the micro-motions involved in gait, is often lost in the spectrogram representation. Additionally, the spectrogram is sensitive to noise and may require preprocessing steps like windowing and overlapping, which could introduce artifacts or distortions in the data. Moreover, the spectrogram-based approach may not be as effective in capturing dynamic changes in gait, such as sudden stops or changes in direction, which are crucial for applications like fall detection or abnormality identification in assisted living scenarios. These limitations suggest the need for more advanced feature extraction techniques that can fully exploit the multi-dimensional nature of radar data for gait analysis.

3.4 Summary

HAR has emerged as a compelling area of research in the domains of human-computer interaction and smart surveillance, with radar technology gaining increasing attention for its unique advantages such as environment-insensitivity and enhanced privacy protection [28]. Deep learning techniques have further propelled this field by automatically extracting deep features, thereby achieving impressive classification performance [79]. Various state-of-the-art deep learning models, each with unique characteristics for identifying human activities, are being increasingly combined to capture features more effectively. In terms of radar systems employed for HAR, Doppler radar specializes in obtaining Doppler information, FMCW radar provides both range and Doppler data. The radar echoes are categorized into 1D, 2D, and 3D forms, each with their own set of challenges and opportunities when applying deep learning techniques. Particularly, 2D radar echoes [143], [148], [154], [176],

especially time-Doppler maps, are most commonly used due to their intuitiveness and rich information, whereas 3D echoes [177], although information-rich, pose greater processing challenges. The simpler 1D echoes [101] present an untapped potential for feature extraction. Convolutional Neural Networks (CNNs) have also been adapted for radar based HAR, with architectural choices like ReLU [114] and ResNet [114] addressing issues like the exploding gradient problem. Despite the current lag in the adoption of radar for HAR compared to vision-based technologies, the field holds significant promise owing to the synergistic advancements in radar technology and deep learning algorithms.

The field of Human Activity Recognition (HAR) using radar and deep learning has undergone several noteworthy advancements. One of the most significant changes is the evolution of deep learning algorithms, with newer architectures like Transformer models [178] and attention mechanisms being adapted for HAR tasks. This has been complemented by the use of advanced data augmentation techniques, such as Generative Adversarial Networks (GANs) [179], which are particularly useful for handling unbalanced datasets. The push towards real-time HAR has also gained momentum, thanks in part to the advent of edge computing and algorithmic efficiencies. Additionally, there is a growing trend towards the development of multi-modal HAR systems that integrate radar data with other types of sensory data, such as visual or inertial information, for a more robust recognition system. As these technologies become more pervasive in both public and private spaces, there has been an increased focus on addressing data privacy and ethical considerations.

The literature reviewed in this section is summarized in Table 5 and compared in terms of sensing approach, Domain used of the experiment (Spectrogram, range profile and etc), number of classes to distinguish, classifier and classification performance.

Table 5 Summary of previous work.

Ref	Year	Sensing Approach	Classification Algorithm	Classes	Radar Data Domain	Performance Metrics	Best Performance
[144]	2009	CW radar	SVM	7	Doppler-Time maps	Accuracy	92.8% for HAR
[172]	2014	Pulse Doppler Radar	Velocity and Step time estimation	13	Doppler-Time maps	Gait Assessment	No
[148]	2015	Simulated radar (from Kinect)	KNN	4	Doppler-Time maps	Accuracy	90.2%
[146]	2016	CW radar	Multiple Classifiers	3	SVD	Accuracy	99.6%

[155]	2016	Doppler Radar	DCNN	7	Doppler-Time maps	Accuracy	90.3%
[22]	2017	CW Radar	SVM	5	Doppler-Time maps and CVD	Accuracy	73%
[160]	2017	FMCW Radar	SAE	4	Doppler-Time maps, Range Maps.	Accuracy	96.4%
[173]	2017	CW Radar	LSTM	3	Doppler-Time maps	Accuracy	89.1%
[150]	2018	FMCW Radar and Simulated Data	Multiple Classifiers	5	Doppler-Time maps	Accuracy	88.74
[154]	2018	Multistatic Pulsed Radar	DCNN	2	Doppler-Time maps	Accuracy	97.42% for PR, 99.63% for GAR
[152]	2018	CW Radar	GRU	6	Doppler-Time maps	Accuracy	92.7%
[159]	2018	FMCW Radar	LSTM	2	Raw Radar Data, Range Maps	Accuracy	99.56%
[175]	2018	CW Radar	CAE	3	Doppler-Time maps, CVD	Accuracy	96.2%
[15]	2019	CW radar	CAE, DNN, SVM	12	Doppler-Time maps	Accuracy	95%
[151]	2019	CW Radar	LSTM	6	Doppler-Time maps	Accuracy	91.8%
[167]	2019	CW Radar	KNN	5	Doppler-Time maps and CVD	Accuracy	93.8%
[161]	2019	FMCW Radar	ANN	8	Doppler-Time maps, Range Maps.	Accuracy	99.21%
[162]	2019	FMCW Radar	CNN	5	Doppler-Time maps, Range Maps.	Accuracy	97.2%
[174]	2019	CW Radar	SVM	74	Doppler-Time maps	Accuracy	94.6%
[180]	2019	FMCW Radar	KNN	6	Doppler-Time maps	Accuracy	95.5%
[29]	2020	FMCW radar	SVM, KNN	6	Doppler-Time maps, CVD	Accuracy	95.4% for HAR, 91.4% for PR
[149]	2020	UWB radar	KNN, SVM, DNN	6	Doppler-Time maps	Accuracy	98%
[156]	2020	FMCW Radar	LSTM+CNN	7	Doppler-Time maps	Precision, Recall, F1-Score, Accuracy	Precision: 98% Recall: 98% F1-score: 98% Acc = 98.28%
[100]	2020	FMCW Radar	Bi-LSTM	6	Doppler-Time maps	Accuracy	96%
[164]	2020	Simulated Radar Data	CNN, SVM	9	Doppler-Range-	Accuracy	91.57%

					Time 3D cube		
[143]	2020	FMCW Radar	CNN (PointNet), LSTM	5	3-D cloud point	Accuracy	90%
[181]	2021	FMCW Radar	CFFN (CNN +MLP)	6	Range-Doppler Surface, Phase	Accuracy	94%
[182]	2021	FMCW Radar	MB-GAN (Multi-Branch GAN)	5	Doppler-Time maps	Accuracy	89.83%
[142]	2021	FMCW Radar	CNN+LSTM	6	Range-Doppler-Angle map	Accuracy Recall	Accuracy:96.2% Recall: 96%
[70]	2021	FMCW Radar	SVM, KNN, RF	6	Cyclostationarity	Accuracy	95.4%
[114]	2023	CW radar	DCNN	7	Doppler-Time maps	Accuracy	97.6% for PR, 90.3% for HAR
[157]	2023	FMCW Radar	Multi-Domain feature attention fusion network	6	Range Time and Doppler Time maps	Accuracy	97.58%
[158]	2023	FMCW radar	CNN-LSTM-Attention Hybrid Multi-Network	5	Sequence of Doppler Time maps	Accuracy	96.9%

Overall, Chapter 3 provides a comprehensive overview and review of the application of machine learning and artificial intelligence techniques in AAL using radar technology. It highlights the significance of AI in enhancing AAL systems and showcases the potential of radar-based features, gait analysis, activity monitoring, and multimodal radar sensing in improving the quality of life for individuals in need of assisted living support.

In addressing the radar technology for healthcare applications, particularly in activity recognition, this chapter identifies some critical research questions that highlight the gaps in current methodologies.

The first question delves into the potential of radar technology, especially micro-Doppler and beyond, in healthcare contexts 'Why use radar, and how can we leverage its capabilities, including different radar representations, for enhanced healthcare applications, specifically in activity recognition?' This question underscores the need to explore and justify the use of radar over other sensing technologies. It also opens up avenues for investigating how advanced radar techniques like micro-Doppler can be optimized for more accurate and efficient activity recognition.

The second question broadens the scope of investigation beyond the spectrogram capabilities of radar. It inquires 'How can it be improved to optimize the classification accuracy in the context of assisted living?' This question goes beyond the technical aspects of radar sensing to consider the optimal signal processing and types of radar representations, the nature of data to be collected, and the balance between data collection and privacy concerns.

In short there is a new question formed based on the literature 'How can the integration of machine learning-assisted techniques in radar-based HAR systems be optimized to significantly enhance classification accuracy across various radar representations?' This question aims to investigate the synergy between advanced signal processing methods and machine learning algorithms in the context of radar based HAR. It focuses on exploring the potential of novel pre-processing methods, such as adaptive thresholding, and their impact on the accuracy and efficiency of HAR systems. Additionally, it seeks to understand the effectiveness of integrating these methods with ML techniques, particularly in terms of performance optimization and computational resource management. This research gap is crucial for advancing the field of HAR, especially in applications where accuracy, efficiency, and resource constraints are of paramount.

4. Individual Human Activity Recognition

Cognitive processes in intelligent beings are fundamentally anchored in the discernment of distinctive characteristics observed in objects or actions. This principle is evident in the natural world, where survival often hinges on the ability to perceive specific signals, such as using olfactory cues for tracking prey or adapting to environmental light conditions for optimal hunting. At its essence, the identification of these unique traits underpins a range of complex behaviours. Similarly, in the realm of computational cognition, the challenge of object identification is approached through the meticulous extraction and analysis of features. However, unlike the tangible biological features observed in nature, features in radar signals are abstract, necessitating sophisticated methods for their derivation and processing.

This chapter explores the initial application in employing handcrafted features for human activity recognition using radar sensors, with a particular emphasis on examining the impact of diverse features. It also addresses the influence of demographic variables—such as age, gender, and physiology—on the effectiveness of activity recognition. Initially, the chapter evaluates the performance of handcrafted features across a diverse dataset. Following this, the focus shifts to the exploration of feature selection and fusion algorithms, aiming to refine and enhance the human activity recognition process.

In the current landscape, where end-to-end classification models are increasingly dominant, the initial research efforts concentrated on extracting salient features from radar data. This extraction is not limited to spectrograms but extended to other data domains as well. These features are broadly categorized into three groups: physical features, image-based features, and transform-based features. The chapter discusses their application in classifying various biological movements, highlighting their relevance and utility in the broader context of radar-based human activity recognition.

This chapter is based on the publication [38] and the chapter is organized as follows: beginning with Section 4.1, which establishes the motivation behind the study. This section highlights the significance of human activity recognition using radar technology. Following this, Section 4.2 details the experimental setup, including the radar system configuration, activity selection, and participant demographics, offering insight into the data collection process. Section 4.3 delves into the handcrafted features utilized in the study, discussing their selection, extraction, and relevance to the task at hand. The results of the study are presented in Section 4.4, where the classification outcomes are analysed, showcasing the performance of various features

and the overall system efficacy. The chapter concludes with Section 4.5, summarizing the key findings and implications of the research.

4.1 Motivation

The exploration of HAR using radar technology has predominantly cantered around mD signatures, as evidenced by the majority of existing research in this domain. However, the emergence of methods utilizing other data domains, such as spectrograms, signals a shift towards a more diverse and comprehensive approach in radar-based cognitive selection. This shift is not merely confined to the extraction of features but extends to the consideration of various data domains, thereby enriching the repertoire of techniques available for HAR.

The investigation of these multiple domains opens up new possibilities for cognitive selection in radar-based HAR. By exploring different data domains, researchers can uncover a broader array of techniques and strategies for feature extraction and selection. This expansion is not just theoretical; the practical implications are significant, as the insights gained from these investigations can be integrated into other algorithms, enhancing their effectiveness and efficiency. One such algorithm that stands to benefit from this expanded approach is hierarchical classification [79]. Hierarchical classification offers a nuanced method to tackle HAR by segmenting activities into several subgroups. Each subgroup can then be analysed using a tailored combination of data domains and features, specific to the characteristics of that subgroup. This methodological refinement allows for a more targeted and accurate classification of activities.

In addition to the exploration of domains, in supervised classification, the standard method involves initially learning predefined or handcrafted features from a training dataset, followed by applying the same feature computation on test data for classification. Conversely, data-driven feature learning methods [74] adapt the feature extraction process using insights from the training data analysis. Figure 21 presents a block diagram outlining the key steps in classification using data-driven learning.

The division of activities into distinct groups also paves the way for the application of domain knowledge and information fusion techniques. By aligning specific radar data domains with the activities, they are best suited to recognize, the overall accuracy of HAR can be significantly improved. This alignment is not a mere matching exercise

but involves a deep understanding of the nature of the activities and the characteristics of the radar data domains.
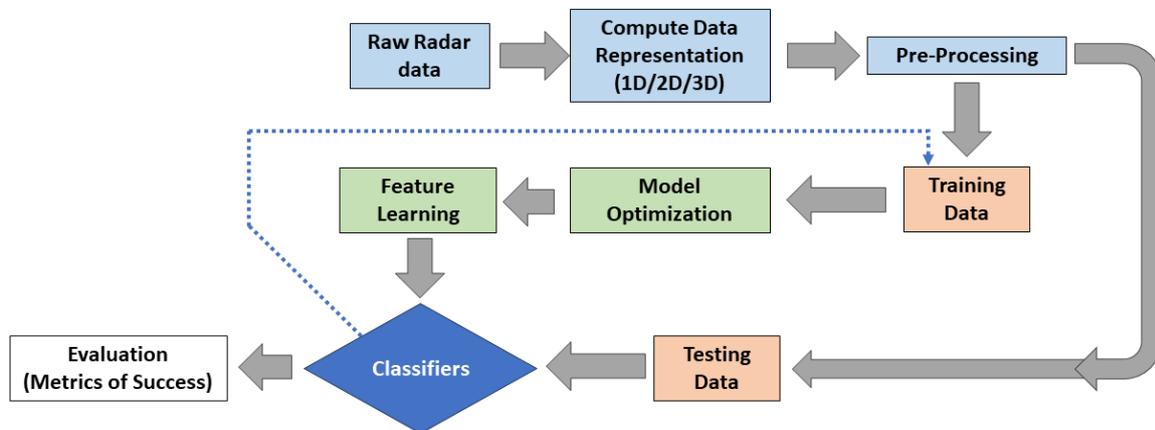
Figure 21 *Flow chart of data-driven classification approach*

In summary, the motivation for this research lies in expanding the scope of radar-based HAR beyond the traditional focus on mD signatures. By exploring a variety of data domains and integrating these insights into advanced algorithms like hierarchical classification, this research aims to enhance the accuracy and effectiveness of HAR systems. This approach not only leverages the strengths of different radar data domains but also aligns them with specific activity groups, thereby optimizing the recognition process and contributing to the advancement of radar-based human activity recognition.

4.2   Experimental Setup

This section is dedicated to the variables integral to the experiments conducted. It encompasses a detailed description of the sensors utilized, encompassing their specifications and operational capabilities. Additionally, the section provides a comprehensive view of the participant demographics, including the total number of participants, their gender distribution, and any other relevant characteristics that may influence the outcomes of the experiments.

4.2.1 Radar Sensor

In this section, an off-the-shell FMCW radar (Ancortek 580B) operating at 5.8 GHz was used to collect the human motion data. The FMCW radar has 400 MHz instantaneous bandwidth and the Pulse Repetition Frequency (PRF) equates to 1 kHz.

The detailed parameters are also shown in Table 6. The radar transmitted power is approximately 18 dBm, whereas the maximal power of the transmitting and receiving Yagi antenna is 100 mW, with a gain equal to approximately 17 dB. Additionally, both transmitting and receiving antennas are vertically polarized with beam width of 24 degree in azimuth and elevation.

Table 6 The Radar parameters

Radar Specification	The details
Center Frequency	5.8 GHz
Bandwidth	400 MHz
Pulse Repetition Frequency	1 KHz
Coherent Processing Interval	1 ms

4.2.2 Environmental, Participant and Activity Setup

The influence of varying locations on a dataset can introduce a layer of complexity that is critical to consider in the context of radar-based activity recognition. Different locations may present unique environmental variables such as background noise, signal interference, or multipath effects, which can significantly impact the quality and reliability of radar data. For instance, indoor settings may involve reflections off walls and furniture. These location-specific factors can introduce variability in the extracted features, thereby affecting the performance of ML models trained on such data. Consequently, it becomes imperative to either design algorithms robust enough to handle this variability or to employ location-specific calibration to ensure accurate activity recognition across diverse settings. In this experiment, the data was collected in various indoor locations. Generally, there are six distinct indoor rooms, with different layouts and decorations. They are the University of Glasgow laboratory room at the James Watt South building, the common room at the School of Engineering at the University of Glasgow, Glasgow NG Homes Room 1-3, and Age UK West Cumbria Room, as shown in Figure 22.

(a) and (c) presents a minimalist environment, characterized by its bareness. Enclosed by four solid walls, the room is devoid of furnishings and equipment, offering a blank environment for experiment setup. The strength of the room is its simplicity of data collection area, providing a controlled environment for experiment. In contrast, room (b) for students is a vibrant and dynamic space, designed to cater to the needs of a diverse students. It is furnished with multiple desks and chairs, arranged to facilitate both individual study and group collaboration, with cupboards placed around the room. Also, it is the largest room in this experiment. (d)-(f) are the normal bedrooms,

offering a personal retreat. It features a bed as the centrepiece, accompanied by wardrobes, desks and chairs to accommodate work and life. (g)-(h) are activity rooms particularly designed for the elders. The activity rooms are dynamic spaces filled with resources for creative and physical activities tailored to the interests and abilities of older adults. It includes craft tables, chairs, sofa, exercise equipment, and areas for games and hobbies, encouraging residents to stay active and pursue their interests in a supportive environment.
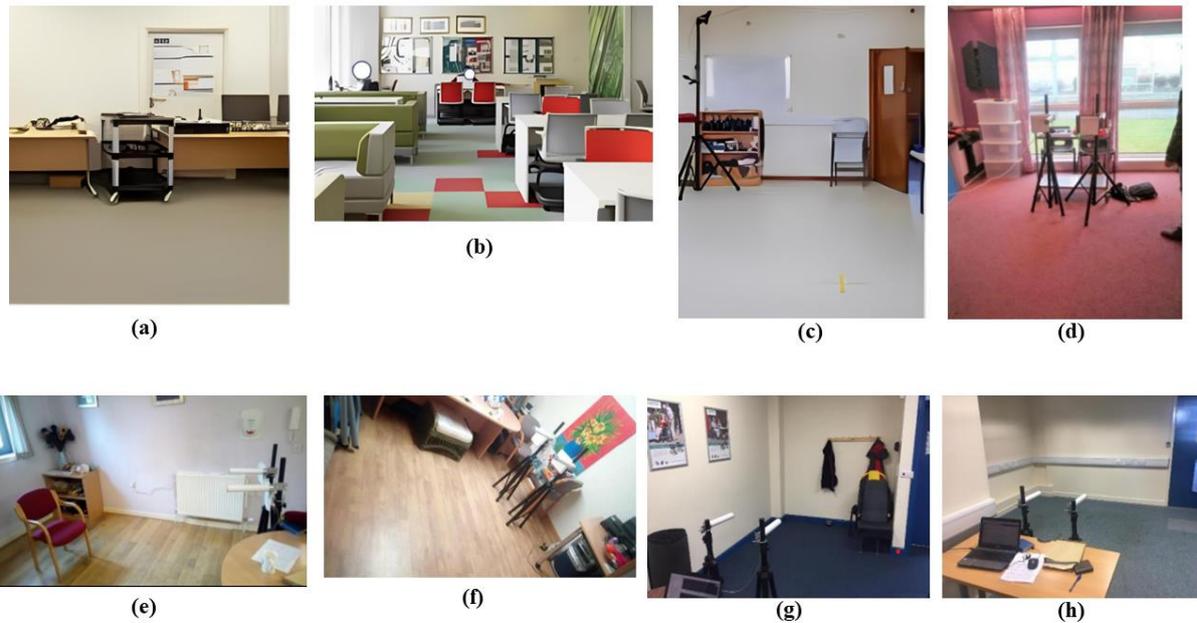
Figure 22 *The environment of experiment. (a), (c) the Laboratory room, (b) the Common room, (d)-(f) NG home room 1-3, (g)-(h) Age UK West Cumbria room*

Table 7 List of the activities

No.	Activity Description	# of samples	Duration (s)
A1	Walking back and forth	312	10s
A2	Sitting down on a chair	312	5s
A3	Standing up from a chair	311	5s
A4	Picking up an object from the ground	311	5s
A5	Drinking water from a glass	310	5s
A6	Fall	198	5s
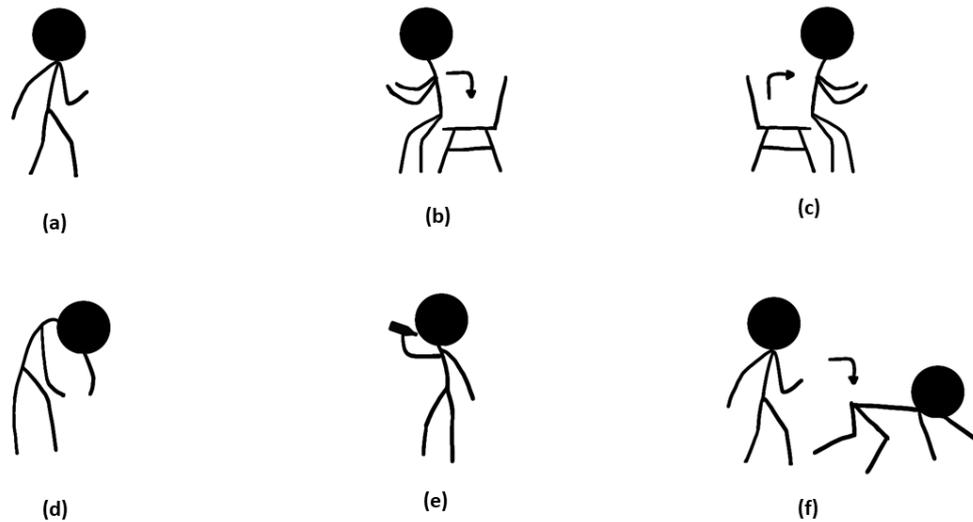
Figure 23 *A pictorial list of activities: these six activities were performed. (a) walking, (b) sitting down, (c) standing up, (d) picking up an object, (e) drinking, (f) fall.*

The participants are aged 21-98 with diversity in terms of gender (27 females, 45 males), height (1.52 to 1.88 m), body shape and dominant hand. The recorded activities are illustrated in the right part of Figure 23 and summarized in Table 7, where some of the activities are designed to be similar in pairs (e.g., 'A2' and 'A3', 'A4' and 'A5'). Those similar activities are deliberately added to create more classification challenges. For each participant, three repetitions of each activity are taken, and this generates a database containing 1754 samples.

In this experiment, the selection of activities - walking, sitting down, standing up, bending to pick up, drinking water, and falling - was made with careful consideration to encompass a range of common, everyday movements that are particularly relevant in the context of assisted living and healthcare monitoring.

The consent process was meticulously designed to adhere to the ethical guidelines stipulated by the Ethics Board of University of Glasgow. Participants were informed in detail about the study's objectives, the nature of the data collection, their rights as participants, including the right to withdraw at any point without penalty, and the measures taken to ensure their privacy and data confidentiality. Consent was obtained through a two-step process: initially, verbal consent was secured during the recruitment phase, followed by written consent at the beginning of the data collection session. This process ensured that participants were fully informed and voluntarily agreed to participate in the study.

The recruitment of older adults was conducted through a multi-faceted approach to ensure a diverse and representative sample. Recruitment channels included senior living communities and health care facilities, as well as social media platforms. To encourage participation, the study was presented not only as an opportunity to contribute to meaningful research but also as a chance to engage in activities promoting physical and cognitive health. Participants were compensated for their time to acknowledge their contribution and to incentivize participation.

The selection of activities for this study was strategically chosen to encompass a broad spectrum of movements that are integral to daily living, particularly focusing on those that present varying degrees of complexity and relevance to the target demographic of older adults. Walking was selected due to its fundamental role in human mobility and the significant variability in gait patterns it exhibits. This variability is essential for assessing an individual's mobility and balance capabilities, which are critical components of functional health. The universal nature of walking makes it a cornerstone activity for analysis in this context. Sitting down and standing up were included as they represent transitional movements that are pivotal in daily life but often pose challenges for the elderly or individuals with physical impairments. These activities serve as key indicators of an individual functional health, providing insights into their ability to perform basic movements that are essential for independence. Bending is incorporated into the study due to its complexity, involving multiple joints and muscle groups. This activity offers valuable information regarding an individual's flexibility and range of motion, which are indicative of their overall physical condition and ability to engage in a variety of daily tasks. Drinking water was chosen as a representative fine motor activity to evaluate the capability of system to recognize subtle yet essential activities. The ability to perform fine motor tasks is crucial for maintaining independence and quality of life, making the inclusion of such activities important for a comprehensive assessment. Lastly, the inclusion of falling addresses a major concern in the care of the elderly, given its potential for causing serious injury. The ability of a system to accurately detect falls is paramount in assisted living environments, where timely intervention can significantly mitigate the risk of injury.

In summary, the chosen activities provide a comprehensive framework for evaluating the system's ability to accurately recognize and differentiate between a wide range of human movements. This selection is aimed at ensuring the system's applicability in real-world scenarios, particularly in supporting the autonomy and safety of older adults in assisted living environments. Together, these activities form a comprehensive set that not only covers a wide spectrum of human motion but also

addresses key aspects of daily life and health monitoring, making them ideal for evaluating the effectiveness of radar-based human activity recognition systems in real-world scenarios.

4.3 Handcraft Features Extraction

For the conventional classifier, prior to the classification, handcrafted features are extracted from the raw data to characterize each class in specific manner. In general, feature extraction reduces the dimension of classifier inputs. These are often identifiable with the human eye and have origins in the literature and communities.

4.3.1 Physical Features of Spectrogram

The physical features of spectrograms are extracted according to the physical properties of mD signatures. Doppler centroid (or Doppler centre of mass) and bandwidth are the most salient features, where the centroid is the peak amplitude of the spectrogram and it represents the translational speed of the human subject, the bandwidth indicates the relative velocities of limbs around the centroid and the spread is largely driven with the activity being performed and the aspect angle of the subject. The centroid and bandwidth for each Doppler slice in a spectrogram are expressed in Eq. 4.1 and Eq. 4.2, respectively [47].

$$f_c(i) = \frac{\sum f(j) S(i,j)}{\sum S(i,j)} \tag{4.1}$$

$$B_c(i) = \sqrt{\frac{\sum_j \big(f(j) - f_c(i)\big)^2 S(i,j)}{\sum_j S(i,j)}} \tag{4.2}$$

where $f_c(i)$ and $B_c(i)$ denotes the Doppler centroid and bandwidth of the i-th time bin, $f(j)$ refers to the Doppler frequency of j-th Doppler bin, $S(i,j)$ is the matrix component of the spectrogram at the i-th time bin and j-th Doppler bin.

The energy curve of the Doppler for each time bin of the spectrogram calculates the energy within a given frequency band. The equation of energy curve $C(i)$ is shown in Eq. 4.3.

$$C(i) = \sum_{j=f_1}^{f_2} |S(i,j)| + \sum_{j=-f_2}^{-f_1} |S(i,j)| \tag{4.3}$$

where f_1 and f_2 are the frequency band of spectrogram. This band can be selected and should be smaller than the range between minimum and maximum Doppler frequency to be of any use, however not as small as a Doppler bin pair.

4.3.2 Transform-based Features [146]

Transform-based features perform mathematical transformations, where these features utilize the transformed domain to easily extract and assess properties such as singular value decomposition, and repetition and frequency of motions within the original spectrogram signature.

4.3.2.1 Singular Value Decomposition

Singular Value Decomposition (SVD) [21] is one of the effective approaches to analyse the Doppler spectrogram to search for useful hidden information, and the SVD of complex matrix M can be formulated as Eq. (4.4):

$$M = U\Sigma V^T \qquad (4.4)$$

where M is the input matrix, Σ is the singular value matrix, U and V are orthogonal basis matrices. Consider a matrix M representing a spectrogram, with dimension $A \times B$. The dimension of these matrices are as follows: U is $A \times A$, V is $B \times B$, and Σ aligns with the dimension of M, i.e., $A \times B$. Each singular vector matrix is normalized, and the corresponding singular value in Σ represents the energy weight of these vectors.

The application of SVD in spectrogram analysis offers several advantages over traditional physical feature extraction methods. Primarily, SVD facilitates the automatic derivation of dominant components within the spectrogram. This is because the singular vectors associated with smaller singular values typically correspond to less significant components or noise, thereby enabling a more focused analysis of the crucial elements in the spectrogram. Furthermore, SVD offers a means to reduce the dimensionality of the input data. By selecting a subset of singular vectors, the entire spectrogram M can be effectively represented in a more compact form.

In addition, a spectrogram M with dimensions $A \times B$, where A is determined by the window length of the STFT and B by the time bin, the left singular vector matrix U can be regarded as the frequency information of the spectrogram. Conversely, the right singular vector matrix V contains the time distribution information. This dimensional reduction and feature extraction capability of SVD make it a valuable

technique in signal processing, particularly for applications requiring efficient and insightful analysis of spectrogram data.

4.3.2.2 Features of Cadence Velocity Diagram [183]

The Fourier transform, when applied to the spectrogram across time for each Doppler bin, reveals the repetitive nature or cadence of the observed activity. This process leads to the formation of the CVD, a tool that uncovers the frequencies at which various velocities recur. The CVD is instrumental in providing insights into the shape, size, and frequency of the patterns formed in the spectrogram by the moving parts of the target. In the CVD, the features extracted are peak-based, focusing on the identification and analysis of significant peaks.

One of the key features derived from the CVD is the step repetition frequency. This is obtained by identifying the local maxima within the CVD and arranging them based on their prominence. The step repetition frequency is specifically determined by locating the cadence frequency index corresponding to the second peak (the first peak typically being at 0 Hz) in the average CVD profile, as depicted in Figure 24. Concurrently, the amplitude of the Doppler bins corresponding to this cadence frequency index can also be used as features. The most significant Doppler frequency is identified as the frequency corresponding to the maximum frequency of this peak in the CVD. Through these methods, the CVD serves as a critical tool in extracting meaningful and actionable insights from radar spectrogram data, particularly in the context of analysing repetitive motions such as walking.
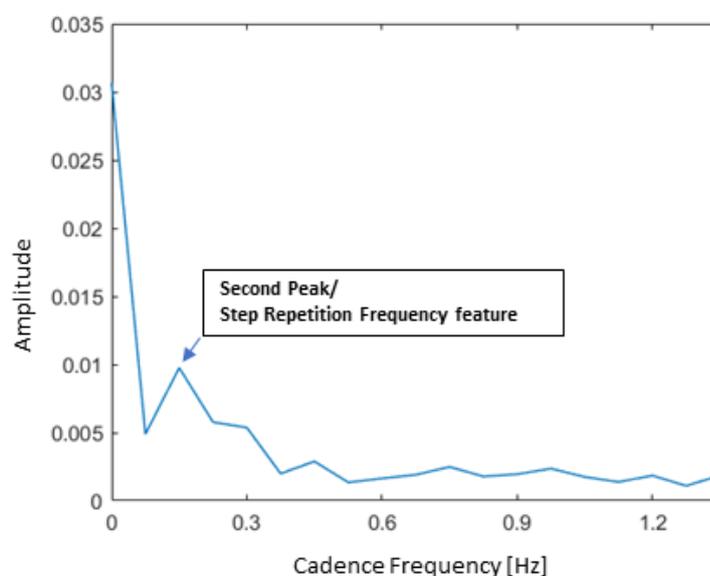
4.3.3 Features from Classic Image Recognition [184]

Treating the spectrogram as a grayscale image opens up a myriad of possibilities for feature extraction using classical image processing techniques. In this approach, the spectrogram, fundamentally a bitmap array, is interpreted in terms of varying shades of grey, each representing different dynamic ranges. This grayscale representation allows for a nuanced interpretation of the radar signals, where the intensities of movements, such as those from the torso and limbs, are translated into distinct segments on the grey spectrum.

By adopting this perspective, the amplitude of movements from different body parts can be effectively quantified and mapped using a histogram. This histogram essentially serves as a tool to capture and represent the distribution of movement intensities within the spectrogram. It provides a visual and quantitative means to analyse the range and Doppler frequency of the radar signal amplitudes, correlating them to specific physical movements. The entropy of a histogram of an image equates to the intensity of the signature. It could also be defined as the average information within an image. It is expressed as Eq. 4.5:

$$H = -\sum p(x) \log p(x) \tag{4.5}$$

where $p(x)$ is probability mass function of the histogram of the image. Complex motions such as picking up an object (which involve many joints and rotations about an axis) will be expected to return a higher value of entropy relative to activities such as standing.

Skewness is also a parameter that could be derived from histograms of spectrogram image. This statistical measure, along with a series of moments such as mean and standard deviation, plays a crucial role in exploring the difference of images of activities. For images have a standard form given as Eq. 4.6:

$$\gamma_n = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{x_i - \mu}{\sigma} \right]^n \tag{4.6}$$

where μ is the mean and σ is the standard deviation of the histogram distribution.

By leveraging these standard forms of statistical analysis, researchers can extract meaningful insights from the complex interplay of radar signals and human activities, enhancing the accuracy and depth of their interpretations.

4.4 Results

In Section 4.2, we introduced a range of features to our study. Building upon this foundation, the subsequent analysis delves into the classification results, structured in a progressive manner. Initially, we explore classification within a single domain, providing a baseline understanding of how each domain independently contributes to the recognition task. This is followed by an examination of feature selection within a single domain, where we identify and utilize the most significant features to enhance classification performance. Then, our analysis is the fusion of information across different domains. This advanced approach aims to leverage the complementary strengths of each domain, thereby achieving a more robust and accurate classification system. Throughout this process, we employ one-vs-one strategies to thoroughly evaluate the classification performance. This approach involves creating and evaluating binary classifiers for each pair of activities, thereby ensuring a comprehensive and detailed analysis of how well each activity is distinguished from every other.

In addition to the techniques applied in this study, we also extend the analysis to encompass a comparative evaluation of activity patterns between adults and the elderly. This comparison is crucial, as it sheds light on the variations in movement and gait that typically occur with age. By analysing and contrasting the activity data from these two distinct age groups, we aim to uncover nuanced differences in their motion characteristics. The results are presented in terms of average performance metrics, providing a comprehensive overview of the classification efficacy across various scenarios. This structured approach not only highlights the individual contributions of each domain but also showcases the synergistic potential when these domains are intelligently combined.

4.4.1 Classification Results

At this stage, an SVM classifier with a quadratic kernel and a weighted KNN classifier with K = 10 were employed for the activity classification. To validate the performance of these classifiers, we employed a 10-fold cross-validation technique. This method involves partitioning the dataset into ten equal subsets, where each subset

is used once as a validation set while the remaining nine subsets are used for training. This process is repeated ten times, with each subset serving as the validation set once. This approach ensures that every data point is used for both training and validation, providing a comprehensive evaluation of the classifier's performance.

Additionally, to assess the robustness of the models, we calculated the standard deviation using each run accuracy and the final average accuracy. This step is crucial as it allows us to evaluate the stability and reliability of the classifiers under varying conditions. A lower standard deviation across runs indicates a more robust model, capable of delivering consistent performance regardless of minor variations in the training and validation sets.

Table 8 Comparison of classification accuracy of algorithms and domains

Accuracy (%)	mD	RT	CVD
SVM	80.3	64.2	**82.4**
KNN	75.2	61.3	**80.4**

Table 9 Comparison of standard deviation of algorithms and domains.

Standard Deviation	mD	RT	CVD
SVM	0.6534	3.4112	2.5813
KNN	2.2886	3.6831	2.4643

Table 8 shows preliminary results of the classification without feature selection. Note that the number of features for mD domain is 21, for RT domain is 21 and for CVD domain is 7. Generally, the SVM algorithm outperforms the KNN algorithm, and CVD obtains the best result overall. The result for mDs is worse than CVD by approximately 2 %. The classification performance of the RT domain is the worst, which is 18 % lower than CVD 18 % with SVM and 19 % with KNN.

Table 9 reveals that the SVM classifier generally shows lower or similar variability in its performance across different feature types compared to the KNN classifier. The Range-Time (RT) features seem to introduce the most variability in performance for both classifiers, which might be more sensitive to the data or that the classifiers struggle to consistently leverage them effectively. In contrast, the micro-Doppler signature with the SVM classifier stands out for its robustness, with relatively low

standard deviation of 0.6534, underscoring its stability and reliability across different scenarios within the domain of activity recognition.

4.4.2 Feature Selection

Achieving high accuracy in classification tasks necessitates the use of discriminative features that enhance the confidence and reliability of the results, as highlighted in [32], [185]. The effectiveness of a classifier is largely contingent on the quality and size of its input data. With the increase in sensory information, the feature space can expand significantly, both spatially and dimensionally, leading to potential redundancy due to the 'curse of dimensionality'. To counteract this, feature selection is a critical step, not only to manage input size but also to enhance model performance by eliminating redundant or irrelevant features.

Feature selection methodologies aim to streamline computational processes and enhance performance by removing features that are either redundant or carry negative/repetitive information. These methodologies can be broadly categorized into three types: filter-based, wrapper, and embedded methods.

1. Filter-based Methods: These methods employ statistical measures (such as entropy, correlation, and Euclidean distance) to rank features. Their primary advantage is that they can operate independently of any classifier, making them versatile and less computationally demanding.
2. Wrapper Methods: These methods involve using a specific classifier to assess different combinations of features. The subset that yields the highest classification accuracy is selected. While effective, wrapper methods can be computationally intensive, particularly when dealing with high-dimensional feature sets.
3. Embedded Methods: These methods integrate feature selection within the classification process. During training, the classifier algorithm calculates a weight for each feature, ranking them based on their importance.

Algorithm 2 Fisher Score

Initialization: C-the number of classes; N-the total number of samples in datasets, and N_i-the number of samples in feature i, S_{Bi}- Between class variance for each class i, S_W- Within class variance. x-the value of each sample

Input N, C, N_i

Set both S_B and S_W to 0;

Calculate the mean μ of the features across all samples

For each feature i **do**

> Calculate the mean μ_i of the feature
>
> $S_{Bi} = N_i \times (\mu_i - \mu)^2$
>
> $S_w = \sum (x - \mu)^2$
>
> Calculate Fisher Score for feature $i = \dfrac{S_{Bi}}{S_w}$

End for

Output Fisher Score for each feature i.

In this work, two filter-based methods, the Chi-Square [185] and Fisher score (F-score) [32], are investigated and implemented to select the optimal feature subsets from the feature pool. Chi-Square test is used in statistics for testing the independence of events. When it is used for feature selection, Chi-Square calculation indicates the dependence between the target and features, where higher Chi-Square values indicate more informative features. Chi-Square is widely used due to its ease of computation and robustness with respect to the independence of data. The Fisher score algorithm operates on the principle of optimizing feature separability. It ranks features based on their ability to minimize the distance between data points within the same class while simultaneously maximizing the distance between data points belonging to different classes. This approach ensures that the selected features are effective in distinguishing between various classes, enhancing the accuracy of classification tasks. The Fisher score is also a computationally simple algorithm, with fast processing speed and generally good performance.

A larger Fisher Score for a feature indicates that the feature has a stronger discriminative power. Specifically, it means that the feature effectively separates data into different classes. This is achieved by maximizing the difference between the means of the feature across different classes while minimizing the variance of the feature within each class.

Algorithm 3 Chi-Square

Initialization: S-the selected feature ranking map; N-the total number of samples; i- the number of features; Y- the labels of classes;

set the initial S is empty;

    for each feature F **do**

        Create a contingency table T for F against Y

        Calculate the Expected Frequency E for F in T.

        Calculate the Chi-Square Value $X^2 = \sum \frac{(\sigma - E)^2}{E}$ for each feature

    end for

    Rank the X^2 in S from highest to lowest

Output S

Table 10 Comparison of feature selection methods

Methods	No. of Features	No. of Features after	Computing time (s)
mD (SVM and FS)	21	13	41.8
mD (KNN and FS)	21	13	7.3
CVD (SVM and FS)	7	5	28.4
CVD (KNN and FS)	7	6	**1.7**
RT (SVM and FS)	21	18	126.4
RT (KNN and FS)	21	13	6.1
mD (SVM and CS)	21	15	56.5
mD (KNN and CS)	21	13	7.5
CVD (SVM and CS)	7	5	20.6
CVD (KNN and CS)	7	5	2.0
RT (SVM and CS)	21	17	97.2
RT (KNN and CS)	21	12	6.4

* Chi-Square: CS, Fisher Score: FS

Table 10 illustrates the results of the Fisher score and Chi-Square (CS). Fisher score method provides a limited improvement in the SVM algorithm, which is

approximately 1 %, whereas it enhances the performance of KNN algorithm by approximately 3 %. The accuracy of the classification results is boosted when the CS is implemented. For the SVM results, it generally improves the accuracy by approximately 4 %, while this enhancement on the classification performance for KNN is about 5 % for the mD domain. Feature selection decreases the number of features to improve the computing time and generally improves accuracy. FS shows that the optimisation occurs at 62-85% feature available. Meanwhile, the optimal results obtained by CS happened around 57-80 % available features. Hence, the CS algorithm outperforms the FS approaches with higher accuracy of classification. Figure 25 illustrates how the classification accuracy changes along with the increase of the number of features.

The purpose is to discuss and investigate the relationship between the domains and specific activities. Table 11 summarises the results. Picking up an object is the most easily misclassified activity, where the highest accuracy is 77.3 %. In the mD domain, most misclassifications are from A4 (picking up an object), which average accuracy is 63.6 %. In the CVD domain, the most confusing pair is picking up an object and drinking water (A5). The walking activity also has errors, while it performs well in the other two domains, with 100 % accuracy. However, the CVD domain has the best overall performance among the three domains, especially for its sensitivity to the sitting and standing, with high accuracies of 92.7 % and 89.0 %, respectively. In the RT domain, walking and falling can be well detected, with approximately 100 % and 87.1 % accuracy, whereas the rest of the activities are problematic.
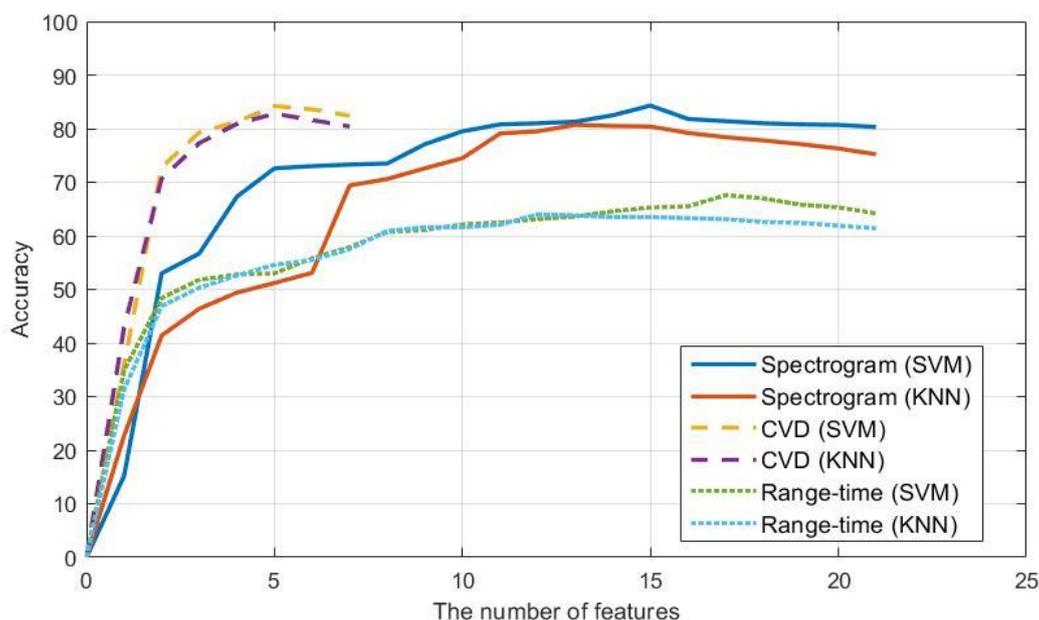
4.4.3 Feature Fusion

To further investigate the overall performance of classification, data fusion approaches will be used, and the fusion results will be compared. Data fusion is the process of integrating multiple data sources to produce information which can overcome limitation caused by any single data source. This can be achieved at many levels of abstraction, such as characteristics and symbols, and it is typically done at either signal, feature, or decision level [186]. In this experiment, the feature level fusion is used on three different data domains. Feature level fusion is aimed to generate a single, larger feature vector samples from different features. Feature selection methods can be used for removing redundant or incorrect features in the new feature space. For feature level fusion, the Chi-Square algorithm is used as feature selection method before feeding features to the classifiers due to its higher accuracy, and the SVM classifier only, as it was found to be the better performing classifier compared with KNN.

The results are shown in Table 11. The overall classification accuracy is increased. To the CVD domain, the accuracy of walking increases to 100 % from 86.5 % when the RT and mD features were applied to fusion. Compared with using RT and mD feature independently, the fusion with CVD also improves the accuracy of standing up (by 38 % and 11 %), sitting down (by 38 % and 5 %), picking up an object (by 20 % and 10 %) and drinking (by 14 % and 4 %) for both RT and mD, respectively. The fusion of CVD with the other two domains could cover the deficiencies of each other. However, when the fusion was applied to features of the mD and RT, it exacerbates the poor accuracy. This exacerbation might be caused by features with similar drawbacks.

Table 11 *Comparison of accuracy for individual activities and averaged across the activities using each data domain independently, with feature fusion, and with customised hierarchical classification.*

Accuracy (%)	A1	A2	A3	A4	A5	A6	Overall Performance
mD domain	100	82.2	88.4	63.6	79.7	91.9	84.3
CVD domain	86.5	92.7	89.0	63.7	80.7	92.8	84.2
RT domain	100	53.1	51.4	53.3	60.9	87.1	67.6
mD + CVD	100	92.7	92.9	74.2	83.8	97.1	90.1
mD + RT	100	51.5	28.6	42.7	58.1	86.2	61.2
RT + CVD	100	90.9	89.3	73.5	75.2	92.8	87.0
mD + RT + CVD	100	94.2	91.7	77.3	84.9	97.1	90.9

** The accuracy of three domains is recorded from SVM classifiers with Chi-Square algorithm, due to their better performance compared to others.

4.4.4 Hierarchical Structure

Based on the previous results, a hierarchical structure is introduced to optimise the overall performance. As the number of features increases, the same feature pool or classifier will not be the optimum for classifying all the activities. Unlike the conventional 'flat' classification, which classifies each example to all its available labels at the same time, the hierarchical classification uses data taxonomy to create a hierarchy of classifiers. In this experiment, the hierarchical structure firstly separates activities into several subgroups, and then each subgroup can implement distinct features and classifiers, which are more suitable for the subgroup, to improve the overall performance. Figure 26 demonstrates the structure of hierarchical classification. The activities were divided into three subgroups (A1, A6; A2, A3 & A4, A5) at the first stage due to their similarity and false alarm rate. The classifiers of all stages were SVM, with CS algorithm used for feature selection at each classification stage. The selections of each stage were independent, which means the discarded features at the first stage would come back to the feature pool for selection at the second stage. The fusion data of mD + CVD + RT was used for the feature sets at the first stage, and at the second stage for A2 and A3, A4 and A5. At the upper second stage, which classifies A1 and A6, the feature fusion of CVD + mD was employed with less computation burden due to the smaller number of features set. Figure 27 and Table 12 present the result of the hierarchical classification. The overall performance of classification is increased to approximately 92 %, which improves by ~1 % the result of the best fusion combination. The trending of specificity, precision and sensitivity are the similar to accuracy. It is observed that the accuracies of all the activities were increased except A4 and A5. The majority of misclassification of A4 and A5 was generated at the second stage. It was observed in previous spectrograms that the RT, mD and CVD of A4 and A5 were similar. This situation became more serious when the participants were the elders. From the observance, the spectrograms of A4 and A5 were almost the same, which means the differences between features extracted from them were little, increasing the possibility of misclassification. Besides, misclassification also happened at the first stage between subgroups (A2, A3 &A4, A5). If an activity was misclassified at the first stage, it would also be an error at the second stage, where the errors were accumulated.
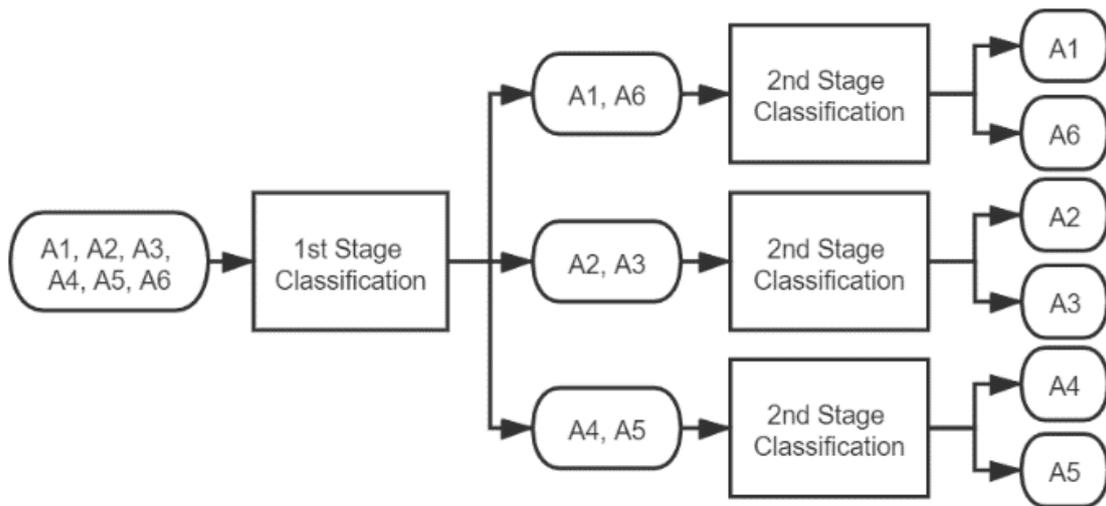
Figure 26 *A custom hierarchical classification structure based on the result and analysis.*

Figure 27 *Confusion matrix of hierarchical classification.*

Table 12 The Results of hierarchical classification.

	Accuracy	Sensitivity	Specificity	Precision
A1	1	1	1	1
A2	0.955	0.955	0.971	0.971
A3	0.952	0.952	0.949	0.949
A4	0.769	0.769	0.823	0.813
A5	0.846	0.846	0.779	0.973
A6	1	1	1	1

4.4.5 Impact of Age on HAR.

In the realm of HAR using radar, understanding the impact of demographic variables, such as age, is crucial for developing robust and universally applicable models. This section delves into the influence of age on the performance of our hybrid information approach. We bifurcate our dataset into two distinct age groups to analyse how age-related variations in movement and physiology affect the classification accuracy and reliability of our model.

Our dataset, encompassing a range of human activities captured via radar, is segmented into two groups based on age. In alignment with standard definitions, individuals aged 60 and above are categorized as older adults. This expanded age range, from the traditional 65 years, ensures a sufficient sample size for meaningful analysis. Out of 72 participants, 25 fall into this older age group. For this study, we have consciously excluded fall detection activities for the older group due to safety concerns. Instead, we focus on the first five actions, comparing their performance metrics between younger and older participants. Our analysis employs the same model used in previous studies, validated through a 10-fold cross-validation process.

Table 13 *Accuracy Assessment: Confusion Matrix for Adult Participants Only*

%	A1	A2	A3	A4	A5
A1	100	0	0	0	0
A2	0	95.5	0	2.5	2.0
A3	0	0.5	96.0	0.5	3.0
A4	0	3.5	3.5	78.5	14.5
A5	0	1.0	5.0	8.0	86.0

Table 14 *Accuracy Assessment: Confusion Matrix for the older Participants Only*

%	A1	A2	A3	A4	A5
A1	100	0	0	0	0
A2	0	92.0	0	0.9	7.1
A3	0	0	91.0	3.6	5.4
A4	0	0	2.7	66.7	30.6
A5	0	1.8	3.6	11.7	82.9

Table 13 and Table 14 indicate our preliminary result with a noticeable difference in classification accuracy between the two age groups. The younger group achieved a classification accuracy of 91.2%, while the older group accuracy was slightly lower at 86.6%. The overall trend observed in the classification results for both adults and elders align with previous findings, where activities such as walking (A1), sitting down (A2), and standing up (A3) are classified with high accuracy. However, activities A4 (bending to pick up an object) and A5 (drinking water) consistently yield lower accuracy. Notably, there is a general decrease in classification accuracy for each activity among the elder group, which can be attributed to subtler movement patterns characteristic of this age group. This is particularly evident in the case of A4, where the accuracy drops to 66.7% for elders. The activity of bending to pick up an object inherently includes bending movements, which are likely to be less pronounced in older individuals. It can be attributed to several factors: reduced speed and amplitude of movements, variability in movement patterns, or even physiological changes as muscle loss and bone deformation. This reduced amplitude and variation in movement make it more challenging for the radar-based HAR system to accurately recognize and classify these activities in the elder population. The subtlety in their movements leads to less distinct radar signatures, thereby impacting the overall classification performance.

4.5 Summary of the Chapter

In this chapter, the results from using feature extraction with human activity detection are presented. Following these results are presented after using the same and extended features to detect lameness in a variety of domestic/farmed animals using radar mD signatures, where promising classification rates have shown to be achievable for both use cases with simple features and classifiers. The work in this chapter was conducted in the initial part of the research and therefore it is inspired heavily by the techniques at the time where features and classification algorithms were the main focus. However, compared to the works in the literature, a deeper view is provided into the differences between the state of the art which was conducted either with simulations or with actors. This provided insight into the different problems a real system would encounter while attempting to automatically detect daily activity or a harmful activity such as a fall. Individual activities in this case could be detected at rates between 76.9% to 100%. By introducing this new domain of problems to be solved with radar and demonstrating the feasibility of using feature-based classification, this broadened the scope of what is possible with radar concerning activity classification.

After that, we discussed the age-impact of dataset based on our proposed method. In conclusion, the analysis of human activity recognition (HAR) using radar technology underscores the significant impact of age-related factors on classification accuracy. The comparative study between adult and elder age groups reveals that the subtler movement patterns of elders pose a challenge for radar-based HAR systems. Activities that involve more nuanced movements, such as bending to pick up an object, exhibit notably lower classification accuracies in the elder group. This highlights the importance of incorporating a diverse age range in the dataset, especially including elder participants, to develop more robust and universally applicable HAR models.

5. Radar-based Human Activity Recognition with Adaptive Thresholding towards Resource Constrained Platforms

This chapter introduces the second novelty of this thesis which is the use of an adaptive thresholding method with information selection and fusion techniques used together to improve activity recognition accuracy.

In general machine learning applications, improvements come through the increased variation of the input data type where, the more diverse the inputs are, the easier it is to identify an object or an action. In this chapter, this concept is used to show the strength of leveraging different information from varying data representation through fusion for activity recognition, while striving to decrease the computational cost and inference time. All of these aforementioned techniques are explored in this chapter with an additional analysis of the effect of classification models having prior information of the target on the recognition accuracy.

The research conducted for this chapter intended to address the problem of how to improve radar data domain representations and adjust parameters to optimise the classification accuracy in the context of assisted living. This was addressed in a twofold manner.

- First, by showing the usefulness of features from different domains with feature fusion to improve activity recognition accuracy. In Chapter 4, there were a number of features using different properties of the input signal being used for radar-based activity monitoring. In this chapter, more features are considered and used in our research for activity recognition. By using them together, we were able to demonstrate higher accuracy in activity recognition.
- Secondly, cooperative use of different radar data domains was found to further increase accuracy when adjusting the parameters during the generation of data domains. Although this was a lateral research direction to the aforementioned feature selection work, it was a result of the question of using different parameters arising as the benefit of increased variance through multi-domain inputs being observed.

This chapter is based on the following publications [36] and [37]. This chapter is organised as follows: It begins with Section 5.1, which lays the foundation by presenting the motivation for the study, elucidating the underlying reasons and

objectives driving the research. Section 5.2, titled 'Dataset Composition and Feature Extraction', delves into the specifics of the dataset used, detailing its composition and the methodologies employed for this experiment. This is followed by Section 5.3, where the focus shifts to the analysis of classification outcomes using adaptive thresholding techniques, alongside a comparative study of resource consumption. Section 5.4 introduces an innovative approach, based on proposed approach in Section 5.3, we deliver a holistic optimization system, employing evolutionary algorithms to enhance the research methodology. The chapter concludes at Section 5.5, summarizing the key findings and insights gleaned from the research and setting the stage for further exploration and application of the study's outcomes.

5.1 Motivation

Radar information in HAR can be presented in multiple domains, including but not limited to range-time, Doppler-time, and range-Doppler. Doppler-time domain or micro-Doppler (mD) signatures are typically used to exploit the small modulations in the received radar signal caused by the relative motions of limbs with respect to the trunk. Numerous studies in the literature have investigated the use of radar for human activity classification [79], [89]. Most works have focused on creating and optimizing feature extraction algorithms that generate salient features (e.g., physical, mathematical, and/or textural). However, most of radar-based HAR research focuses on spectrograms, i.e., the amplitude of micro-Doppler signatures [43], [67]. Radar data can be represented in a wide range of formats in addition to spectrograms. Finding the optimal radar data domains, as well as the most suitable combination of salient features for a given classification problem becomes an intractable problem.

More recently, deep learning and related classification techniques have gained considerable interest in radar-based HAR as they automatically extract salient features from the radar signatures. However, deep learning methods require a large amount of training data, which is more difficult to gather experimentally for radar systems than for other sensing modalities. Furthermore, radar data processing may have high computational costs because of the pre-processing steps of raw data, making it challenging to process in real time, especially if multiple radar sensors are involved. While general-purpose compute engines, especially graphics processing units (GPUs), have been the mainstay for much processing, less work is done on investigating non-tensor-based computation on resource constraint platforms. Real-world platforms, such as mobile embedded systems, are inevitably constrained by the hardware. The consideration of the balance between efficiency and performance has emerged when

exploring the most suitable algorithms. This aspect of real-time implementation of radar-based HAR approaches in constrained platforms has attracted increasing attention, as the natural yet crucial step after classification algorithms development.

Although various solutions have been developed for radar-based human activity classification in indoor scenarios, some important research questions are still not fully answered. First, most current approaches would require a long latency even at the inference/testing stage, because of complex data processing methods or deep neural networks. These research works did not consider the computational cost, focusing on classification accuracy only, so the results were satisfactory but not always suitable for embedded platforms. It is paramount for realistic deployment to focus on decreasing the footprint of the algorithms in terms of energy consumption as well as on silicon to drive the price of the product for the end-users down. Moreover, many works tend to apply the same algorithm (e.g., using the same feature) to recognize all activities in a multi-class problem.

We propose an adaptive thresholding pre-processing method to focus on the region of interest (ROI) for classification based on patented innovations. This approach is designed to reduce the computational load by outlining the ROI, i.e., the most relevant part of a spectrogram also named 'mask'. Afterwards, these 'masks' are also applied to the phase, unwrapped phase, and magnitude of the mD signature to highlight the ROI in those domains. A series of specifically designed features for the adaptive thresholding method is also introduced. To increase accuracy and reduce computational loading concurrently, we investigate feature selection and information fusion techniques to optimize performances.

This work considers and investigates two new domains of radar information, namely phase and unwrapped phase, which are seldom considered in the literature. Moreover, we expand the implementation of our feature extraction algorithm to new domains, which was not considered in our previous study. In addition, we present a detailed analysis of the effect of the thresholding value selection. Since our new experiments involve a series of new features from different domains, a hierarchical classification model, which divides the standard classification into several stages, is introduced to improve the overall performance by combining different features and domains for each stage. A comprehensive comparison between our methods and other popular neural network-based approaches is also shown.

To summarize, the specific contributions which distinguish this work from the current state of the art are summarized here:

- A novel pre-processing method with adaptive thresholding is proposed for radar-based HAR which automatically generates ROI from human mD signatures, with a set of specifically designed features for classification on different domains.
- A comprehensive evaluation of the effect of this adaptive thresholding method on the classification accuracy of individual activities and overall accuracy for the data domains under consideration (mask, masked spectrogram, masked phase, masked unwrapped phase) is provided.
- The optimization of the performance is further analysed with the fusion of data domains and selection strategies, the use of different parameters of the support vector machine, and the usage of a hierarchical method. These optimizations prove to be very beneficial to boost performances.
- The method is benchmarked against deep learning methods using the same dataset, considering metrics of training time, inference time, model size, number of parameters, accuracy, and memory footprint. This comparison shows that the proposed method can outperform deep learning methods while being computationally efficient and reduce the memory footprint.

5.2 Dataset Composition and Feature Extraction

In this experiment, we use the same dataset as in Chapter 4 [114]. A total number of 1754 motion captures were recorded from 72 participants aged 21 to 98 years old. This dataset comprises six types of daily human activities, including walking, sitting, standing, picking up an object, drinking and falling. Note that the dataset is not completely balanced, as the older individuals did not participate in the 'falling' activity recording for obvious safety concerns.

5.2.1 Adaptive Thresholding Methods

The aim of the proposed adaptive thresholding approach is to focus only on an ROI containing the contribution of the moving targets in spectrograms for subsequent classification. Areas of the spectrogram that do not convey salient information, such as the portion with low energy (dark blue in the chosen colour scale) in Figure 28, should be discarded.

From Figure 28 and the samples in the database, we can observe that the intensity varies depending on the activities being performed and the individual performing the activities. This means that it is suboptimal to apply a fixed threshold for all samples as

shown in [181]. An adaptive thresholding method is necessary to extract the ROI of each spectrogram.
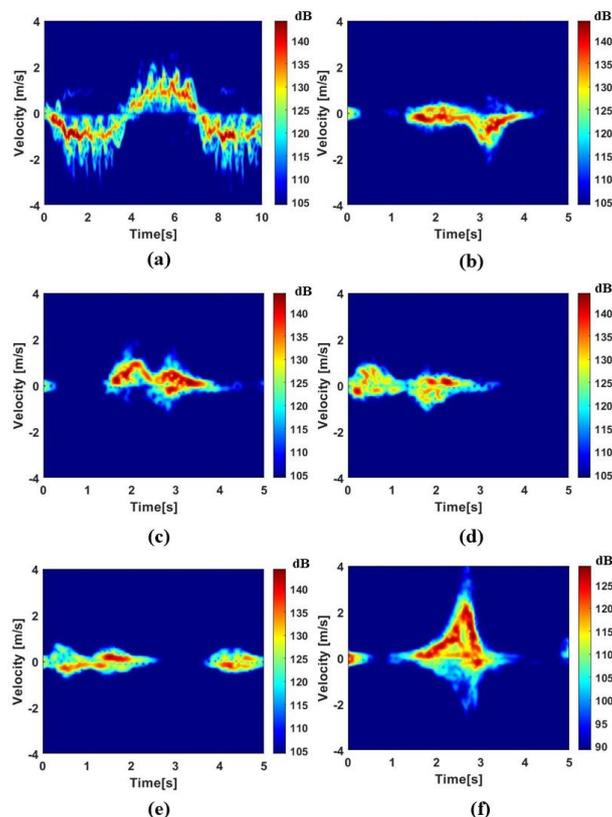
Figure 28 *The micro-Doppler signatures of typical samples of the dataset. (a)~(f) represent activities A1~A6 micro-Doppler spectrogram.*

The proposed technique uses a specific threshold T to binarize the grayscale mD signature image. This approach focuses on the ROI adaptively by selecting a threshold and then updating it based on the information contained in the window being processed. First, the spectrogram image is transformed into a grayscale image. Suppose that the grayscale image S contains N pixels, and the value of each pixel is represented as $I(x, y)$. Then the initial threshold μ is defined as in Eq. 5.1.

$$\mu = \frac{1}{N} \sum_{I(x,y) \in S} I(x, y) \tag{5.1}$$

The grayscale spectrogram image is separated into two portions based on the initial threshold value μ: P1 and P2, where P1 is the image area with a pixel value greater than μ and P2 is the image area that has a pixel value less than μ. Then, a new threshold T can be determined as in Eq. 5.2.

$$T = \frac{1}{2}\left[\frac{1}{N_1}\sum_{I(x,y)\in p_1} I(x,y) + \frac{1}{N_2}\sum_{I(x,y)\in p_2} I(x,y)\right] \tag{5.2}$$

Where N_1 and N_2 are the number of pixels in part p_1 and part p_2, respectively.

After both μ and T are obtained, their difference will be compared to a specific parameter: V, the default value of V is 0.1. If the difference is greater than V, then T will replace μ to segment the grayscale spectrogram image and a new T will be calculated using equation (2). This process is repeatedly performed until the difference is smaller than V, preserving as much of the ROI as possible. The final T value is implemented to binarize the grayscale spectrogram image, as shown in Eq. 5.3.

$$b(x,y) = \begin{cases} 1, & I(x,y) \geq T \\ 0, & I(x,y) < T \end{cases} \tag{5.3}$$

where $b(x,y)$ is the pixel value of the mask.

The binarized image, called 'mask', can be used for feature extraction. A mask is applied for this reason on the magnitude, phase, and unwrapped phase of the spectrogram, which are named 'masked spectrogram' (amplitude), 'masked phase', and 'masked unwrapped phase' images, respectively. The flow chat of how to access those domains are illustrated in Figure 29 .The 'Mask' samples for each activity are shown in Figure 30.
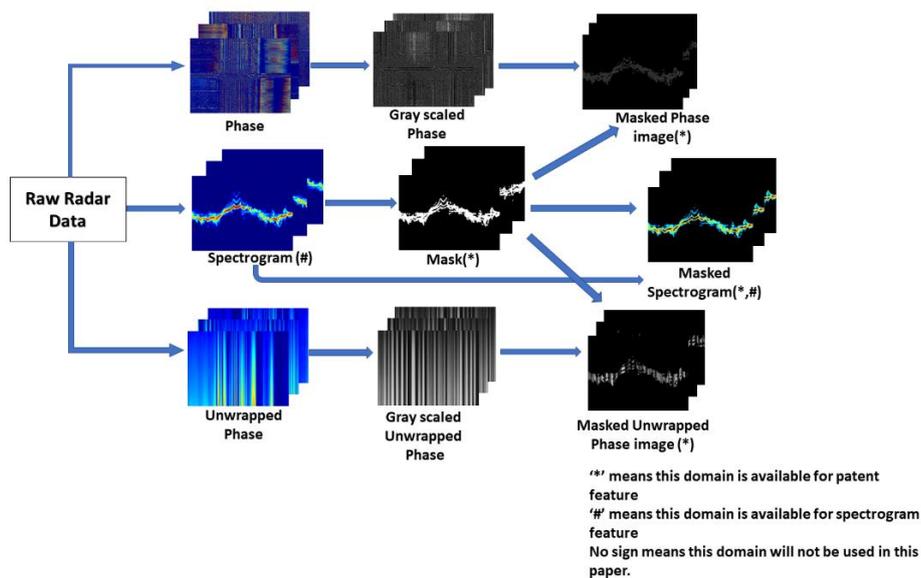
Figure 29 *An example of calculating the binary mask to generate masked phase, masked unwrapped phase, and masked spectrogram.*

The features used in this experiment are divided into two groups: group 1 features, also known as 'patent' features, due to their correspondence with our patents, whereas group 2 features are the 'radar' features, which will be referred to as 'radar' features in the following section.

- Group 1 ('patent') features: 68 features are evaluated, of which two categories are considered: the properties of the ROI and the texture of the image [187]. The first category captures the geometrical properties of the ROI, such as centroid, perimeter, and area. The second category is characterized by the spatial distribution of intensity levels within a neighbourhood of pixels, which contains information on the spatial arrangements of intensities in an ROI. All the features calculated in this experiment are listed in Table 2.

- Group 2 ('radar') features: different types of features are suggested for the spectrograms and masked spectrograms inspired from the previous literature [4, 16, 26] and from our preliminary results [27]. This includes a total of 21 features, and they are listed in Table 3.

Note that the data domains of these two groups of features are also listed in Table 15 and Table 16, with a brief description in Table 17.
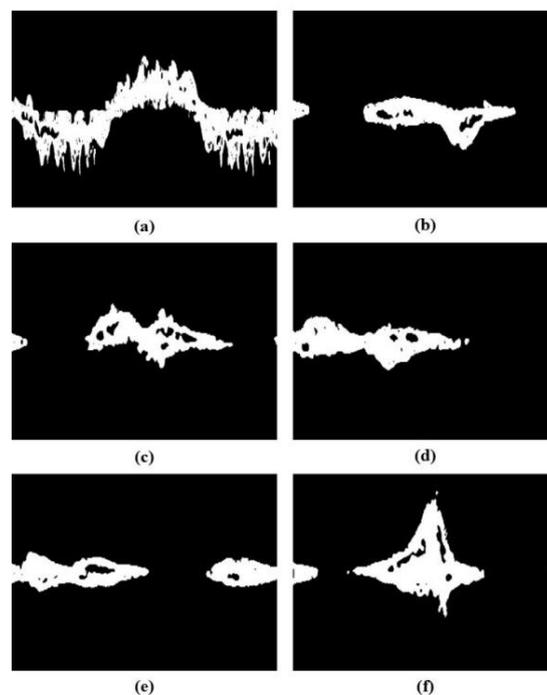
Figure 30 The binary mask of typical samples of the dataset. (a)~(f) represent activities A1~A6 micro-Doppler spectrograms.

Table 15 68 patent features and their data domains.

ROI Features	Feature Dimensions	Applicable Domains
Perimeter of ROI	1×1	
Area of ROI	1×1	
Centroid of ROI	1×2	
Eccentricity of ROI	1×1	Mask
Orientation of ROI	1×1	Masked Phase
Major and Minor Axis Length of ROI	1×2	Masked Unwrapped Phase
Textural Features	**Feature Dimensions**	Masked Spectrogram
Local Binary Pattern of image	1×59	
Moment of image	1×1	
The Number of Total Features	**68**	

Table 16 21 radar features and their data domains.

Radar Spectrogram Features	Feature Dimensions	Applicable Domains
Entropy of spectrogram	1×1	
Skewness of spectrogram	1×1	
Centroid of spectrogram (mean & variance)	1×2	
Bandwidth of spectrogram (mean & variance)	1×2	Spectrogram
Energy Curve (mean & variance & Trapezoidal numerical integration)	1×3	Masked Spectrogram
Singular Vector Decomposition (mean and variance of the first three vectors of components)	1×12	
The Number of Total Features	**21**	

Table 17 Description of New extracted features

Category	Brief description
Region of Interest features	The features described are specific in characterizing distinct areas within an image. They include measurements such as the perimeter of the boundary, the overall area encompassed, and the centroid or centre of mass. Additionally, they assess the elongation and orientation of the region, along with the dimensions of the major and minor axes of an ellipse optimally fitted to the area.
Textural Features	These features are employed to delineate the texture of an image. They analyse local texture patterns using Local Binary Patterns (LBP), a method that encodes the texture by comparing each pixel with its neighbourhoods. Additionally, they involve measuring statistical descriptors, such as moments, to capture the spatial distribution of intensity values, providing a comprehensive understanding of the image's textural characteristics.

5.2.2 Feature Selection, Fusion, and Hierarchical Structure.

Feature selection approaches are applied to further improve the performance and reduce the computational complexity.

In this case, we evaluate a wrapper method - sequential floating forward selection (SFFS), which is based on sequential forward selection (SFS). SFS determines the optimal feature combinations by ranking the features in accordance with a classifier and its accuracy as a measure. In feature search, it begins with the empty feature subset and increase the dimension one-by-one. Unlike the traditional SFS, SFFS not only adds features progressively but also eliminates features from the selected subset when the classifier deems it to improve performances after eliminating a specific feature. The algorithm is given below:

Algorithm 4 Sequential Floating Forward Selection

Initialization: the optimal features set Z

set the initial Z is empty;

    while the accuracy of Z is increasing **do**

        Find feature F which yields the best predefined criterion with Z;

        Update F to feature set Z;

        Evaluate the potential removal of each feature in the current Z.

Until the criterion stop becoming better

End

Output *Z*

Data fusion is the process of integrating multiple data sources to produce information which can overcome limitation caused by any single data source. It could be attained through different levels of abstraction [148], which are commonly divided into three levels - signal, feature, and decision. Feature level fusion cascades the same-labelled features from various sources, as in Eq. 5.4, where ∩ represents the concatenation of features from different domains.

$$F_{fusion} = F_{mask} \cap F_{phase} \cap F_{unwrap} \tag{5.4}$$
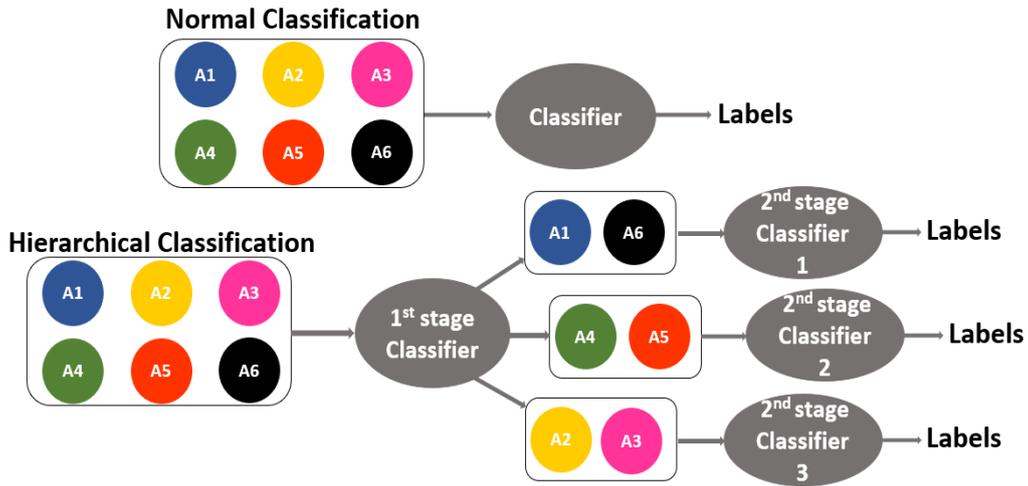
Figure 31 *Example of conventional and hierarchical classification.*

Unlike traditional supervised classification approaches, which feed all activities into the classifier simultaneously, the proposed hierarchical structure classifies the activities into several sub-groups based on their similarity or misclassification rate. As is shown in Figure 31, the hierarchical model permits the use of distinct feature sets and algorithms at different stages, and therefore improves the overall performance.

5.3 Classification Results of the Adaptive Thresholding Methods and Resource Consumption Comparison

5.3.1 Classifiers.

We begin with an evaluation of the proposed threshold-based approach on spectrograms, phase images and unwrapped phase images, followed by the extraction of features and comparisons among different feature domains. Then, information fusion and feature selection are utilized to improve performances. Finally, we design a hierarchical classification structure based on the prior results to boost the overall performances. Based on the features listed in Table 15 and Table 16, the classification models are trained using several support vector machine (SVM) classifiers. SVM is one of the machine learning methods which can be used for the classification task. SVM can provide a unique hyperplane to separate learning samples for different classes. This process depends on the choices of kernel functions and hyper-parameters. To analyse which kernel function would be suitable for our data, different kernel functions, namely linear, polynomial (quadratic and cubic), and radial basis kernel (RBF), are implemented and compared with a 10-fold cross-validation method.

Table 18 Classifications accuracy in % for different SVM kernel functions and different data domains. The green colour indicates the best performance amongst the different methods. The quadratic polynomial SVM kernel consistently provides the best performance.

%	Linear	Polynomial (quadratic)	Polynomial (cubic)	RBF
Mask	81.0	84.9	83.0	79.9
Masked phase	82.3	83.1	80.6	80.9
Masked unwrapped phase	72.9	73.1	71.1	72.0
Spectrogram	78.0	80.3	79.5	78.2
Masked spectrogram (patent)	83.4	83.6	80.7	80.0
Masked spectrogram (Radar)	84.1	85.7	82.3	84.1
Average across data domains	80.3	81.8	79.5	79.2

According to these preliminary results in Table 18, the SVM model with the quadratic kernel (second-degree polynomial) achieves the highest accuracy consistently across all domains. Furthermore, when comparing the spectrogram to the masked spectrogram (Table 4), the usage of our proposed adaptive thresholding method improves the overall accuracy by 3.3% (from 80.3% to 83.6%) and 5.4% (from 80.3% to 85.7%) when 'patent' and 'radar' features are used, respectively. Afterwards, an

analysis of thresholding values is conducted to further improve the prediction performance, as well as to achieve a better understanding of the interactions between thresholding values, various domains, and the data.

5.3.2 Thresholding Evaluation.

To investigate the impact of the adaptive threshold T, seven values ranging from $T - 10$ to $T + 20$ to obtain the binary masks are applied on the spectrogram, phase, and unwrapped phase data. These different data domains are analyzed separately to determine their contribution to classification. They are divided into three types in terms of features: for binary mask and masked (unwrapped) phase, the 'patent' features are implemented. For the spectrogram, the 'radar' features are used, and for the masked spectrogram data, both 'patent' and 'radar' features are implemented. At this stage, a robust quadratic-kernel support vector machine (Q-SVM) algorithm with 10-fold cross-validation is adopted for activity classification.

Table 19 Classification results for the mask data domain, with patent features and different threshold values.

%	A1	A2	A3	A4	A5	A6	Avg
T-10	96.8	91.3	87.7	63.0	77.1	89.9	84.3
T-5	97.2	92.1	90.4	63.4	77.6	89.3	85.0
T	97.8	90.7	89.3	62.5	76.7	92.4	84.9
T+5	98.7	93.3	86.9	64.7	69.3	89.9	83.8
T+10	96.1	89.7	90.0	61.2	74.8	90.4	83.7
T+15	97.4	90.4	90.3	64.4	75.1	91.2	84.8
T+20	97.8	91.0	88.4	62.0	74.5	90.4	84.0
Avg	97.4	91.2	89.0	63.0	75.1	90.5	

Table 20 Classification results for the masked phase data domain, with patent features and different threshold values.

%	A1	A2	A3	A4	A5	A6	Avg
T-10	98.4	88.8	85.8	65.7	70.3	92.6	83.6
T-5	99.4	86.6	87.4	68.0	71.6	91.9	84.1
T	99.7	87.8	86.7	63.4	69.6	91.4	83.1
T+5	98.4	89.1	84.5	61.5	67.8	91.9	82.2
T+10	99.7	87.1	84.8	61.8	67.1	93.9	82.4
T+15	99.7	87.5	84.2	59.9	65.1	93.4	81.6
T+20	100	86.2	85.1	59.5	69.3	94.4	82.4
Avg	99.3	87.6	85.5	62.9	68.7	92.8	

Table 21 Classification results for the masked unwrapped phase data domain, with patent features and different threshold values.

%	A1	A2	A3	A4	A5	A6	Avg
T-10	98.4	67.6	65.5	50.2	70.3	85.4	72.9
T-5	99.4	69.2	65.1	54.7	66.1	85.4	73.3
T	98.0	70.1	59.2	54.0	69.3	87.9	73.1
T+5	98.4	65.7	61.6	53.0	68.7	86.6	72.3
T+10	98.4	71.1	61.9	50.2	67.7	86.9	72.7
T+15	98.7	67.3	59.6	50.5	68.7	87.4	72.0
T+20	99.4	68.9	60.3	53.0	69.3	84.9	72.6
Avg	98.7	68.5	61.8	52.2	68.6	86.3	

Table 19 to Table 21 illustrate the initial results using the mask, masked phase, and masked unwrapped phase data domains, with different threshold values. Table 19 shows that an average accuracy of 85% is achieved when the binary mask is used with threshold $T - 5$. The result of the masked unwrapped image has ~10% performance degradation compared with mask and masked phase images. It is mainly because the accuracy decreases greatly in both A2 and A3 activities and slightly in A4 and A5 activities, which are reduced by approximately 20%, 26%, 10% and 5%, respectively. Different thresholds yield the best accuracy for individual activities. For instance, 100% accuracy is achieved for walking with $T + 20$ in the masked phase domain.

Table 22 to Table 24 show the initial results using spectrogram and masked spectrogram data domains, with thresholding ranging from $T - 10$ to $T + 20$. Spectrograms with both patent and radar features achieve the highest accuracy at 90.0% with the threshold value T. Compared to using both 'patent' and 'radar' features together, implementing only one of them has a negative effect on performances causing a ~5% drop in accuracy. For spectrograms with 'radar' features and both 'radar' and 'patent' features, the maximum average accuracy is obtained with threshold T unaltered, which are 85.7% and 90.0%, respectively. The spectrogram with 'patent' features reaches its peak accuracy of 84.8% with a threshold value of $T + 5$. Comparing the use of 'radar' and 'patent' features separately on spectrograms, 'radar' features yield better performances with ~1% improvement overall. However, it should be noted that the 'patent' features can be applied on all data domains and not just on mD signatures, so they are in a sense more versatile.

Table 22 Classification results for the spectrogram ('No Mask' highlighted in orange) and masked spectrogram data domains, with radar features and different threshold values.

%	A1	A2	A3	A4	A5	A6	Avg
No Mask	94.0	80.2	82.4	63.6	73.7	87.9	80.3
T-10	100	87.1	86.3	61.2	70.8	93.1	83.1
T-5	100	89.3	87.9	62.6	72.1	92.4	84.1
T	100	91.3	89.2	63.9	75.5	94.4	85.7
T+5	100	89.6	88.4	62.1	71.8	94.4	84.4
T+10	100	88.0	87.7	61.5	70.3	93.9	83.6
T+15	99.7	87.1	86.3	63.0	69.6	92.4	83.0
T+20	99.7	86.7	85.6	64.4	67.9	91.7	82.6
Avg	99.9	88.4	87.3	61.7	71.1	93.2	

Table 23 Classification results for the masked spectrogram data domain, with patent features and different threshold values.

%	A1	A2	A3	A4	A5	A6	Avg
T-10	100	91.9	89.1	62.9	73.8	90.4	84.7
T-5	99.7	91.6	88.8	61.5	74.8	91.9	84.7
T	100	89.8	87	59.8	72.7	92.4	83.6
T+5	100	91.9	88.8	61.5	74.1	92.4	84.8
T+10	100	87.1	84.8	61.8	67.1	93.9	82.5
T+15	100	90.9	87.4	65.4	73.8	90.4	84.7
T+20	100	90.5	88.8	60.5	70.6	90.9	83.6
Avg	99.9	90.5	87.8	61.2	72.4	91.8	

Table 24 Classification results for the masked spectrogram data domain, with both patent and radar features and different threshold values.

%	A1	A2	A3	A4	A5	A6	Avg
T-10	100	94.7	93.0	80.1	78.4	93.4	89.9
T-5	100	93.7	94.4	79.0	77.3	94.9	89.9
T	100	94.7	92.3	77.6	80.1	95.4	90.0
T+5	100	95.1	90.9	76.6	80.1	95.4	89.7
T+10	100	93	93.7	77.3	79.4	96.4	89.9
T+15	100	91.6	91.2	78.3	79.4	94.4	89.2
T+20	100	90.4	93.3	75.9	79.0	95.4	89.0
Avg	100	93.3	92.7	77.8	79.1	95.0	

In summary, from this analysis the overall accuracies of mask, masked phase, and masked spectrogram data domains with patent features are increased when the threshold value changes, which means the exploration in this range of threshold values has positive effects on the results. The masked spectrogram with both patent and radar features outperforms other domains, which has achieved the highest accuracy of 90.0%.

5.3.3 Feature Level Fusion and Feature Selection

After analysing the performances of the binary mask, masked phase, masked unwrapped phase, and masked spectrogram individually, these data domains are combined with feature level fusion. In each domain, the group with the best overall performance is selected as features for the feature fusion. Based on the previous results, in the fusion for the binary mask and masked phase/unwrapped phase data domains, we only choose the features extracted with threshold $T - 5$. For spectrogram and masked spectrogram domains, both features with threshold T are chosen.

At this stage, the Q-SVM and 10-fold cross-validation are still used. In this case, we provide seven combinations of features, which are mask + masked phase (Comb 1), mask + masked unwrapped phase (Comb 2), masked phase + masked unwrapped phase (Comb 3), mask + masked spectrogram (Comb 4), mask + masked phase +masked spectrogram (Comb 5), mask + masked unwrapped phase + masked spectrogram (Comb 6) and all together (Comb 7). These are shown in Table 25.

To further improve the accuracy, reduce the computational load, and evaluate the feature selection approach, the SFFS is applied to both individual results with the best average accuracy and the fusion results of all combos listed above. The individual results are shown in Figure 32, and the combination results are shown in Figure 33. These results are also summarized in Table 26.

Table 25 Classification results for the different combinations of data domains without feature fusion.

%	A1	A2	A3	A4	A5	A6	Avg
Comb 1	99.7	92.3	93.5	80.3	82.3	95.5	90.6
Comb 2	100	88.8	88.2	73.1	82.9	93.4	87.7
Comb 3	98.7	88.8	87.1	67.3	74.5	92.9	84.9
Comb 4	100	95.4	96.1	78.2	78.2	94.5	90.4
Comb 5	99.7	92.6	93.5	80.9	83.5	94.2	90.7
Comb 6	98.7	92.3	92.6	79.1	81.2	93	89.9
Comb 7	100	95.2	95.0	80.3	80.8	95.9	91.2

To further improve the accuracy, reduce the computational load, and evaluate the feature selection approach, the SFFS is applied to both individual results with the best average accuracy and the fusion results of all combos listed above. The individual results are shown in Figure 32, and the combination results are shown in Figure 33. The t-SNE images are used to illustrate the distribution of features before and after feature selection, which is also shown in Figure 34. These results are also summarized in Table 26.

The accuracy increase provided by the SFFS is limited. However, the dimension of

the feature pool is significantly decreased. Generally, the number of features is reduced by up to ~80% compared to the starting count. The accuracy increases by ~1% to ~4% for individually used data and by ~1 to ~2% for fusion results.
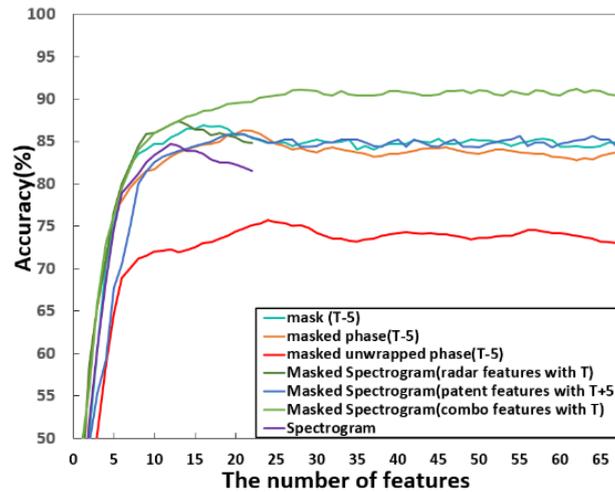
Figure 32 *Feature selection with SFFS, results for individual data domains.*
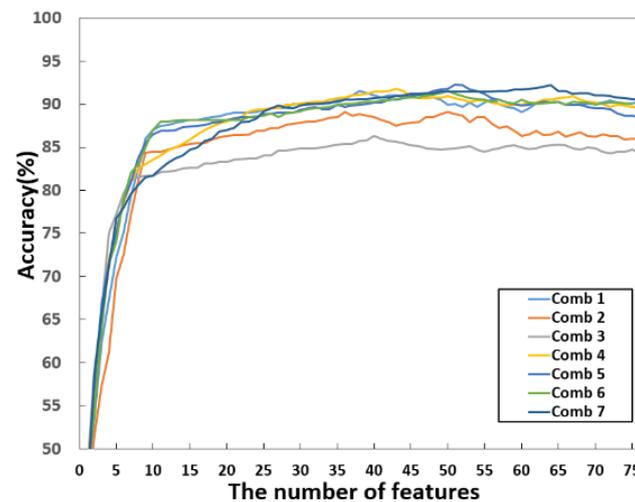
Figure 33 *Feature selection with SFFS for feature-level fusion approaches across different data domains.*

Table 26 Performance comparison using feature selection via SFFS (across different data domains and their combinations)

Methods	No. of Original Features	No. of Selected Features	Accuracy (%)
Mask	68	16	86.9
Masked Phase	68	21	86.3
Masked Unwrapped Phase	68	24	75.7
Spectrogram	21	12	84.7

Masked Spectrogram (Radar features)	21	13	87.4
Masked Spectrogram (Patent features)	68	20	85.9
Masked Spectrogram (Both features together)	89	28	91.1
Data Domain Comb 1	136	38	91.5
Data Domain Comb 2	136	36	89.1
Data Domain Comb 3	136	40	86.3
Data Domain Comb 4	157	43	91.8
Data Domain Comb 5	225	51	91.3
Data Domain Comb 6	225	50	91.5
Data Domain Comb 7	**314**	**64**	**92.2**
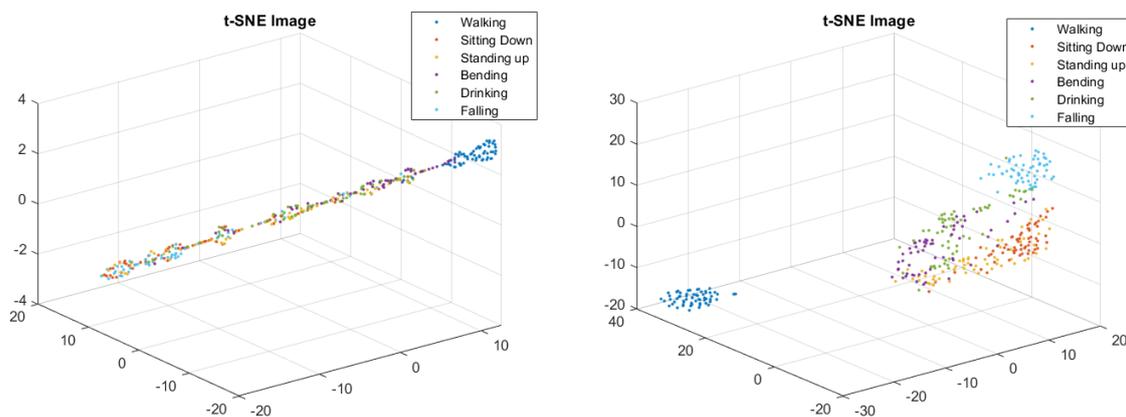
Figure 34 *The t-SNE image before and after feature selection process.*

Note that the binary mask provides the most lightweight implementation with the highest accuracy 86.9% for individual data domains with 16 features. The masked spectrogram data provides the highest accuracy 91.1% for the use of a single radar data domain with 28 features. For combined domains, Comb 7 achieves the highest accuracy among all combinations of domains by cascading all types of features, which yields the best accuracy of 92.2% with 64 features. Compared to using single domain 94.7features without feature selection, this improvement is from ~2% (masked spectrogram) to ~18.9% (masked unwrapped phase). However, misclassification events remain, especially for activities A4 and A5.

5.3.4 The Hierarchical Structure

The hierarchical structure is applied for optimizing classification. The activities are

grouped as in our previous study based on their similarity and false alarm rate. The six activities are first divided into three groups: A1 and A6, A2 and A3, A4 and A5, as shown in Figure 31.

The reason why division of activities into three groups—walking and falling, standing and sitting, and picking and drinking—can be rationalized from both physical and t-SNE image perspectives. This classification leverages the inherent similarities and differences in the physical dynamics and the data representation space of these activities.

From physical perspective, walking and falling are grouped together due to their dynamic motion and significant changes in body posture and centre of gravity, with walking characterized by controlled forward motion and falling by uncontrolled descent. This contrasts with the activities of standing and sitting, which involve transitions between two primary static postures—upright and seated—highlighted by controlled vertical movements without the significant lateral or forward motion observed in walking or falling. Similarly, bending to pick up an object and drinking are paired due to their reliance on precise, controlled movements and a focus on maintaining balance, requiring fine motor skills and coordination for different purposes: bending engages the lower back and legs in a vertical motion, while drinking involves arm and hand coordination. These activities are distinguished from the more gross motor skills involved in walking, falling, standing, and sitting, underscoring their shared emphasis on stability and precise movement control.

From t-SNE image perspective, it can be observed from Figure 34 that, standing and sitting are clustered together in t-SNE visualizations because the transition between standing and sitting involves a relatively predictable change in posture and sensor readings, especially when compared to the more erratic patterns seen in falling or the varied pace of walking.

In addition, bending to pick up an object and drinking could cluster together due to the subtle specific movements involved, which may not exhibit as wide a range of sensor data variability as walking or falling but are distinct from the simple posture transitions of standing and sitting. The fine-grained movements and the specific sequence of actions in both activities lead to a pattern of data points that are more similar to each other.

In the t-SNE visualization, walking and falling activities are distinctly separated from each other and from the other four classes, leading to their categorization into a single

group. In summary, the hierarchical classification of these six activities into three groups is supported by both the physical characteristics of the activities and their representation in t-SNE visualizations. This approach not only reflects the underlying physical dynamics and control mechanisms involved in each activity but also aligns with the patterns observed in the data when analysed through dimensionality reduction techniques like t-SNE.

These three pairs will go through the first classification stage, and this is followed for each pair by a binary classification. In the first stage, Comb 7 is implemented with Q-SVM and SFFS (64 features). Comb 7 is also used in the second stage of binary classification for A1&A6. For A2&A3, Comb 4 is used with Q-SVM and SFFS algorithm (43 features). For A4&A5, Comb 5 is applied, with Q-SVM and SFFS (51 features). The confusion matrices of the two classification stages are shown in Figure 35.

Table (a):

%	A1&A6	A2&A3	A4&A5
A1&A6	98.6	1.1	0.3
A2&A3	0	94.7	5.3
A4&A5	0	0.6	99.4

(True Class / Predicted Class)

(a)

Table (b):

%	A1	A2	A3	A4	A5	A6
A1	100					
A2		95.2		2.9	1.9	
A3			94.2	2.5	3.3	
A4			0.7	84.2	15.1	
A5			0.6	12.9	86.5	
A6		0.5	2.5	0.5		96.5

(True Class / Predicted Class)

(b)
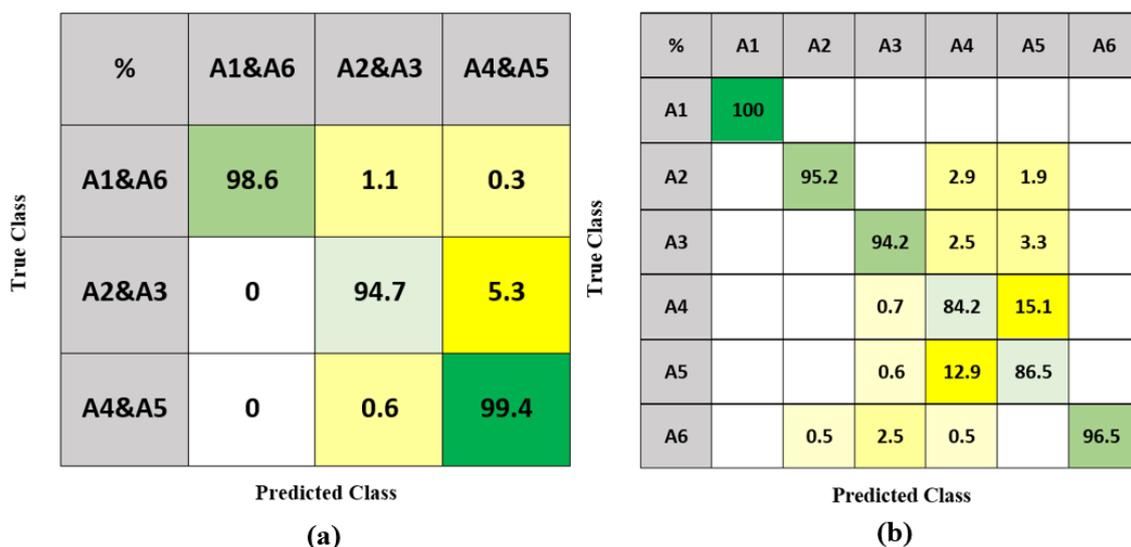
Figure 35 Confusion matrix of 1st stage classification (a) and result of the 2nd hierarchical classification (b).

The custom hierarchical structure has an average accuracy of 93.1%, which 0.9% compared to feature-level fusion. Although the accuracies of A2 (-2.8%), A3 (-3.2%) and A6 (-0.8%) decreased by 2.8, 3.2, and 0.8%, respectively, this approach still has the best overall performance. A4 and A5 have the largest improvements with 1.9% and 4.4%, respectively. The accuracy for A1 remains at 100%. A1 consistently has the best performance over the six activities in our experiments. We hypothesize that this is happening because A1, which is walking, is much more diverse than other in-place activities (A2-A6). As a periodic and translational activity, it generates richer Doppler signatures than in-place activities, leading to more distinct features, which make it easier to recognize the activity and thus achieve the best performance.

5.3.5 Discussion

To evaluate the performance of our methods, different alternative classification models are used with the same dataset including those based on deep learning approaches. The compared models include K-Nearest Neighbour (KNN) model with K=10, VGG19 [188], Resnet50 [189], NASNet-Mobile [190], Densenet201 [191], and ShuffleNet [192]. The performance of the models can be analysed according to three categories –

- *Time*, which refers to how long the model takes to train and to produce an inference,

- *Memory footprint*, which deals with how much memory the model occupies,

- A*ccuracy*, which presents the ability to infer the correct class of activities.

Specifically, for the time performance, both training time and inference time are assessed separately, and for the memory footprint, the assessment investigates both the number of parameters and the model size. We implement the benchmark analysis on a workstation with an Intel Core I5-9400F CPU 2.9 GHz and NVIDIA GeForce RTX 2060 GPU. The result of this benchmark is shown in Table 13.

The inference time shown in Table 27 is an average per data inference of over 30 runs for all models. In general, the time required to train a deep learning model varies depending on the number of network layers. ShuffleNet is the fastest deep learning model in the list, taking 232 seconds. In comparison to alternative network-based approaches, our approach has the fastest training time of 20.58 seconds, which is only ~9% of the training time of ShuffleNet. VGG-19 is the fastest deep learning method in terms of inference time with 16.243 ms. Our proposed method achieves an inference time of 15.646 ms, which is comparable.

The relevant parameters in this analysis are weights that are learnt during training. They are weight matrices that contribute to the model's predictive capability, changed during the back-propagation process. There are millions of parameters produced at the learning stage, and hence the parameters are counted in millions (M). From the comparison of the model sizes in Table 13, we can deduce that the larger the size of the deep learning models, the more parameters they had. The size of VGG19, ResNet50, and DenseNet201 are 558.48 MB, 94.82 MB, and 75.08 MB. On the other

hand, NASNet-Mobile and ShuffleNet are much smaller in size at 19.42 MB and 3.97 MB, respectively. The size of our model is only 2.06 MB, which is a 48.11% reduction compared to ShuffleNet and a 99.6% reduction compared to VGG19.

Table 27 also illustrates the accuracy and memory usage of the models using the same dataset which is used in this article. From the memory footprint reported, deep learning models require a considerable memory footprint. ShuffleNet has the lowest footprint in the listed deep learning algorithms. However, our method requires 89.13 MB, which is only 6.21% of the footprint required for ShuffleNet. In addition, our method requires less than one-tenth of the training time compared to the fastest deep learning method while yielding the highest accuracy at 93.10%, which is 1.15% higher than the most accurate deep learning method. Meanwhile, the KNN model with our adaptive thresholding method achieves an accuracy of 85.2%. The result shows that our adaptive thresholding method can also achieve good accuracy with other classifiers instead of SVM, which demonstrates that our method for pre-processing and multi-domain exploration is salient and versatile. This experiment proposed a combination of the adaptive thresholding algorithm with the Q-SVM (machine learning based) model, which is more suitable for resource constrained platforms because of its reduced footprint while maintaining speed and increasing accuracy.

Table 27 Computational metrics and accuracy comparison of proposed adaptive thresholding method and alternative approaches.

Model	Training Time (s)	Inference Time (ms)	Model Size (MB)	Params (M)	Accuracy (%)	Memory Footprint (MB)
KNN	9.25	10.20	1.50	0	85.20	40.21
VGG 19[35]	2173	16.24	558.48	139.60	73.99	2870.69
ResNet50[36]	330	20.11	94.82	23.53	87.93	1468.53
NASNet-Mobile[37]	1889	105.11	19.42	4.28	86.07	1558.85
DenseNet201[38]	2199	87.85	75.08	17.86	91.95	1590.28
ShuffleNet[39]	232	22.26	3.97	1.02	91.02	1435.55
Adaptive Thresholding + Hierarchical	**20.58**	**15.64**	**2.06**	**0**	**93.10**	**89.13**

5.3.6 The Difference Between the Proposed Methods and Traditional Methods.

The proposed method and traditional background/foreground separation techniques, such as histogram-based approaches [193], are both utilized in image processing for segmenting images into background and foreground components [194]. While they

share the common goal of image segmentation, their methodologies and underlying principles exhibit notable differences alongside their similarities.

For similarities, both methods aim to separate the foreground (region of interest) from the background, facilitating tasks like object detection and image analysis. Besides, the thresholding value is automated extracted, reducing the need for manual intervention in choosing the threshold for segmentation, which is essential for processing large datasets or real-time applications [195].

However, there are also many differences between two approaches:

1. Thresholding Technique:

The Proposed Method: It is a global thresholding technique that calculates the optimum threshold by minimizing intra-class variance or, equivalently, maximizing inter-class variance. This means it finds a single threshold that best separates the pixel values into two classes, assuming the image contains two classes (foreground and background).

Traditional Histogram-Based Approaches [194]: These can be more varied and might not specifically aim to optimize the variance between classes. For example, a simple histogram-based method might set a threshold at a fixed intensity level or use multiple thresholds based on the histogram's peaks and valleys without explicitly considering the statistical properties of the classes.

2. Assumptions:

The Proposed Method: Assumes that the image contains two distinct classes and seeks to find a threshold that minimizes the overlap between them. It works best when the histogram of the image has a bimodal distribution, clearly indicating two classes.

Traditional Background/Foreground Separation [194]: Depending on the specific technique, these methods may not make as strong assumptions about the number of classes or the distribution of pixel intensities. Some methods might be more heuristic and less statistical in nature.

3. Robustness and Complexity:

The Proposed Method: Offers a more robust approach in scenarios where the assumption of two distinct classes holds true, as it systematically calculates the optimal threshold. However, its reliance on the bimodal distribution can be a limitation in complex images.

Traditional Histogram-Based Approaches [194]: These methods can vary in complexity and robustness. Some might be simpler but less effective in diverse conditions, while others might incorporate additional rules or criteria to handle more complex scenarios but at the cost of increased computational complexity.

In conclusion, while both separation techniques aim to segment images, they differ in their approach, assumptions, and best use cases. The proposed approach provides a statistically grounded solution for images with clear bimodal distributions, whereas traditional methods offer a broader toolkit with varying levels of complexity and assumptions about the image's content.

5.3.7 Summary

In Section 5.3, the research meticulously evaluates the effectiveness of adaptive thresholding in enhancing classification accuracy while also considering the computational and resource implications. This detailed analysis reveals both the strengths and limitations of the thresholding approach, particularly in terms of scalability and adaptability to diverse and complex datasets. While adaptive thresholding shows promise in specific contexts, its application highlights the need for a more flexible and comprehensive optimization strategy that can dynamically adjust to varying data characteristics and requirements. This shift is rooted in the realization that emerged from a challenging endeavour of applying similar methodologies to lameness detection of cattle. The initial approach, while theoretically sound, faced practical limitations in its application to livestock. Thus, Section 5.4 marks a significant transition in the research journey, where the lessons learned from the practical challenges encountered in the field guide the adoption of a more sophisticated, adaptive, and holistic approach to optimization. This section delves into the intricacies of the Evolutionary Algorithm, illustrating how this innovative method revolutionizes the research approach, enhancing its applicability and effectiveness in complex, real-world scenarios.

5.4 A Holistic Optimization Using AI Techniques

Building on previous radar-based human activity recognition, we expand the micro-Doppler signature (MDS) to 6 domains and exploit each domain with a set of handcrafted features derived from the literature and our patents. We employ an adaptive thresholding method to isolate the region of interest, which is then applied in the other domains. To reduce the computational burden and accelerate the convergence to an optimal solution for classification accuracy, a holistic approach to

human activity recognition optimization is proposed by using a surrogate model-assisted differential evolutionary algorithm (SADEA-I) to jointly optimize signal processing, adaptive thresholding and classification parameters for human activity recognition.

Optimizing the pre-processing parameters for thresholding, MDS generation and the selection of the best representations to maximize classification accuracy is a challenge for parametric sweeping, since it is intractable for such problems with ever-expanding parameters to tune, which is computationally expensive and time-consuming. Furthermore, some off-the-shelf modern optimization techniques, such as differential evolution (DE) [196] and particle swarm optimization [197], are still too expensive to utilize in this case. This experiment investigates a holistic HAR optimization using Surrogate Model-Assisted DE Algorithm (SADEA-I) in which the MDS domain is used as a proof of concept before expanding this technique to a wider range of parameters and other radar data domains to include the pre-processing, domain selection, feature selection, and machine learning/deep learning joint optimization. The specific contributions which distinguish this work from the current state of the art are summarized here:

- SADEA-I, as a machine learning-assisted Evolutionary Algorithm (EA), is employed for the efficient joint optimization of the signal pre-processing parameters and a combination of radar representations to minimize the classification error rate through an effective global search and performance predictions to reduce the computational load.

- A novel radar-based HAR information processing method using adaptive thresholding that automatically produces ROI for human MDS yielding 6 different domains – Unfiltered MDS amplitude, Binary Mask, Masked phase, Masked unwrapped phase, Masked spectrogram (patent), Masked MDS complex matrix, with a set of handcrafted features for classification.

- Given that the domains can be visualized in an image format, we also implement a Convolutional Neural Network (CNN). This serves to showcase the enhanced integration of our optimization strategies with deep learning algorithms, highlighting the advancement made in our approach.

5.4.1 Optimization Problem Formulation and Challenges

The 10-fold classification error rate $P_x(t)$ serves as the minimization objective. Mathematically, the minimization problem can be expressed as Eq. (5.5),

$$\min_x P_x(t) \qquad\qquad (5.5)$$

where x is a set of parameters in the HAR system that define the system and t is the validation sample which is randomly chosen for every function evaluation.

At the optimization stage, there are 12 parameters in total. They are divided into two groups according to whether they are continuous or not.

The first group parameter is discrete, including 'Window', 'Mask', 'Masked phase', 'Masked unwrapped phase', 'Spectrogram', 'Masked spectrogram (patent)', and 'Masked spectrogram (radar)'. They have two options: 0 and 1. For 'Window', 0 means rectangular window and 1 means Hamming window. For the other six parameters, 0 means combination excludes it, and 1 means combination includes it.

The second group parameter is continuous, including overlapping factor F (range from 0.5 to 0.95), time window length W (range from 100 to 1000), V (range from 0.01 to 1) and T_e (range from -20 to 20) in thresholding method, clipping time K (range from 1.5 s to 5 s). Note that T_e has to be integer and other parameter can be any value in the range.

Parametric sweeping is a widely used method to find the set of parameters x to optimize classification accuracy. However, this method only works for small-scale problems due to its large computing overhead. For the targeted 12-dimensional problem with wide search ranges, where each function evaluation costs 3 minutes in our workstation, the computing overhead is still unaffordable even using modern intelligent optimization techniques, e.g., DE algorithm [196], genetic algorithm [198], and particle swarm optimization [197]. Hence, obtaining optimal global solutions efficiently becomes the central challenge.

5.4.2 The SADEA algorithm and Parameter Settings.

To the best of our knowledge, there are few off-the-shelf methods to address the above challenge in the signal processing field. The SADEA-I [199], [200], usually applied to antenna design optimization, is adapted for HAR. SADEA-I can obtain comparable optimization ability with the DE algorithm, which shows excellent

optimization ability, while reducing by a factor of 5 to 10 the necessary number of function evaluations compared to using the standard DE [13]. The flowchart of SADEA-I is illustrated in Figure 36. The adapted SADEA-I algorithm for HAR is summarized as follows.

- Step 1: Latin Hypercube Sampling (LHS) is applied to generate *NP* numbers of samples to form an initial database with *NP* exact function evaluations.
- Step 2: The classification error rate is minimized until a suitable *x* with the lowest classification error rate among those in the database satisfies the preset error rate threshold or the total optimization time exceeds the preset threshold.
- Step 3: Obtain λ best sets of solutions as the parent population to form a child population by applying the DE mutation and crossover operations [196].
- Step 4: For each child solution, obtain its training data by collecting their nearest τ known samples from the database and train Gaussian Process (GP) models [201]. Predict classification error rate and uncertainty for each child solution.
- Step 5: Prescreen the child population with the predicted values and prediction uncertainty using the lower confidence bound (LCB) method [202], [203] and obtain the estimated most promising candidate solution. Carry out an exact function evaluation for it. The candidate solution and its function value are appended to the database and go back to Step 2.

GP models in this case are trained to predict and suggest the most promising candidate parameter set in each iteration. Hence, it significantly reduces the number of exact function evaluations needed for the search. To make GP and DE work harmoniously, the surrogate model-aware evolutionary framework is used and more details can be found in [204].

The algorithm parameters are set following [199], [205]: *NP* is set to be 50. Both mutation rate and crossover rate are set to be 0.8. λ is set to be the same as *NP*. The number of training samples τ, is also set to be 50 to train GP models. In LCB prescreening, LCB parameter is set to be 2.
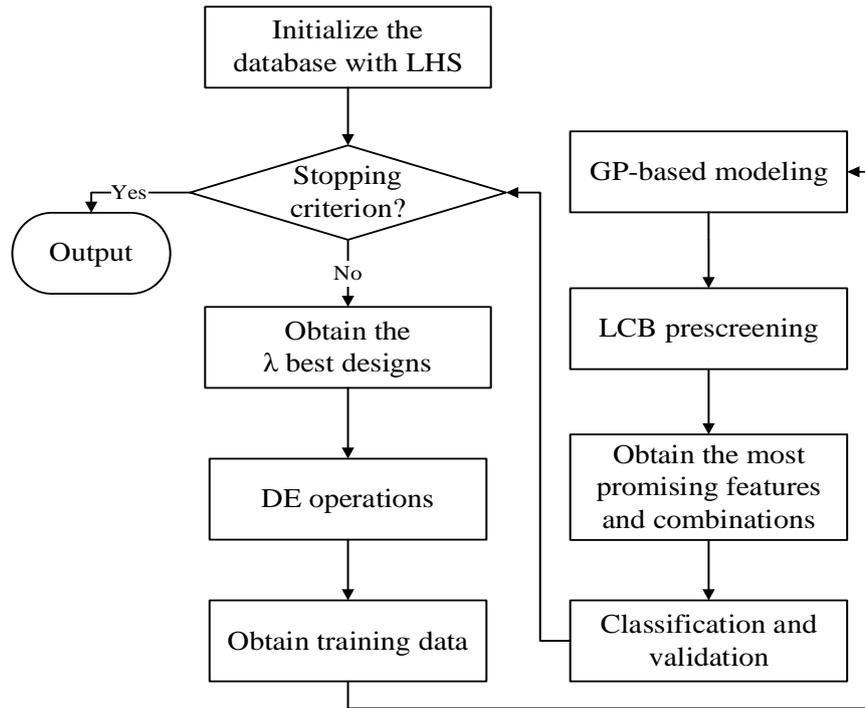
Figure 36 *Hierarchical structure of the solution algorithm.*

5.4.3 Results of Optimization

The entire AI-driven holistic HAR system optimization was implemented using a workstation with an AMD Ryzen 9 3900X 12-core processor (3.8GHz) and an NVIDIA GeForce GT 710 GPU. In this study, binary parameters are treated as continuous variables with values from 0 to 1 and rounded off to the nearest integers. A reference design was based on our previous experiment [206], produced by the adaptive thresholding method with parameter sweeping on adaptive thresholding T_e and feature fusion of six domains. The reference performance in this experiment is based on the human operator achieved classification error rate – manually tuned results - which is based on the operator's domain knowledge.

5.4.3.1 SVM classifier with Adaptive thresholding for SADEA-I

The optimization convergence trend is shown in Figure 37. The red dotted line represents the manual design result, where the error rate is 21.98%. After 2000 function evaluations using SADEA-I, the system is optimized and achieves an error rate from 16.50% to 10.59% in three independent optimization runs. Compared to the manually tuned reference system, the holistic approach decreases the error rate by 5.48% to 11.39%.

Table 28 illustrates the details of the optimal parameters and combinations when the optimization algorithm converges. The result shows that for our human activity dataset, the error rate reaches the minimum with the highest overlapping factor and minimum time window length. In addition, the clipping time is 4.58 s and the adaptive thresholding parameters V and T_e are 0.9166 and 3, respectively. These values were in line with the general trend found in a parametric sweeping on the different domains which identified T_e to be optimal between 0 and 5 of the adaptive threshold value for $V = 0.1$. The effectiveness and necessity of research on thresholding are demonstrated in an improvement of ~11% using the SADEA-I method in only 2000 function evaluations (3 minutes per evaluation). All MDS representations were selected but Domain 1. Domain 1 (Unfiltered MDS Amplitude) is excluded from the most suitable combination, which identifies that the "raw" spectrogram is not optimal and the research on adaptive thresholding and its application in different MDS representations is essential as they all contain salient information for classification although more difficult to interpret visually.
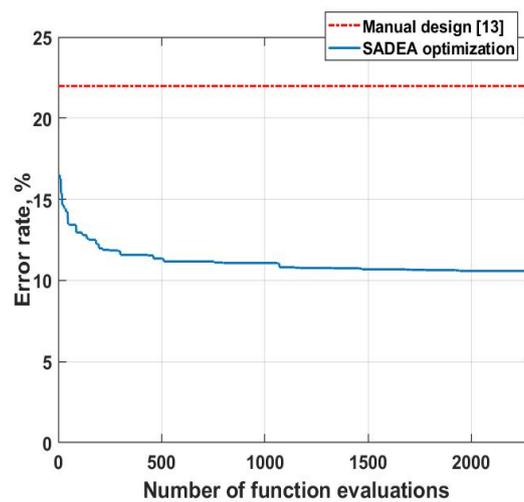
Figure 37 *Convergence trend of SADEA-I with SVM and comparison with manual design results.*

Table 28 Tuned parameters and selected domains

Time window Length	Overlapping Factor	Clipping Time	Difference Value V
154 ms	0.95	4.58s	0.9166
Adaptive Thresholding Te	**FFT Window type**	**Selected Domains of Radar**	
3	Hamming	Domains 2 to 6	

5.3.4.2 AlexNet with Adaptive thresholding for SADEA-I

For this experiment, the dataset was partitioned into three subsets: training, validation, and testing, with a ratio of 7:2:1, respectively. The performance of our experiment is

evaluated on the basis of the testing set results. In alignment with the previous 10-fold cross validation results, we also carried out this experiment 10 times, considering the mean accuracy of iterations as our final performance of the algorithm.

Figure 38 demonstrates the optimization convergence trend of SADEA-I with AlexNet. It also presents the result of manual design parameters on AlexNet. When combined with AlexNet, the SADEA-I is optimized in 700 function evaluations, with an average error rate decrease from 17.83% to 6.46%. Compared with manual design result with an error rate of 9.14%, the holistic approach can yield a worse result with an extra 8.69% error rate or decrease the error rate by 2.68%. Table 29 illustrates the details of convergence trend, along with the optimal parameters and combinations realized at the point of convergence of the optimization algorithm. The result is slightly different from the previous findings. The Hamming window continues to present the most effective option, yielding superior accuracy. The time window length has escalated to 390 ms. The adaptive thresholding parameters, V and T_e, are determined to be 0.8249 and -13, respectively. The V changes to a minor but close value, while T_e has exhibited significant alteration. This result indicates the success of our adaptive thresholding algorithm.
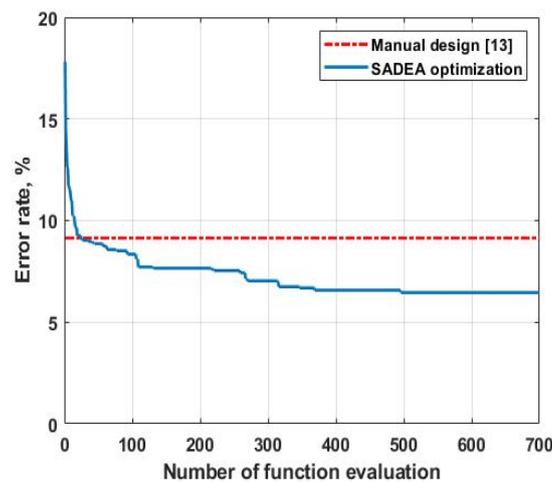
Figure 38 *Convergence trend of SADEA-I with AlexNet and comparison with manual design results.*

Table 29 Tuned parameters and selected domains

Time window Length	Overlapping Factor	Clipping Time	Difference Value V
390 ms	0.87	5 s	0.8249
Adaptive Thresholding Te	**FFT Window type**	**Selected Domains of Radar**	
-13	Hamming	Domains 1 to 5	

5.4.4 Discussion

By leveraging the capabilities of SADEA, we ensure that the algorithm iteratively explores and evaluates a wide range of potential values for v, ultimately converging on the value that maximizes the performance of our system. This evolutionary approach not only automates the optimization process but also enhances the robustness and efficacy of the solution by adapting to the intricacies of the problem space, thereby finding the best possible value for v that aligns with the objectives of our holistic optimization strategy.

Compared with SADEA-I and SVM results, SADEA-1 with AlexNet achieves further improvement of approximately 5% over merely 700 function evaluations, which reduces more than half of the evaluations to convergence compared to the SVM implementation. All radar representations were chosen, including domain I, which highlights the enhanced interpretative capability provided by the AlexNet model. However, there is a significant change observed in the length of the time window. This variance could potentially be attributed to the influence of a higher Doppler frequency resolution, which results in a more intricate graphical representation of the data. The finer resolution provides enriched data for the automatic extraction of features within the deep learning model, thereby augmenting its ability to identify distinctive patterns.

Table 30 concludes the comparison of statistics of SADEA algorithms paired with different classifiers. For SADEA-I with SVM, the optimized error rate decreases from 16.50% to 10.59% with 2000 function evaluations, compared to 21.98% from a human operator. This means that SADEA-I with SVM can boost the performance of our classification by 5.48% to 11.39% in 4 days compared to 3 months of manual tuning by the human operator. The second method SADEA-I with AlexNet exhibits an error rate reduction from 17.83% to 6.46% with 700 function evaluations, compared to 9.14% from a human operator. We observe that the error rate with SADEA-I with AlexNet can reach a minimum error rate of 6.46% which is 4.13% better than SADEA-I with SVM. However, the maximum error of SADEA-I with SVM is 1.33% better than SADEA-I with AlexNet. These statistics indicate the efficacy of the SADEA-I algorithm when interfaced with various classifiers. Deep learning models tend to be more computationally intensive, but with SADEA, it appears that the number of function evaluations are drastically reduced by 2.85 times in this case. This suggests that the SADEA algorithm boosts convergence towards optimal solutions

more efficiently when combined with deep learning algorithm, thus leading to potentially improved performance with fewer evaluations. However, note that the error rate for SADEA-I with AlexNet is more volatile (11.37%) than with SVM (5.91%).

The comparative analysis reveals that the CNN achieved an accuracy of 90.86%, surpassing the SVM which recorded an accuracy of 78.02%, marking a significant difference of 12.84%. This differential initially highlights the superior performance of CNN over SVM in this context. However, the implementation of our proposed optimization strategy yields noteworthy improvements: the optimized CNN accuracy escalates to 93.54%, while the SVM accuracy significantly increases to 89.41%. Consequently, the performance gap between SVM and CNN narrows from 12.84% to a mere 4.13%. This substantial reduction in the disparity not only underscores the effectiveness of CNN but also demonstrates the remarkable potential of our holistic optimization approach to enhance the performance of both models, particularly emphasizing the enhanced capability of SVM when subjected to strategic optimization. This outcome suggests that while CNN inherently offers a robust framework for accuracy, the gap in performance relative to SVM can be considerably minimized through targeted optimization techniques, thereby illustrating the versatility and adaptability of SVM in achieving competitive accuracy levels.

Moreover, SADEA-I with SVM did not incorporate feature selection to enhance performance at this stage. Feature selection can reduce the overall complexity and improve the performance when combined with the adaptive thresholding method [207]. As we continue to refine our holistic techniques, feature selection will be included in our future plan. However, SADEA-I could not handle this many variables to tune at once. The next version of SADEA will be able to handle more tuning parameters for further optimization.

Table 30 Statistics of the best function values obtained by SADEA with different classifiers.

Method	Best error rate (%)	Worst error rate (%)	Human error rate (%)	Function Evaluations
SADEA-I with SVM	10.59	16.50	21.98	2000
SADEA-I with AlexNet	6.46	17.83	9.14	700

5.5 Summary

The rationale for favouring a non-deep learning method, specifically SVM, over deep learning approaches, despite the latter's potential for superior performance with

adequate tuning, is multifaceted. This decision is grounded in considerations of the nature of the data, the computational cost, and the specific requirements of the study, which aims to classify activities based on radar signal features.

Deep learning models are known for their computational intensity, requiring substantial hardware resources and energy consumption, especially during the training phase [208], [209], [210]. This is further exacerbated when tuning the models to achieve optimal performance, a process that involves extensive experimentation with various architectures, hyperparameters, and training datasets. In contrast, SVMs are computationally more efficient, requiring less intensive hardware resources and shorter training times. This efficiency makes SVMs particularly appealing for applications where computational resources are limited or where rapid deployment is necessary [211], [212].

Moreover, the feature selection process undertaken in this study exemplifies the strategic tuning of the SVM classifier [213], [214]. By meticulously selecting and engineering features from the radar signals that are most indicative of the activities of interest, the study enhances the SVM's ability to discriminate between different activities. This process of feature selection not only improves the classifier's performance but also reduces the model's complexity, further contributing to the computational efficiency of the non-deep learning approach.

In conclusion, the choice to employ a non-deep learning method over deep learning is motivated by the specific challenges and requirements of activity recognition from radar signals. The ability of SVMs to handle variable-length data, combined with their computational efficiency and the targeted feature selection process, makes them a compelling choice for this application. This approach ensures that the study can accurately classify a range of activities with varying durations, while also addressing practical considerations of computational cost and resource availability.

6. A Short-range personnel recognition using Radar Signals with LSTM

Coming to the end of the thesis and after having overviewed the many challenges and solutions for activity recognition, the research interest has been progressing towards using edge artificial intelligence (AI) techniques, with deep neural networks becoming more and more prominent. This chapter mirrors this trend as the process of how automatic feature generation methods are utilized for personnel recognition by generating salient features for the classification of human gait sequences is reviewed. Utilising a temporal neural network design, specifically LSTM, a spectrogram signature composed of sequences of six participants is classified without handcrafted feature extraction. The main contributions of this chapter are as follows:

- We evaluate the performance of phase information and range profile compared with the traditional methods which only uses the magnitude information of spectrogram.

- We propose a novel and robust human recognition approach using the combination of spectrogram and range-time domain with both magnitude and phase information.

This work has been previously covered in the publication [39] and the remainder of this chapter is organized as follows. Initially the discussion on the motivations for using LSTM networks is conducted, followed by the description of the experimental setup, data collection, and data pre-processing. Section 6.2 and 6.3 presents a description of the algorithms used and the results obtained with LSTM networks used for and offers some insight on optimizing performances. Finally, section 6.4 concludes the chapter and outlines possible future avenues of research.

6.1 Training/Testing Set Composition and Learning

In this experiment, the data collection introduced in Chapter 4. We only consider the walking activity for the personnel recognition problem. Each walking data is a 10-second-long recording, and each participant repeats it three times. We randomly chose six adults (labelled C1 to C5 and aged between 21 to 60) from the participants' pool to compose the dataset for this experiment. An additional older person (labelled C6 and aged over 60) is also considered to increase the diversity of the dataset. Due to the limitation of the number of samples, the original samples are processed with data

augmentation to expand the size. The original data is cut using a sliding window with a fixed duration of 1 second. It starts from 0s to 1s, and then shifts in time of 0.1s each step. For instance, the second cut is 0.1s to 1.1s, and the third cut is 0.2s to 1.2s. By using the approach, the total number of samples expands from 18 to 1638.

The motivation of the pre-processing of the raw signal data is to generate essentially low noise data for further application. For the raw radar signal, the processing steps are followed. Firstly, a 128-point Hamming-window is proposed to reduce the sidelobes in range-bin. Then, a Fast Fourier Transform (FFT) method is applied to the raw data matrix to convert it into Range-Time domain, which is also known as the high range resolution profile. Next, a high-pass Butterworth notch filter with cut-off frequencies at 0.0075 Hz is utilized to remove static clutter caused by stationary objects such as furniture and walls. After that, Short-Time Fourier Transform (STFT) is implemented with a 0.2 s Hamming window with 95% overlap on the Range-Time data to generate micro-Doppler signatures.

The utilization of only six subjects in this research is primarily justified by its pilot study nature, allowing for the preliminary testing of the methodology and identification of areas for improvement in a controlled setting. Limited resources, such as time, funding, and access to specialized equipment or expertise necessary for radar data collection and analysis, further necessitate a smaller sample size. Additionally, a more homogeneous group of subjects helps mitigate variability in the early stages of method development, crucial for studies where individual differences could significantly influence outcomes. The complexity and time-intensive nature of radar signal processing and analysis also favour a smaller sample size, enabling thorough in-depth analysis and methodological refinement.

However, utilizing a small sample size of only six subjects introduces significant limitations to the study, primarily affecting the generalizability of the findings to a broader population, given the difficulty in ensuring the results are representative across varied individual characteristics such as age, gender, and physical fitness. This limitation is compounded by a reduction in statistical power, heightening the risk of not detecting existing effects, and questioning the methodology's efficacy on a larger scale. Moreover, the small sample size does not adequately test the methodology's robustness against the wide spectrum of inter-individual variability, essential for real-world applications. Additionally, there is a heightened risk of model overfitting, where the methodology might perform well for the small, specific sample but fail to generalize to new, unseen data, thereby limiting its practical applicability and reliability in broader contexts.

6.2 Motivation for personnel recognition and temporal classification networks.

The non-deep learning methodology was initially developed and applied to the domain of personnel recognition, with a specific focus on evaluating its efficacy in recognizing individuals based on their gait patterns. This approach, detailed in Chapter 5, employs a suite of non-deep learning techniques, culminating in the utilization of a SVM classifier to distinguish between different personnel based on gait analysis. Upon implementation and testing of these methods, the resultant overall accuracy achieved was 62.5%. This level of performance, while indicative of the method's potential, falls short of the desired benchmarks for reliable personnel recognition. The modest accuracy underscores the challenges inherent in capturing the nuanced variations in gait patterns with non-deep learning approaches and an SVM classifier.

Table 31 Personnel recognition with proposed non-deep learning methods.

	C1	C2	C3	C4	C5	C6
Accuracy	70.2	58.7	67.5	64.8	61.4	52.1

Given the less than satisfactory outcome achieved with the initial strategy, the research direction pivoted towards exploring the capabilities of deep learning technologies. In the past few years, a series of techniques have been proposed for personnel recognition in order to enhance public security, where most of the approaches are based on optical devices [215] and biometric technology [216]. However, vision-based method and biological features have their own limitations. For optical devices, there are possible invasion of privacy and disputes over image rights. People may feel their privacy is violated when their whereabouts are monitored by a camera all the time. Also, the performance degrades highly when the field of view is narrow, and in adverse lighting conditions. On the other hand, biological features, such as fingerprint or retina scans, are also highly private and require the compliance of people, which cannot always be taken for granted. Radar has potential advantages over the sensors mentioned above, making it a relevant technology in personnel recognition.

Typically, radar-based personnel recognition uses gait analysis from spectrograms [147], [217], [218]. Human gait can provide clear and detailed micro-Doppler signatures of different people. The recognition and classification based on the micro-Doppler signatures are generally performed by extracting hand-crafted features, such

as bandwidth and Doppler mean speed. However, the performance of classification or recognition based on the features are highly dependent on the robustness of those features. However, the traditional feature extraction methods based on experience and statistical characteristics still have many limitations in capability and flexibility, which limits the achievable accuracy with the spectrogram. Deep learning methods are therefore introduced to address the issues.

Deep learning has become a popular research topic in radar fields because it can automatically extract salient features from radar data. It aims to find the mapping relationship between the training data and labels through supervised and testing of a large number of samples. Compared with the traditional hand-picked features, using deep learning technologies can achieve a higher accuracy of classification. Rather than akin to images, these radar data are akin to sequences of temporal signals, as well. Due to this reason, the recurrent neural network architectures used in the work in the audio/speech processing community were taken as inspiration and explored in this work. Specifically, the focus was on the Long Short-Term Memory (LSTM) networks.

The LSTM was already adopted in [102], [151]. The main property of the LSTM is its memory capability to capture long-term dependency between data separated by a significant duration [113]. This is relevant in speech, where two correlated words can be separated by other words (e.g., auxiliary verb and past participles in Germanic languages, nouns, and adjectives where many adjectives are utilised). Radar data, therefore, can resemble speech as different actions performed at different time steps are correlated by human kinematics (e.g., an individual can stand up only after sitting down, but a variable duration can separate these two actions). However, the difference between radar data and speech or audio data arises because they do not encode any kinematic information or constraint, which usually are instead the main feature of the radar data and what radar-based classification algorithms aim to utilize.

LSTM networks have been chosen for radar data personnel recognition over other potential alternatives such as RNN, Gated Recurrent Units (GRU), Transformers, and CNN due to several compelling reasons that align with the specific requirements and challenges of radar signal processing for personnel recognition tasks. Each of these alternatives has its strengths, but LSTMs offer a unique combination of features that make them particularly well-suited for this application:

Advantages of LSTM Over RNN [219], [220], [221], [222]: While RNNs are theoretically capable of handling long-term dependencies, in practice, they struggle due to the vanishing gradient problem. LSTMs address this issue with their

specialized architecture, including memory cells that can maintain information over longer sequences, making them more effective for radar data sequences where the relevance of information can span across long time intervals. Besides, LSTMs are designed to better capture temporal dynamics, which is crucial for recognizing patterns in radar data that evolve over time, such as the movements of personnel.

Advantages of LSTM Over GRU [222], [223], [224]: Although GRUs are a simplified version of LSTMs that can perform comparably in many tasks with fewer parameters, LSTMs provide finer control over the memory. This is due to their additional gate (the output gate), allowing for a more nuanced handling of information flow. This complexity can be beneficial in radar signal processing, where the distinction between relevant and irrelevant temporal features is critical.

Advantages of LSTM Over Transformer [115], [116], [117]: Transformers, despite their success in various domains, primarily excel in parallel processing and require significant computational resources. LSTMs, on the other hand, are inherently sequential, making them more naturally aligned with the time-series nature of radar data. Additionally, LSTMs can be more data-efficient and require less computational power, making them suitable for scenarios where real-time processing is essential, and resources may be limited.

Advantages of LSTM Over CNN [209], [225]: CNNs excel in extracting hierarchical spatial features and are predominantly used in image processing tasks. In contrast, LSTMs are tailored for sequential data, making them more adept at capturing the temporal dynamics essential for personnel recognition in radar data. While CNNs can process time-series data by treating time as a spatial dimension, they may not capture long-term dependencies as effectively as LSTMs.

In summary, the choice of LSTM for radar data personnel recognition is motivated by its superior ability to handle long-term dependencies, its efficient processing of sequential time-series data, and its nuanced control over memory. These characteristics make LSTMs particularly adept at capturing the complex temporal patterns in radar signals associated with human movement, providing a robust framework for accurate personnel recognition.

6.3 Experiment Results and Performance Analysis

In this section experimental results using different LSTM network architectures are provided, together with discussions on changes in performances due to the format of

input data used (e.g., spectrograms vs range-time plots), and on significant hyperparameters of the networks (e.g., learning rate).

CNN-based architectures do not include the memory unit. Hence, the network processes each window of the spectrogram as independent inputs. This may cause much overlap when the time interval is small. The response of RNN-based structure to new data is decided by the current and the past input, which acts on the memory of the network. When the time interval is small, it can feed small pieces of the spectrogram into the network saving on computational load compared to CNN.

Overfitting is a problem that often happens in deep learning applications due to the robust learning ability of neural networks which only focus on training data. This negatively impacts the result when the network processes new or unseen testing data. This problem usually occurs when the database is small, or the model is complex, which is the case here.

Different methods have been proposed to prevent overfitting problems [226]. For example, the early-stop method can stop the learning process when the performance begins to degrade on the validation set. In this experiment, the dropout [121] method is proposed to address the overfitting problem. Dropout means dropping out units, which abandons a part of the output randomly in one layer, yielding an improved generalization.

To improve the performance, a hybrid information fusion method, which is the combination of neural network fusion and a soft fusion at the decision level, is then considered. The architecture of the network is shown in Fig. 1. It consists of two parts, a feature extraction network and a fusion network including a deep fusion part and a soft fusion part. The feature extraction network contains the Range-LSTM and the Doppler-LSTM, which are both composed of two LSTM layers extracting temporal features from magnitude and phase separately. Afterwards, the temporal features from both magnitude and phase information in the same domain are aggregated, and then a series of dense layers are integrated with a SoftMax classifier to generate the prediction of class for each domain. Finally, a soft fusion method is employed to combine the outputs of the previous networks to improve performances.

Soft fusion [100], [227] aims at generating the new prediction of classes by combining the scores which are generated in the last layer of network with SoftMax activation. In the SoftMax layer, the classifier generates a scoring matrix with regard to the posterior probability, which represents the confidence level. The class with the highest probability will be chosen as the output class. The following equation illustrates how

the combination works mathematically, where W_D is the weight of the fused Doppler network, and W_R is the weight of the fused range network. S_D and S_R are the score matrix of the fused Doppler network and fused range network, respectively.

$$S_F = W_D \cdot S_D + W_R \cdot S_R \tag{6.1}$$

In the first experiment, we investigate the performance of both networks using magnitude and phase separately. The Range-LSTM and the Doppler-LSTM networks are investigated with a 1310 samples training set (80%) and a 328 samples validation set (20%), where those datasets are randomly picking data from the entire database, as mentioned before. The network structure used in this experiment is a two-layers LSTM, with 128 neurons in each hidden layer, and the dropout probability between two LSTM layers is 0.6 for the Doppler-LSTM and 0.5 for the Range-LSTM. The output of LSTM layers is passed to fully connected (FC) layers. The first FC layer uses ReLU as the activation function due to its low computational cost. The second FC layer uses the same activation function. Then, a SoftMax layer is connected to the second FC layer since it can output the final labels. The block diagram of the network is shown in Figure 39.
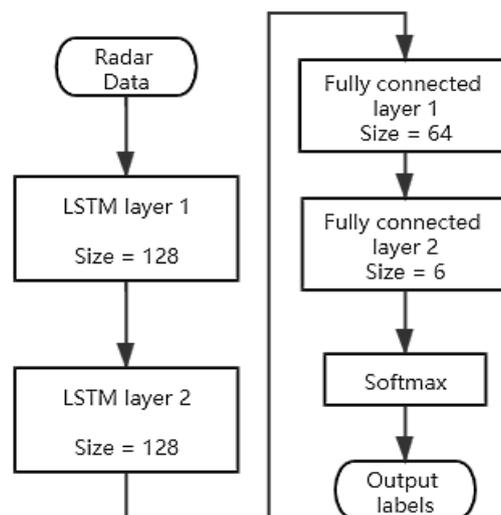
Figure 39 *The LSTM architecture for the recognition.*

The networks are trained in 200 epochs using magnitude and phase separately, with the Adam optimizer and fixed initial learning rate of 0.001. Figure 40 and Figure 41 demonstrate the loss curves as a function of epochs. The validation accuracy is illustrated in Table 32 and the training time consumption is shown in Table 33.

The result shows that, in both the spectrogram and range-time domain, using phase information can accelerate the convergence of the network compared with traditional

methods using amplitudes. In Doppler-LSTM, both training and validation of phase information converge within 50 epochs. To the magnitude information, both training and validation converge at around 100 epochs, which is approximately twice longer than for the phase. In range-LSTM, the convergence finishes in a short time, which is ~40 epochs for phase and ~75 epochs for magnitude. In addition, the loss of both remains at an acceptable range at the end of the process, which means the dropout method succeeds, and the network limits overfitting problems.
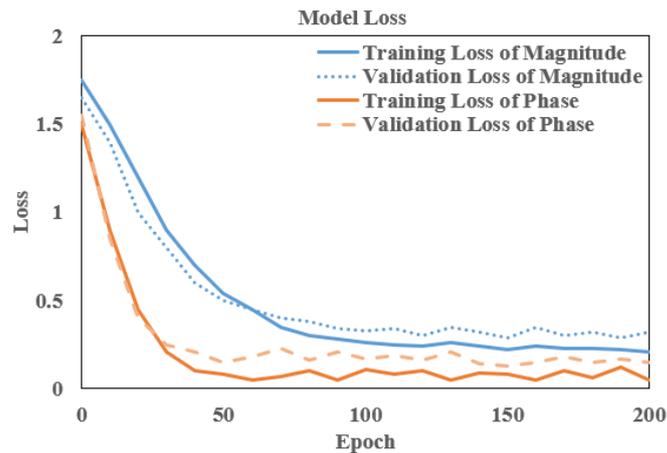
Figure 40 *Loss evaluation of the Doppler-LSTM.*
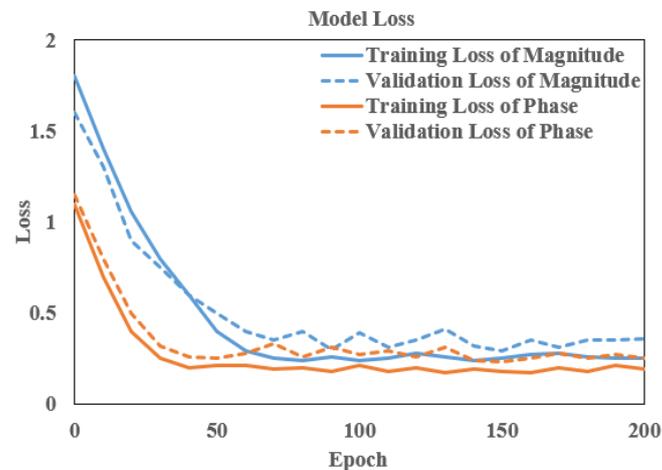
Figure 41 *Loss evaluation of the Range-LSTM.*

In the second experiment, a stratified ten-fold cross-validation approach is used to assess our approach. Compared with the normal k-fold cross-validation, the stratified one extracts the validation set in terms of the ratio of classes, which makes the validation more comprehensive. To further improve the accuracy of the recognition, the hybrid information fusion method is employed, where the phase and magnitude information from the same domain are fused using deep fusion methods at the first stage, and then the results are used for the second stage with soft fusion. The hyperparameters of the networks remain the same as the first experiment, and the

network is still trained with 200 epochs. The weight ratio between the Doppler-LSTM and Range-LSTM sets from 1:5 to 5:1, to appraise the performance of this method. The result is shown in Figure 42.
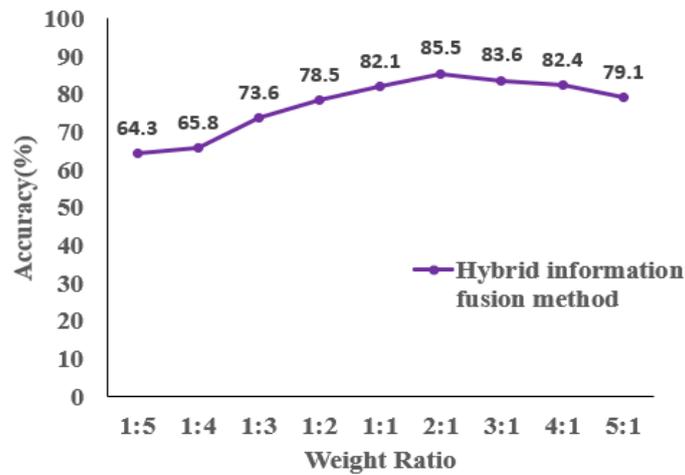
Figure 42 *Hybrid solution evaluation with different weight ratios.*

The Figure 42 shows that the accuracy reaches 85.5% for the hybrid information fusion methods when the weight ratios ($W_D : W_R$) are 2:1.

To further analyze the influence caused by the phase information and the differences between the two domains, the neural network fusions of Range-LSTM and Doppler LSTM are independently used. The results are shown in Table 32 and the time consumption is demonstrated in Table 33. The experiment shows the Doppler-LSTM performs better than the Range-LSTM, with both higher average accuracy (76.0% for the Doppler and 64.7% for the range-time) and better results in the fusion performance (81.9% and 71.0%, respectively). Generally, both networks achieve outstanding results in distinguishing C6. One possible reason is that the cycle of walking of the oldest person is longer than the younger adults, which means they have a smaller and slower pace. In addition, the extent of the body motion for the aged person is smaller than for younger adults, which results in an easily distinguishable micro-Doppler signature, leading to the recognition of that person easier. The performance of the Range-LSTM is not satisfying. In our perspective, the information contained in the range-time domain, which is the relative location from the target to radar, is not adequate for personnel recognition. Besides, the poor performance of the range-LSTM is possibly due to the low radar bandwidth. A higher bandwidth could result in a more satisfactory range resolution (range resolution < 10cm), which gives rise to the better performance. Another possible reason for the unsatisfactory result is that the size of the database is too limited to provide satisfactory performance. The recording of each subject is only 30 s, which is not enough to provide an adequate number of samples.

The performance could be better with a more extensive database in terms of longer recording for each subject.

Table 32 Comparison of accuracy for human recognition using Doppler-LSTM, Range-LSTM and Hybrid mode.

Accuracy (%)	C1	C2	C3	C4	C5	C6	Overall Performance
Magnitude of spectrogram	75.4	73.5	76.9	77.3	71.2	95.7	78.3%
Phase of spectrogram	67.5	72.4	68.6	73.8	74.3	85.2	73.6%
Magnitude of range-time	70.5	62.7	55.8	61.4	67.9	87.1	67.6%
Phase of range-time	64.8	51.5	62.9	70.4	53.8	66.7	61.7%
Hybrid information of spectrogram	80.7	78.1	77.5	81.5	79.3	94.1	81.9%
Hybrid information of range-time	64.1	74.7	73.1	70.5	65.5	78.3	71.0%
Hybrid information of two domains fusion	81.6	79.7	84.2	87.8	83.2	96.5	85.5%

Table 33 Total Training time consumption

Network	Time Consumption for 200 epochs (second)
Doppler-LSTM with magnitude	1483
Doppler-LSTM with phase	964
Range-LSTM with magnitude	1038
Range-LSTM with phase	702
Hybrid information of spectrogram	1221
Hybrid information of range-time	865
Hybrid information of two domains fusion	1892

Time consumption is shown in Table 33. It is obvious that phase information improves computation time by ~35.0% in Doppler and ~32.4% in Range profile compared to the magnitude information. Besides, the time consumption of the Range-LSTM is generally lower than the computation time of Doppler-LSTM, which is lower by ~30.0% using magnitude information and ~27.2% using phase information. To both the spectrogram and range-time domain, the training time is improved when the phase information is fused with magnitude information, compared with the

independent use of magnitude information. The hybrid information fusion improves the performance of the recognition, which achieves the best result with the accuracy of 85.5%. However, in the multi-domain fusion, the computation time increases. One possible reason is that two deep fusion networks are implemented at the same time, which largely increases the computational load of the processor, leading to the degradation in computing speed.

6.4 Summary

In this experiment, we proposed the use of the recurrent neural network on both the spectrogram and range-time domain of radar signal for the identification of individual subjects. The preliminary experiment results show that without a hybrid information fusion, the Doppler-LSTM and Range-LSTM can achieve accuracy of approximately 78.3% and 67.6%, respectively. When the hybrid information fusion is applied in each domain, whereby the magnitude information is fused with the phase information, the accuracy can reach up to 81.9% and 71.0%. Meanwhile, the computational speeds are also improved by ~17.7% and ~16.7% compared with using magnitude information alone in Doppler-LSTM and Range-LSTM, respectively. This suggests that the phase information of radar signal is as effective as the magnitude information, with better convergence rate. On the other hand, the combination of phase information and magnitude information can reduce the rate of abnormal prediction and improve accuracy. When the Hybrid information fusion continues to be applied in multi-domain level, the performance improves to 85.5%, but the computational time increases largely. This might be caused by the simultaneous running of two neural network fusion, which places additional computing burden to the processor.

For future work, further advanced Range-LSTM and Doppler-LSTM scheme will be carried out, with other types of the layer such as Bi-LSTM layer. Besides, the phase information used in this experiment is wrapped. The performance of the algorithm with wrapped vs unwrapped phase information should be investigated to determine whether this can improve accuracy. Also, designing an adaptive algorithm for fusing Range-LSTM and Doppler-LSTM could have the potential to improve performance. Furthermore, the complex number can be directly used as input in the neural network, and thus we can use one network with complex numbers of radar signal instead of two separate networks. Additionally, a thresholding method is proposed to upgrade the performance of phase information, which would facilitate network training by focusing more on regions of interest in the phase data. The dataset in this experiment is still limited, and thus more data from different participants repeating the same

action should be collected, including different aspect angles with respect to radar and various measurement environments.

7. Conclusion and Future Work

Radar technology, traditionally associated with defence applications, has seen a significant expansion in civilian domains [228], [229], particularly in activity recognition and monitoring of living targets. This shift is evident in the growing body of research and the increasing market applications of radar beyond its conventional uses. This thesis provides a comprehensive overview of the evolution of radar technology in human activity recognition, highlighting its journey from inception to the current state of advanced applications.

This research, validated through experiments at the University of Glasgow CSI Lab, focused on classifying human behaviours, identifying individuals using radar data integrated into different frameworks, and advancing beyond traditional spectrogram-based analyses. Machine learning classifiers trained on this data can distinguish activities, like falls, by their unique patterns. This thesis evaluates various classification methods for their robustness and efficiency in recognizing human movements, with radar offering privacy advantages over vision-based methods by tracking movement trajectories without creating identifiable images. Micro-Doppler signatures, derived from time-frequency analysis of range-time maps, are pivotal in characterising micro-motions during activities, aiding in differentiating behaviours and detecting presence. The thesis also explores radar signal processing, discussing the impact of various parameters on performance and comparing radar technologies in AAL applications. This work highlights the potential and challenges of radar in human activity recognition, setting a foundation for future advancements in the field.

In this chapter, the contributions to knowledge are summarised looking back at the research questions set at the beginning of this thesis, with limitations and future work proposed. The future works are also provided to guide and inspire people. The future directions outlined in this thesis aim to inspire and guide further research in the field of radar technology and recognition systems.

7.1 Summary of Contributions

The research presented in this thesis addresses two primary inquiries, pivotal to the advancement of machine learning (ML) applications in activity recognition, particularly in the context of radar technology.

The Impact of Subject Physiology on Classification Accuracy: Chapter 4 delves into the influence of subject physiology on classification rates, a critical aspect often overlooked in ML applications. This investigation aligns with the contemporary trend in the ML community that emphasizes the significance of input data over other factors like algorithm selection or hyperparameter tuning. By assessing the effects of age, gender, and physiological diversity on activity detection accuracy, the research addresses a crucial question: how do these variations in the target population impact the efficacy of hand-crafted feature extraction and, consequently, the performance of classification algorithms? This exploration not only contributes to a deeper understanding of the subject matter but also guides the development of more inclusive and robust activity recognition systems.

Exploration Beyond Spectrogram-Based Radar Classification: The second inquiry, discussed in Chapters 4 and 5, examines the limitations of traditional spectrogram-based radar classification and explores innovative alternatives. The introduction of diverse radar representations marks a significant advancement in activity recognition. By employing a multi-domain approach, integrating features from various domains through feature selection and fusion, the research demonstrates a notable improvement in activity recognition accuracy. This approach not only optimizes classification accuracy in assisted living contexts but also highlights the potential of radar technology in broader applications. The use of masking pre-processing techniques to focus on relevant information in spectrograms further enhances classification rates, achieving up to 93.1% accuracy. This segment of the research showcases the transformative potential of radar-based Human Activity Recognition (HAR), expanding its applicability and generating renewed interest in the field.

Optimization of the Proposed Model (Chapter 5) and Exploration of Gait Analysis (Chapter 6): In Chapter 5, the thesis also discusses the optimization of the proposed model through engineering approaches. However, recognizing the time-intensive nature and limited scope of improvement through these methods, the research went towards more robust tools like Evolutionary Algorithms (EA). The implementation of SADEA-I, for instance, allows for a more comprehensive optimization, encompassing not just the radar domains but also parameters within the Short-Time Fourier Transform (STFT) process, such as overlapping factor and time window length. The application of different classifiers, including SVM and AlexNet, validates this holistic optimization approach, demonstrating significant improvements in accuracy.

Additionally, the thesis explores personnel recognition through gait analysis in Chapter 6, an essential aspect of identity verification. By employing temporal classification algorithms, typically used in audio AI, the research achieves improved recognition rates without the need for explicit feature extraction. This approach leverages the similarities between radar and audio signals, both of which represent physical movements as spectral information. The use of Long Short-Term Memory (LSTM) based classifiers on spectrograms containing activity sequences enables the classification of complex signatures, encompassing micro-Doppler (mD) variations of gaits and range-time maps. This comprehensive analysis culminates in presenting the optimal performance of a single radar sensor for personnel recognition, based on realistic gait sequences, without the necessity for predefined feature generation.

In essence, this thesis not only addresses fundamental questions in the realm of radar-based activity recognition but also pushes the boundaries of current methodologies, offering innovative solutions and demonstrating their practical efficacy through empirical validation. The thesis stands out for its holistic approach to addressing the challenges in radar-based HAR, offering innovative solutions that are both theoretically sound and practically viable. The research work significantly advances the understanding and application of radar technology in human activity recognition, setting a new benchmark in the field.

7.2 Limitations

1. Performance Under Varying Noise Conditions

One notable limitation of this work is the absence of testing the model's performance across different Signal-to-Noise Ratios (SNRs). The robustness of machine learning models, particularly in real-world applications, is often gauged by their ability to maintain high accuracy under varying levels of noise. By not evaluating the model across a spectrum of SNR conditions, the study may not fully represent the potential fluctuations in performance that could occur in noisy environments. This oversight limits the generalizability of the findings to scenarios where noise levels are controlled or minimal.

2. Evaluation Across Different Aspect Angles

The study did not explore the impact of different aspect angles on model performance. Aspect angle, the angle between the radar line of sight and the target orientation, can significantly affect the radar cross-section and, consequently, the signal characteristics.

The variation in aspect angles could introduce additional complexities in signal patterns, potentially influencing the model's accuracy. The exclusion of this variable from the analysis restricts the understanding of how changes in target orientation might affect the detection and classification capabilities of the model.

3. Testing for Multiple Human Conditions

Another limitation is the study's focus on a limited set of human conditions, without testing the model's efficacy in recognizing multiple, potentially overlapping human conditions. Human targets in radar data can exhibit a wide range of behaviors, activities, and physiological states, each presenting unique signal characteristics. The failure to test the model against a diverse array of human conditions may overlook its ability to differentiate between complex and nuanced patterns, limiting its applicability in scenarios requiring fine-grained classification of human states.

4. Handling of Unknown Classes

The research did not address the model's response to unknown classes, a critical aspect of machine learning models deployed in dynamic environments. The ability to accurately identify or appropriately handle signals that do not belong to any of the trained categories is essential for practical applications. Without this consideration, the model's reliability, and adaptability in real-world situations, where unforeseen categories are likely, remain untested.

5. Cross Subject Validation

Cross-subject validation involves training the model on data from a subset of subjects and testing it on data from entirely different subjects. This method is crucial for assessing the model's generalizability across individuals with diverse physiological and behavioural characteristics. The primary reason for not utilizing cross-subject validation in our study stems from the initial scope and design of the research, which aimed to explore the feasibility and effectiveness of activity recognition techniques within a more controlled and homogeneous sample. This approach allowed for a focused analysis on the technical capabilities and limitations of the proposed models without the added variability introduced by cross-subject differences.

However, the absence of cross-subject validation presents a notable limitation. It restricts our ability to claim robustness and applicability of the model across a broader population. Without testing the model on unseen subjects, there's a risk that the developed system might perform well on the training subjects but fail to generalize to

new individuals, due to variations in movement patterns, body shapes, and other individual-specific factors. Acknowledging this limitation, future work should consider incorporating cross-subject validation to ensure the model's effectiveness and reliability across diverse subjects. This would not only enhance the model's generalizability but also provide a more comprehensive understanding of its applicability in real-world scenarios, where variability among individuals is a given.

6. Implications of Odd Number of Classes on Hierarchical Classification

Finally, the study did not investigate the implications of having an odd number of classes in hierarchical classification systems. Hierarchical classification, a method used to categorize data into a tree-like structure of classes and subclasses, can be particularly sensitive to the number of classes at each level. An odd number of classes could pose challenges in balancing the classification tree, potentially impacting the efficiency and accuracy of the model. This oversight leaves unanswered questions regarding the optimal structuring of classes for hierarchical classification in similar studies.

In conclusion, while this study makes significant contributions, these limitations highlight areas for future research to explore. Addressing these gaps could enhance the robustness, applicability, and understanding of machine learning models in radar signal processing and beyond.

7.3 Future Trends

1. Co-Learning of Multi-Modality

The concept of multi-modal machine learning, as explored in this thesis, draws a parallel to the human cognitive process of perceiving the environment through multiple sensory modalities. This approach in machine learning seeks to process and correlate information from various data modalities, thereby enhancing the robustness and accuracy of Human Activity Recognition (HAR). Multi-modal learning primarily encompasses two methodologies: fusion and co-learning.

Fusion, as discussed in this thesis, involves the integration of information from two or more domains for both training and validation. This method [230] capitalizes on the collective strengths and capabilities of different data modalities, offering a more comprehensive understanding and interpretation of the data. By amalgamating diverse sensory inputs, fusion-based multi-modal learning can achieve a higher level of

accuracy and reliability in HAR, surpassing the limitations of single-modality approaches [231].

Co-learning [232], on the other hand, is an innovative approach that focuses on the transfer of knowledge between different data modalities. This method explores how insights gained from auxiliary modalities can aid in the learning process of a model focused on a primary modality. The key advantage of co-learning lies in its ability to overcome the constraints of a single data modality by leveraging the strengths of others. Unlike fusion methods, co-learning requires the data from auxiliary modalities only during the training, not during testing. This characteristic is particularly advantageous in scenarios where certain modalities may be unavailable during the testing phase. Furthermore, co-learning can significantly benefit the learning process of a modality with limited data samples by utilizing correlated modalities that possess a richer dataset. This cross-modal knowledge transfer enhances the overall performance and adaptability of the learning model. An excellent example is [233], a knowledge distillation framework was proposed to distil knowledge from a teacher network taking depth videos as input to an RGB-based student network. The knowledge distillation was achieved by forcing the feature maps and prediction scores of the student network to be similar to the teacher network.

2. Expanding on Algorithm and Domains

The research conducted in this thesis, while comprehensive, represents only a subset of the myriad possibilities in the field of radar-based human activity recognition (HAR). Alternative research avenues and solutions, some of which are gaining traction in recent literature, offer intriguing prospects for future exploration.

Another potential research direction involves the exploration of multi-dimensional radar data representations. While this thesis primarily utilizes 2D radar representations, the extension to higher-dimensional data structures, such as 3-D and even 4-D radar information [234], could offer deeper insights and more nuanced understanding of human activities. These multi-dimensional representations can capture additional aspects of the target's movements and environment, potentially leading to more accurate and detailed activity recognition.

The transition to higher-dimensional radar data necessitates the development of advanced processing techniques and algorithms capable of handling the increased complexity and volume of data. This shift could also open up new possibilities in feature extraction, data fusion, and classification methodologies, further pushing the boundaries of what is achievable in radar-based HAR. Our holistic optimization

strategy encompasses the entire process of representation production, feature extraction, selection, fusion and learning process, extending its control to the. This approach not only streamlines the workflow but also ensures consistency and precision in the transformation of raw data into complex, multi-dimensional representations, crucial for advanced analysis and interpretation.
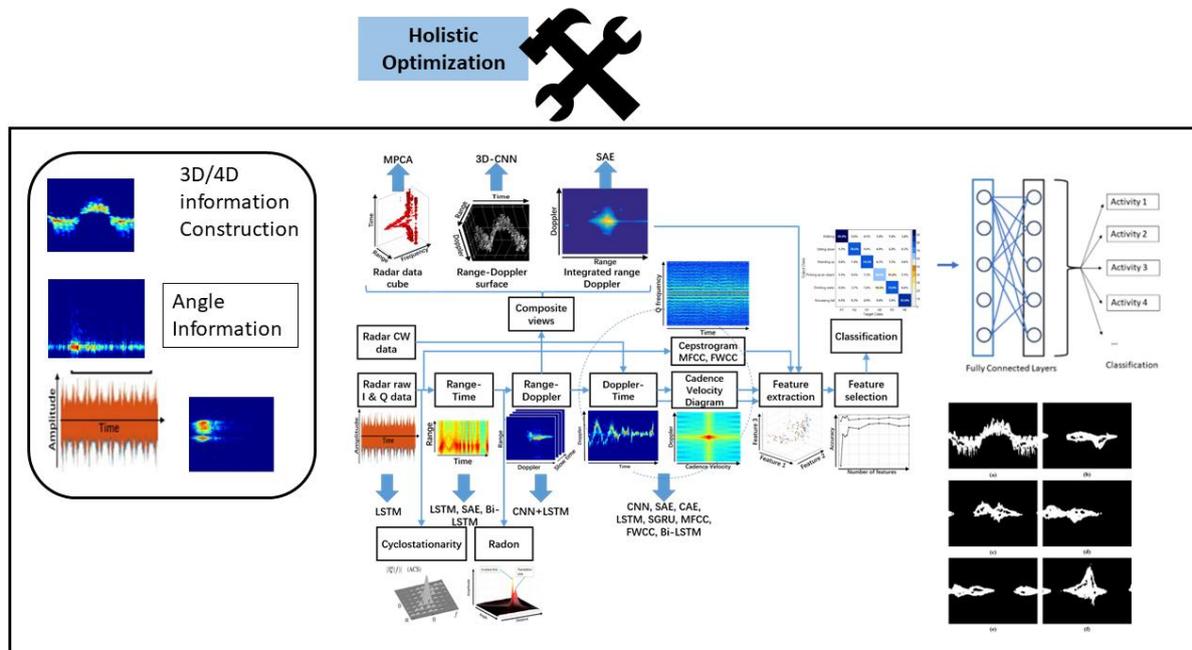
Figure 43 *The target of combination of holistic optimization procedures, controlling every domain generation*

3. Interpretability and Trust for ML methods

A significant hurdle in the broader adoption of ML techniques in HAR is the 'Black Box' problem, particularly prevalent in complex methods like deep learning [235]. The opaque nature of these algorithms, where the decision-making processes are not readily interpretable by humans, poses risks in applications where safety and security are paramount. This lack of transparency can lead to trust issues [236], complicating verification processes and creating legal ambiguities in the event of system failures or accidents. The issue of diversity in datasets is highlighted by our findings in Chapter 4, where distinct variations in machine learning model predictions were observed between datasets of adults and elders. This underscores the necessity of addressing diversity in dataset composition to ensure the robustness and generalizability of predictive models.

As ML methods gain traction in HAR, the need for interpretability becomes increasingly critical. It is essential for users, be they system operators or end-users, to

understand the behaviour of these systems. This understanding not only fosters trust but also opens avenues for user-based control, justification, and performance enhancement [235], [236]. Recent advancements in tools for visualizing the activity of neurons in hidden layers of neural networks offer a glimpse into potential solutions. Such tools can demystify the training processes of neural networks, making them more accessible to non-experts. This approach could serve as a model for making other ML-based methods more interpretable, thereby addressing potential liability issues and enhancing the overall reliability and trustworthiness of HAR systems.

References

[1]   World Health Organization, 'WHO Global Report on Falls Prevention in Older Age', 2007.

[2]   UK government, 'Healthy Ageing', 2022.

[3]   N. D. Centre, 'Help at Home - Use of assistive technology for older people', Dec. 2018. doi: 10.3310/themedreview-03385.

[4]   National Institute for Health and Care Excellence, 'Falls in older people: assessing risk and prevention', *Clinical guideline*, 2013.

[5]   M. Mubashir, L. Shao, and L. Seed, 'A survey on fall detection: Principles and approaches', *Neurocomputing*, vol. 100, pp. 144–152, Jan. 2013, doi: 10.1016/j.neucom.2011.09.037.

[6]   L. Z. Rubenstein, 'Falls in older people: epidemiology, risk factors and strategies for prevention', *Age Ageing*, vol. 35, no. suppl_2, pp. ii37–ii41, Sep. 2006, doi: 10.1093/ageing/afl084.

[7]   M. Terroso, N. Rosa, A. Torres Marques, and R. Simoes, 'Physical consequences of falls in the elderly: a literature review from 1995 to 2010', *European Review of Aging and Physical Activity*, vol. 11, no. 1, pp. 51–59, Apr. 2014, doi: 10.1007/s11556-013-0134-8.

[8]   World Health Organization, 'WHO World Health Statistics 2022', 2022.

[9]   Office for National Statistics, 'Census 2021: Ageing analysis plans', 2021.

[10]   D.-H. Park, H.-C. Bang, C. S. Pyo, and S.-J. Kang, 'Semantic open IoT service platform technology', in *2014 IEEE World Forum on Internet of Things (WF-IoT)*, IEEE, Mar. 2014, pp. 85–88. doi: 10.1109/WF-IoT.2014.6803125.

[11]   I. H. Lopez-Nava and A. Munoz-Melendez, 'Wearable Inertial Sensors for Human Motion Analysis: A Review', *IEEE Sens J*, vol. 16, no. 22, pp. 7821–7834, Nov. 2016, doi: 10.1109/JSEN.2016.2609392.

[12]   C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer, 'Monitoring Activities of Daily Living in Smart Homes: Understanding human behavior', *IEEE Signal Process Mag*, vol. 33, no. 2, pp. 81–94, Mar. 2016, doi: 10.1109/MSP.2015.2503881.

[13]   F. Erden, S. Velipasalar, A. Z. Alkar, and A. E. Cetin, 'Sensors in Assisted Living: A survey of signal and image processing methods', *IEEE Signal Process Mag*, vol. 33, no. 2, pp. 36–44, Mar. 2016, doi: 10.1109/MSP.2015.2489978.

[14]   R. Igual, C. Medrano, and I. Plaza, 'Challenges, issues and trends in fall detection systems', *Biomed Eng Online*, vol. 12, no. 1, p. 66, 2013, doi: 10.1186/1475-925X-12-66.

[15]   S. Z. Gurbuz and M. G. Amin, 'Radar-Based Human-Motion Recognition With Deep Learning: Promising Applications for Indoor Monitoring', *IEEE Signal Process Mag*, vol. 36, no. 4, pp. 16–28, Jul. 2019, doi: 10.1109/MSP.2018.2890128.

[16]   K. Chaccour, R. Darazi, A. H. El Hassani, and E. Andres, 'From Fall Detection to Fall Prevention: A Generic Classification of Fall-Related Systems', *IEEE Sens J*, vol. 17, no. 3, pp. 812–822, Feb. 2017, doi: 10.1109/JSEN.2016.2628099.

[17]   T. R. Bennett, J. Wu, N. Kehtarnavaz, and R. Jafari, 'Inertial Measurement Unit-Based Wearable Computers for Assisted Living Applications: A signal processing perspective', *IEEE Signal Process Mag*, vol. 33, no. 2, pp. 28–35, Mar. 2016, doi: 10.1109/MSP.2015.2499314.

[18]   P. Pierleoni, A. Belli, L. Palma, M. Pellegrini, L. Pernini, and S. Valenti, 'A High Reliability Wearable Device for Elderly Fall Detection', *IEEE Sens J*, vol. 15, no. 8, pp. 4544–4553, Aug. 2015, doi: 10.1109/JSEN.2015.2423562.

[19]   G.-M. Jeong, P. H. Truong, and S.-I. Choi, 'Classification of Three Types of Walking Activities Regarding Stairs Using Plantar Pressure Sensors', *IEEE Sens J*, vol. 17, no. 9, pp. 2638–2639, May 2017, doi: 10.1109/JSEN.2017.2682322.

[20]   S. Tao, M. Kudo, and H. Nonaka, 'Privacy-Preserved Behavior Analysis and Fall Detection by an Infrared Ceiling Sensor Network', *Sensors*, vol. 12, no. 12, pp. 16920–16936, Dec. 2012, doi: 10.3390/s121216920.

[21]   F. Fioranelli, M. Ritchie, and H. Griffiths, 'Classification of Unarmed/Armed Personnel Using the NetRAD Multistatic Radar for Micro-Doppler and Singular Value Decomposition Features', *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1933–1937, Sep. 2015, doi: 10.1109/LGRS.2015.2439393.

[22]   S. Z. Gurbuz, C. Clemente, A. Balleri, and J. J. Soraghan, 'Micro-Doppler-based in-home aided and unaided walking recognition with multiple radar and sonar systems', *IET Radar, Sonar & Navigation*, vol. 11, no. 1, pp. 107–115, Jan. 2017, doi: 10.1049/iet-rsn.2016.0055.

[23]   X. Yang *et al.*, 'Freezing of Gait Detection Considering Leaky Wave Cable', *IEEE Trans Antennas Propag*, vol. 67, no. 1, pp. 554–561, Jan. 2019, doi: 10.1109/TAP.2018.2878081.

[24]   LeadingAge CAST, 'State of technology in aging services.', 2007.

[25]   M. Al-Faris, J. Chiverton, D. Ndzi, and A. I. Ahmed, 'A Review on Computer Vision-Based Methods for Human Action Recognition', *J Imaging*, vol. 6, no. 6, p. 46, Jun. 2020, doi: 10.3390/jimaging6060046.

[26]   Z. Xue, D. Ming, W. Song, B. Wan, and S. Jin, 'Infrared gait recognition based on wavelet transform and support vector machine', *Pattern Recognit*, vol. 43, no. 8, pp. 2904–2910, Aug. 2010, doi: 10.1016/j.patcog.2010.03.011.

[27]   Youngwook Kim and Hao Ling, 'Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328–1337, May 2009, doi: 10.1109/TGRS.2009.2012849.

[28]   J. Le Kernec *et al.*, 'Radar Signal Processing for Sensing in Assisted Living: The Challenges Associated With Real-Time Implementation of Emerging Algorithms', *IEEE Signal Process Mag*, vol. 36, no. 4, pp. 29–41, Jul. 2019, doi: 10.1109/MSP.2019.2903715.

[29]   X. Li, Z. Li, F. Fioranelli, S. Yang, O. Romain, and J. Le Kernec, 'Hierarchical Radar Data Analysis for Activity and Personnel Recognition', *Remote Sens (Basel)*, vol. 12, no. 14, p. 2237, Jul. 2020, doi: 10.3390/rs12142237.

[30]   S. Björklund, H. Petersson, and G. Hendeby, 'Features for micro-Doppler based activity classification', *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1181–1187, Dec. 2015, doi: 10.1049/iet-rsn.2015.0084.

[31]   Y. He, Y. Yang, Y. Lang, D. Huang, X. Jing, and C. Hou, 'Deep Learning based Human Activity Classification in Radar Micro-Doppler Image', in *2018 15th European Radar*

Conference, EuRAD 2018, IEEE, Sep. 2018, pp. 230–233. doi: 10.23919/EuRAD.2018.8546615.

[32]   S. Z. Gürbüz, B. Erol, B. Çağlıyan, and B. Tekeli, 'Operational assessment and adaptive selection of micro-Doppler features', *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1196–1204, Dec. 2015, doi: 10.1049/iet-rsn.2015.0144.

[33]   Y. He, X. Li, and X. Jing, 'A Mutiscale Residual Attention Network for Multitask Learning of Human Activity Using Radar Micro-Doppler Signatures', *Remote Sens (Basel)*, vol. 11, no. 21, p. 2584, Nov. 2019, doi: 10.3390/rs11212584.

[34]   C. Clemente, A. Balleri, K. Woodbridge, and J. J. Soraghan, 'Developments in target micro-Doppler signatures analysis: radar imaging, ultrasound and through-the-wall radar', *EURASIP J Adv Signal Process*, vol. 2013, no. 1, p. 47, Dec. 2013, doi: 10.1186/1687-6180-2013-47.

[35]   F. Fioranelli, J. Le Kernec, and S. A. Shah, 'Radar for Health Care: Recognizing Human Activities and Monitoring Vital Signs', *IEEE Potentials*, vol. 38, no. 4, pp. 16–23, Jul. 2019, doi: 10.1109/MPOT.2019.2906977.

[36]   Z. Li, J. Le Kernec, Q. Abbasi, F. Fioranelli, S. Yang, and O. Romain, 'Radar-based human activity recognition with adaptive thresholding towards resource constrained platforms', *Sci Rep*, vol. 13, no. 1, p. 3473, Mar. 2023, doi: 10.1038/s41598-023-30631-x.

[37]   Z. Li, Y. Liu, B. Liu, J. Le Kernec, and S. Yang, 'A holistic human activity recognition optimisation using AI techniques', *IET Radar, Sonar & Navigation*, Sep. 2023, doi: 10.1049/rsn2.12474.

[38]   Z. Li *et al.*, 'Multi-domains based human activity classification in radar', in *IET International Radar Conference (IET IRC 2020)*, Institution of Engineering and Technology, 2021, pp. 1744–1749. doi: 10.1049/icp.2021.0557.

[39]   Z. Li, J. Le Kernec, F. Fioranelli, O. Romain, L. Zhang, and S. Yang, 'An LSTM Approach to Short-range personnel recognition using Radar Signals', in *2021 IEEE Radar Conference (RadarConf21)*, IEEE, May 2021, pp. 1–6. doi: 10.1109/RadarConf2147009.2021.9455218.

[40]   Z. Li, F. Fioranelli, S. Yang, J. Le Kernec, Q. Abbasi, and O. Romain, 'Human Activity Classification with Adaptive Thresholding using Radar Micro-Doppler', in *2021 CIE International Conference on Radar (Radar)*, IEEE, Dec. 2021, pp. 1511–1515. doi: 10.1109/Radar53847.2021.10028630.

[41]   M. A. Richards, *Fundamentals of Radar Signal Processing*, 3rd Editio. New York: McGraw-Hill Education, 2022.

[42]   M. I. Skolnik, *Introduction to Radar Systems*. in Electrical engineering series. McGraw-Hill, 2001.

[43]   B. Zhou *et al.*, 'Simulation framework for activity recognition and benchmarking in different radar geometries', *IET Radar, Sonar and Navigation*, vol. 15, no. 4, pp. 390–401, 2021, doi: 10.1049/rsn2.12049.

[44]   Y. Richter, J. Gerasimov, N. Balal, and Y. Pinhasi, 'Tracking of Evasive Objects Using Bistatic Doppler Radar Operating in the Millimeter Wave Regime', *Remote Sens (Basel)*, vol. 14, no. 4, p. 867, Feb. 2022, doi: 10.3390/rs14040867.

[45]   T. Araki, T. Nagase, M. Araki, and H. Ono, 'A Two-Frequency CW Radar System for Short-Range Distances Measurements', *IEEJ Transactions on Sensors and Micromachines*, vol. 128, no. 7, pp. 302–305, 2008, doi: 10.1541/ieejsmas.128.302.

[46]   F. Gini, 'Grand Challenges in Radar Signal Processing', *Frontiers in Signal Processing*, vol. 1, Mar. 2021, doi: 10.3389/frsip.2021.664232.

[47]   V. Chen, *The Micro-Doppler Effect in Radar, Second Edition*. 2019.

[48]   M. A. Richards, J. Scheer, and W. A. Holm, *Principles of Modern Radar*, no. v. 3. in Principles of Modern Radar. SciTech Pub., 2010.

[49]   A. Coluccia, G. Parisi, and A. Fascista, 'Detection and Classification of Multirotor Drones in Radar Sensor Networks: A Review', *Sensors*, vol. 20, no. 15, p. 4172, Jul. 2020, doi: 10.3390/s20154172.

[50]   W. Lv, W. He, X. Lin, and J. Miao, 'Non-Contact Monitoring of Human Vital Signs Using FMCW Millimeter Wave Radar in the 120 GHz Band', *Sensors*, vol. 21, no. 8, p. 2732, Apr. 2021, doi: 10.3390/s21082732.

[51]   B. Welp *et al.*, 'Versatile Dual-Receiver 94-GHz FMCW Radar System With High Output Power and 26-GHz Tuning Range for High Distance Applications', *IEEE Trans Microw Theory Tech*, vol. 68, no. 3, pp. 1195–1211, Mar. 2020, doi: 10.1109/TMTT.2019.2955127.

[52]   T.-Y. Lim, S. A. Markowitz, and M. N. Do, 'RaDICaL: A Synchronized FMCW Radar, Depth, IMU and RGB Camera Data Dataset With Low-Level FMCW Radar Signals', *IEEE J Sel Top Signal Process*, vol. 15, no. 4, pp. 941–953, Jun. 2021, doi: 10.1109/JSTSP.2021.3061270.

[53]   J. Wang, 'CFAR-Based Interference Mitigation for FMCW Automotive Radar Systems', *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 12229–12238, Aug. 2022, doi: 10.1109/TITS.2021.3111514.

[54]   J. Gamba, *Radar Signal Processing for Autonomous Driving*. in Signals and Communication Technology. Singapore: Springer Singapore, 2020. doi: 10.1007/978-981-13-9193-4.

[55]   E. Antolinos, F. García-Rial, C. Hernández, D. Montesano, J. I. Godino-Llorente, and J. Grajal, 'Cardiopulmonary Activity Monitoring Using Millimeter Wave Radars', *Remote Sens (Basel)*, vol. 12, no. 14, p. 2265, Jul. 2020, doi: 10.3390/rs12142265.

[56]   V. L. Petrovic, M. M. Jankovic, A. V. Lupsic, V. R. Mihajlovic, and J. S. Popovic-Bozovic, 'High-Accuracy Real-Time Monitoring of Heart Rate Variability Using 24 GHz Continuous-Wave Doppler Radar', *IEEE Access*, vol. 7, pp. 74721–74733, 2019, doi: 10.1109/ACCESS.2019.2921240.

[57]   A. Stumpf, N. Lachiche, J.-P. Malet, N. Kerle, and A. Puissant, 'Active Learning in the Spatial Domain for Remote Sensing Image Classification', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2492–2507, May 2014, doi: 10.1109/TGRS.2013.2262052.

[58]   T. Blumensath and M. E. Davies, 'Normalized Iterative Hard Thresholding: Guaranteed Stability and Performance', *IEEE J Sel Top Signal Process*, vol. 4, no. 2, pp. 298–309, Apr. 2010, doi: 10.1109/JSTSP.2010.2042411.

[59]   M. Song, J. Lim, and D.-J. Shin, 'The velocity and range detection using the 2D-FFT scheme for automotive radars', in *2014 4th IEEE International Conference on Network Infrastructure and Digital Content*, IEEE, Sep. 2014, pp. 507–510. doi: 10.1109/ICNIDC.2014.7000356.

[60] A. Stumpf, N. Lachiche, J.-P. Malet, N. Kerle, and A. Puissant, 'Active Learning in the Spatial Domain for Remote Sensing Image Classification', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2492–2507, May 2014, doi: 10.1109/TGRS.2013.2262052.

[61] J. Pegoraro, J. O. Lacruz, M. Rossi, and J. Widmer, 'SPARCS: A Sparse Recovery Approach for Integrated Communication and Human Sensing in mmWave Systems', in *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, IEEE, May 2022, pp. 79–91. doi: 10.1109/IPSN54338.2022.00014.

[62] F. Qi, H. Lv, F. Liang, Z. Li, X. Yu, and J. Wang, 'MHHT-Based Method for Analysis of Micro-Doppler Signatures for Human Finer-Grained Activity Using Through-Wall SFCW Radar', *Remote Sens (Basel)*, vol. 9, no. 3, p. 260, Mar. 2017, doi: 10.3390/rs9030260.

[63] Boualem Boashash, 'Time-Frequency Methods in Radar, Sonar, and Acoustics', in *Time-Frequency Signal Analysis and Processing*, Elsevier, 2016, pp. 793–856. doi: 10.1016/B978-0-12-398499-9.00014-5.

[64] H. Li, A. Shrestha, H. Heidari, J. Le Kernec, and F. Fioranelli, 'A Multisensory Approach for Remote Health Monitoring of Older People', *IEEE J Electromagn RF Microw Med Biol*, vol. 2, no. 2, pp. 102–108, Jun. 2018, doi: 10.1109/JERM.2018.2827099.

[65] V. C. Chen, D. Tahmoush, and W. J. Miceli, Eds., *Radar Micro-Doppler Signatures: Processing and Applications*. Institution of Engineering and Technology, 2014. doi: 10.1049/PBRA034E.

[66] M. Li and J. Lin, 'Wavelet-Transform-Based Data-Length-Variation Technique for Fast Heart Rate Detection Using 5.8-GHz CW Doppler Radar', *IEEE Trans Microw Theory Tech*, vol. 66, no. 1, pp. 568–576, Jan. 2018, doi: 10.1109/TMTT.2017.2730182.

[67] Y. Dong, W. Qu, T. Gao, P. Wang, and H. Jiang, 'Analysis and Research on Radar Gesture Micro-motion Feature Extraction Method', *J Phys Conf Ser*, vol. 2031, no. 1, p. 012012, Sep. 2021, doi: 10.1088/1742-6596/2031/1/012012.

[68] M. Jia, S. Li, J. Le Kernec, S. Yang, F. Fioranelli, and O. Romain, 'Human activity classification with radar signal processing and machine learning', in *2020 International Conference on UK-China Emerging Technologies (UCET)*, IEEE, Aug. 2020, pp. 1–5. doi: 10.1109/UCET51115.2020.9205461.

[69] S. Björklund, H. Petersson, and G. Hendeby, 'Features for micro-Doppler based activity classification', *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1181–1187, Dec. 2015, doi: 10.1049/iet-rsn.2015.0084.

[70] Y. Du *et al.*, 'Radar-based Human Activity Classification with Cyclostationarity', in *2021 CIE International Conference on Radar (Radar)*, IEEE, Dec. 2021, pp. 1483–1487. doi: 10.1109/Radar53847.2021.10027946.

[71] A. Napolitano, *Cyclostationary processes and time series: theory, applications, and generalizations*. Academic Press, 2019.

[72] B. Erol and M. G. Amin, 'Radar Data Cube Processing for Human Activity Recognition Using Multisubspace Learning', *IEEE Trans Aerosp Electron Syst*, vol. 55, no. 6, pp. 3617–3628, Dec. 2019, doi: 10.1109/TAES.2019.2910980.

[73] Y. He, P. Molchanov, T. Sakamoto, P. Aubry, F. Le Chevalier, and A. Yarovoy, 'Range-Doppler surface: a tool to analyse human target in ultra-wideband radar', *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1240–1250, Dec. 2015, doi: 10.1049/iet-rsn.2015.0065.

[74] F. Fioranelli, H. Griffiths, M. Ritchie, and A. Balleri, Eds., *Micro-Doppler Radar and Its Applications*. Institution of Engineering and Technology, 2020. doi: 10.1049/SBRA531E.

[75] M. G. Amin and R. G. Guendel, 'Radar classifications of consecutive and contiguous human gross-motor activities', *IET Radar, Sonar & Navigation*, vol. 14, no. 9, pp. 1417–1429, Sep. 2020, doi: 10.1049/iet-rsn.2019.0585.

[76] G. Zhang, X. Geng, and Y.-J. Lin, 'Comprehensive mPoint: A Method for 3D Point Cloud Generation of Human Bodies Utilizing FMCW MIMO mm-Wave Radar', *Sensors*, vol. 21, no. 19, p. 6455, Sep. 2021, doi: 10.3390/s21196455.

[77] U. Chipengo, A. Sligar, and S. Carpenter, 'High Fidelity Physics Simulation of 128 Channel MIMO Sensor for 77GHz Automotive Radar', *IEEE Access*, vol. 8, pp. 160643–160652, 2020, doi: 10.1109/ACCESS.2020.3021362.

[78] E. Alpaydin, *Introduction to Machine Learning*, 2nd ed. The MIT Press, 2010.

[79] X. Li, Y. He, and X. Jing, 'A Survey of Deep Learning-Based Human Activity Recognition in Radar', *Remote Sens (Basel)*, vol. 11, no. 9, p. 1068, May 2019, doi: 10.3390/rs11091068.

[80] D. Maulud and A. M. Abdulazeez, 'A Review on Linear Regression Comprehensive in Machine Learning', *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140–147, Dec. 2020, doi: 10.38094/jastt1457.

[81] J. R. Quinlan, 'Induction of decision trees', *Mach Learn*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/BF00116251.

[82] C.-C. Chang and C.-J. Lin, 'LIBSVM', *ACM Trans Intell Syst Technol*, vol. 2, no. 3, pp. 1–27, Apr. 2011, doi: 10.1145/1961189.1961199.

[83] M. Ahmed, R. Seraj, and S. M. S. Islam, 'The k-means Algorithm: A Comprehensive Survey and Performance Evaluation', *Electronics (Basel)*, vol. 9, no. 8, p. 1295, Aug. 2020, doi: 10.3390/electronics9081295.

[84] X. Ran, Y. Xi, Y. Lu, X. Wang, and Z. Lu, 'Comprehensive survey on hierarchical clustering algorithms and the recent developments', *Artif Intell Rev*, vol. 56, no. 8, pp. 8219–8264, Aug. 2023, doi: 10.1007/s10462-022-10366-3.

[85] V. Sundararaj, 'Optimised denoising scheme via opposition-based self-adaptive learning PSO algorithm for wavelet-based ECG signal noise reduction', *Int J Biomed Eng Technol*, vol. 31, no. 4, p. 325, 2019, doi: 10.1504/IJBET.2019.103242.

[86] X. Li, X. Jing, and Y. He, 'Unsupervised Domain Adaptation for Human Activity Recognition in Radar', in *2020 IEEE Radar Conference (RadarConf20)*, IEEE, Sep. 2020, pp. 1–5. doi: 10.1109/RadarConf2043947.2020.9266433.

[87] K. K. Verma, B. M. Singh, and A. Dixit, 'A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system', *International Journal of Information Technology*, vol. 14, no. 1, pp. 397–410, Feb. 2022, doi: 10.1007/s41870-019-00364-0.

[88] X. Zhu and A. B. Goldberg, *Introduction to Semi-Supervised Learning*. in Synthesis Lectures on Artificial Intelligence and Machine Learning. Cham: Springer International Publishing, 2009. doi: 10.1007/978-3-031-01548-9.

[89] X. Li, Y. He, F. Fioranelli, and X. Jing, 'Semisupervised Human Activity Recognition With Radar Micro-Doppler Signatures', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022, doi: 10.1109/TGRS.2021.3090106.

[90] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018.

[91] M. Li, J. Gao, L. Zhao, and X. Shen, 'Deep Reinforcement Learning for Collaborative Edge Computing in Vehicular Networks', *IEEE Trans Cogn Commun Netw*, vol. 6, no. 4, pp. 1122–1135, Dec. 2020, doi: 10.1109/TCCN.2020.3003036.

[92] J. Kober, J. A. Bagnell, and J. Peters, 'Reinforcement learning in robotics: A survey', *Int J Rob Res*, vol. 32, no. 11, pp. 1238–1274, Sep. 2013, doi: 10.1177/0278364913495721.

[93] S. A. Shah and F. Fioranelli, 'Human Activity Recognition : Preliminary Results for Dataset Portability using FMCW Radar', in *2019 International Radar Conference (RADAR)*, IEEE, Sep. 2019, pp. 1–4. doi: 10.1109/RADAR41533.2019.171307.

[94] Z. Zhang, 'Introduction to machine learning: k-nearest neighbors', *Ann Transl Med*, vol. 4, no. 11, pp. 218–218, Jun. 2016, doi: 10.21037/atm.2016.03.37.

[95] C. Cortes and V. Vapnik, 'Support-vector networks', *Mach Learn*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[96] S. Mannor *et al.*, 'K-Means Clustering', in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 563–564. doi: 10.1007/978-0-387-30164-8_425.

[97] J. L. Puga, M. Krzywinski, and N. Altman, 'Bayes' theorem', *Nat Methods*, vol. 12, no. 4, pp. 277–278, Apr. 2015, doi: 10.1038/nmeth.3335.

[98] F. Ordóñez and D. Roggen, 'Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition', *Sensors*, vol. 16, no. 1, p. 115, Jan. 2016, doi: 10.3390/s16010115.

[99] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, 'Continuous Human Activity Classification from FMCW Radar with Bi-LSTM Networks', *IEEE Sens J*, vol. 20, no. 22, pp. 13607–13619, 2020, doi: 10.1109/JSEN.2020.3006386.

[100] H. Li, A. Shrestha, H. Heidari, J. Le Kernec, and F. Fioranelli, 'Bi-LSTM Network for Multimodal Continuous Human Activity Recognition and Fall Detection', *IEEE Sens J*, vol. 20, no. 3, pp. 1191–1201, Feb. 2020, doi: 10.1109/JSEN.2019.2946095.

[101] S. Yang, J. Le Kernec, F. Fioranelli, and O. Romain, 'Human Activities Classification in a Complex Space Using Raw Radar Data', in *2019 International Radar Conference (RADAR)*, IEEE, Sep. 2019, pp. 1–4. doi: 10.1109/RADAR41533.2019.171367.

[102] J. Zhu, H. Chen, and W. Ye, 'A Hybrid CNN–LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar', *IEEE Access*, vol. 8, pp. 24713–24720, 2020, doi: 10.1109/ACCESS.2020.2971064.

[103] H.-U.-R. Khalid, A. Gorji, A. Bourdoux, S. Pollin, and H. Sahli, 'Multi-View CNN-LSTM Architecture for Radar-Based Human Activity Recognition', *IEEE Access*, vol. 10, pp. 24509–24519, 2022, doi: 10.1109/ACCESS.2022.3150838.

[104] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[105] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, 'Learning representations by back-propagating errors', *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, doi: 10.1038/323533a0.

[106] L. Alzubaidi *et al.*, 'Review of deep learning: concepts, CNN architectures, challenges, applications, future directions', *J Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[107] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Incorporated, 2019.

[108] Y. LeCun *et al.*, 'Backpropagation Applied to Handwritten Zip Code Recognition', *Neural Comput*, vol. 1, no. 4, pp. 541–551, Dec. 1989, doi: 10.1162/neco.1989.1.4.541.

[109] Y. LeCun, Y. Bengio, and G. Hinton, 'Deep learning', *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[110] Y. Kim and B. Toomajian, 'Hand Gesture Recognition Using Micro-Doppler Signatures With Convolutional Neural Network', *IEEE Access*, vol. 4, pp. 7125–7130, 2016, doi: 10.1109/ACCESS.2016.2617282.

[111] J. J. Hopfield, 'Neural networks and physical systems with emergent collective computational abilities.', *Proceedings of the National Academy of Sciences*, vol. 79, no. 8, pp. 2554–2558, Apr. 1982, doi: 10.1073/pnas.79.8.2554.

[112] Y. Bengio, P. Simard, and P. Frasconi, 'Learning long-term dependencies with gradient descent is difficult', *IEEE Trans Neural Netw*, vol. 5, no. 2, pp. 157–166, Mar. 1994, doi: 10.1109/72.279181.

[113] S. Hochreiter and J. Schmidhuber, 'Long Short-Term Memory', *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[114] S. Yang *et al.*, 'The Human Activity Radar Challenge: Benchmarking Based on the "Radar Signatures of Human Activities" Dataset From Glasgow University', *IEEE J Biomed Health Inform*, vol. 27, no. 4, pp. 1813–1824, Apr. 2023, doi: 10.1109/JBHI.2023.3240895.

[115] Z. Niu, G. Zhong, and H. Yu, 'A review on the attention mechanism of deep learning', *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021, doi: 10.1016/j.neucom.2021.03.091.

[116] A. Vaswani *et al.*, 'Attention is all you need', in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, in NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 6000–6010.

[117] M.-H. Guo *et al.*, 'Attention mechanisms in computer vision: A survey', *Comput Vis Media (Beijing)*, vol. 8, no. 3, pp. 331–368, Sep. 2022, doi: 10.1007/s41095-022-0271-y.

[118] X. Chen, 'The Advance of Deep Learning and Attention Mechanism', in *2022 International Conference on Electronics and Devices, Computational Science (ICEDCS)*, IEEE, Sep. 2022, pp. 318–321. doi: 10.1109/ICEDCS57360.2022.00078.

[119] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', Oct. 2018.

[120] T. B. Brown *et al.*, 'Language Models are Few-Shot Learners', May 2020.

[121] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting', *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014.

[122] D. Bashir, G. D. Montañez, S. Sehra, P. S. Segura, and J. Lauw, 'An Information-Theoretic Perspective on Overfitting and Underfitting', 2020, pp. 347–358. doi: 10.1007/978-3-030-64984-5_27.

[123] K. H. Zou *et al.*, 'Statistical validation of image segmentation quality based on a spatial overlap index1', *Acad Radiol*, vol. 11, no. 2, pp. 178–189, Feb. 2004, doi: 10.1016/S1076-6332(03)00671-8.

[124] K. M. Ting, 'Confusion Matrix', in *Encyclopedia of Machine Learning and Data Mining*, Boston, MA: Springer US, 2017, pp. 260–260. doi: 10.1007/978-1-4899-7687-1_50.

[125] C. Goutte and E. Gaussier, 'A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation', 2005, pp. 345–359. doi: 10.1007/978-3-540-31865-1_25.

[126] D. Ogwok and E. M. Ehlers, 'Jaccard Index in Ensemble Image Segmentation: An Approach', in *Proceedings of the 2022 5th International Conference on Computational Intelligence and Intelligent Systems*, New York, NY, USA: ACM, Nov. 2022, pp. 9–14. doi: 10.1145/3581792.3581794.

[127] S. Z. Gurbuz and M. G. Amin, 'Radar-Based Human-Motion Recognition With Deep Learning: Promising Applications for Indoor Monitoring', *IEEE Signal Process Mag*, vol. 36, no. 4, pp. 16–28, Jul. 2019, doi: 10.1109/MSP.2018.2890128.

[128] J. Le Kernec *et al.*, 'Radar signal processing for sensing in assisted living', *IEEE Signal Process Mag*, no. 201806070002, 2019.

[129] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer, 'Monitoring Activities of Daily Living in Smart Homes: Understanding human behavior', *IEEE Signal Process Mag*, vol. 33, no. 2, pp. 81–94, Mar. 2016, doi: 10.1109/MSP.2015.2503881.

[130] K. Chaccour, R. Darazi, A. H. El Hassani, and E. Andres, 'From Fall Detection to Fall Prevention: A Generic Classification of Fall-Related Systems', *IEEE Sens J*, vol. 17, no. 3, pp. 812–822, Feb. 2017, doi: 10.1109/JSEN.2016.2628099.

[131] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer, 'Monitoring Activities of Daily Living in Smart Homes: Understanding human behavior', *IEEE Signal Process Mag*, vol. 33, no. 2, pp. 81–94, Mar. 2016, doi: 10.1109/MSP.2015.2503881.

[132] K. Chaccour, R. Darazi, A. H. El Hassani, and E. Andres, 'From Fall Detection to Fall Prevention: A Generic Classification of Fall-Related Systems', *IEEE Sens J*, vol. 17, no. 3, pp. 812–822, Feb. 2017, doi: 10.1109/JSEN.2016.2628099.

[133] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. C. D. Ho, 'Radar Signal Processing for Elderly Fall Detection: The future for in-home monitoring', *IEEE Signal Process Mag*, vol. 33, no. 2, pp. 71–80, Mar. 2016, doi: 10.1109/MSP.2015.2502784.

[134] H. Saner, 'Wearable Sensors for Assisted Living in Elderly People', *Frontiers in ICT*, vol. 5, Feb. 2018, doi: 10.3389/fict.2018.00001.

[135] E. Cippitelli, F. Fioranelli, E. Gambi, and S. Spinsante, 'Radar and RGB-Depth Sensors for Fall Detection: A Review', *IEEE Sens J*, vol. 17, no. 12, pp. 3585–3604, Jun. 2017, doi: 10.1109/JSEN.2017.2697077.

[136] A. Ledergerber and R. D'Andrea, 'A Multi-Static Radar Network with Ultra-Wideband Radio-Equipped Devices', *Sensors*, vol. 20, no. 6, p. 1599, Mar. 2020, doi: 10.3390/s20061599.

[137] B. Vandersmissen *et al.*, 'Indoor Person Identification Using a Low-Power FMCW Radar', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3941–3952, Jul. 2018, doi: 10.1109/TGRS.2018.2816812.

[138] F. Luo, S. Poslad, and E. Bodanese, 'Human Activity Detection and Coarse Localization Outdoors Using Micro-Doppler Signatures', *IEEE Sens J*, vol. 19, no. 18, pp. 8079–8094, Sep. 2019, doi: 10.1109/JSEN.2019.2917375.

[139] H. Sadreazami, M. Bolic, and S. Rajan, 'CapsFall: Fall Detection Using Ultra-Wideband Radar and Capsule Network', *IEEE Access*, vol. 7, pp. 55336–55343, 2019, doi: 10.1109/ACCESS.2019.2907925.

[140] S. Z. Gurbuz, C. Clemente, A. Balleri, and J. J. Soraghan, 'Micro-Doppler-based in-home aided and unaided walking recognition with multiple radar and sonar systems', *IET Radar, Sonar & Navigation*, vol. 11, no. 1, pp. 107–115, Jan. 2017, doi: 10.1049/iet-rsn.2016.0055.

[141] B. Jokanovic, M. Amin, and B. Erol, 'Multiple joint-variable domains recognition of human motion', in *2017 IEEE Radar Conference (RadarConf)*, IEEE, May 2017, pp. 0948–0952. doi: 10.1109/RADAR.2017.7944340.

[142] Z. Yang, H. Wang, P. Ni, P. Wang, Q. Cao, and L. Fang, 'Real-time Human Activity Classification From Radar With CNN-LSTM Network', in *2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA)*, IEEE, Aug. 2021, pp. 50–55. doi: 10.1109/ICIEA51954.2021.9516401.

[143] Z. Meng *et al.*, 'Gait Recognition for Co-Existing Multiple People Using Millimeter Wave Sensing', *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 849–856, Apr. 2020, doi: 10.1609/aaai.v34i01.5430.

[144] Youngwook Kim and Hao Ling, 'Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328–1337, May 2009, doi: 10.1109/TGRS.2009.2012849.

[145] S. S. Ram, C. Christianson, Youngwook Kim, and Hao Ling, 'Simulation and Analysis of Human Micro-Dopplers in Through-Wall Environments', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 4, pp. 2015–2023, Apr. 2010, doi: 10.1109/TGRS.2009.2037219.

[146] F. Fioranelli, M. Ritchie, and H. Griffiths, 'Performance Analysis of Centroid and SVD Features for Personnel Recognition Using Multistatic Micro-Doppler', *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 5, pp. 725–729, May 2016, doi: 10.1109/LGRS.2016.2539386.

[147] X. Li, Z. Li, F. Fioranelli, S. Yang, O. Romain, and J. Le Kernec, 'Hierarchical Radar Data Analysis for Activity and Personnel Recognition', *Remote Sens (Basel)*, vol. 12, no. 14, p. 2237, Jul. 2020, doi: 10.3390/rs12142237.

[148] S. Z. Gürbüz, B. Erol, B. Çağlıyan, and B. Tekeli, 'Operational assessment and adaptive selection of micro-Doppler features', *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1196–1204, Dec. 2015, doi: 10.1049/iet-rsn.2015.0144.

[149] Y. Yang, C. Hou, Y. Lang, T. Sakamoto, Y. He, and W. Xiang, 'Omnidirectional Motion Classification With Monostatic Radar System Using Micro-Doppler Signatures', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3574–3587, May 2020, doi: 10.1109/TGRS.2019.2958178.

[150] Y. Lin, J. Le Kernec, S. Yang, F. Fioranelli, O. Romain, and Z. Zhao, 'Human Activity Classification With Radar: Optimization and Noise Robustness With Iterative Convolutional Neural Networks Followed With Random Forests', *IEEE Sens J*, vol. 18, no. 23, pp. 9669–9681, Dec. 2018, doi: 10.1109/JSEN.2018.2872849.

[151] M. Wang, Y. D. Zhang, and G. Cui, 'Human motion recognition exploiting radar with stacked recurrent neural network', *Digit Signal Process*, vol. 87, pp. 125–131, Apr. 2019, doi: 10.1016/j.dsp.2019.01.013.

[152] M. Wang, G. Cui, X. Yang, and L. Kong, 'Human body and limb motion recognition via stacked gated recurrent units network', *IET Radar, Sonar & Navigation*, vol. 12, no. 9, pp. 1046–1051, Sep. 2018, doi: 10.1049/iet-rsn.2018.5054.

[153] Y. Kim and T. Moon, 'Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural Networks', *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 8–12, Jan. 2016, doi: 10.1109/LGRS.2015.2491329.

[154] Z. Chen, G. Li, F. Fioranelli, and H. Griffiths, 'Personnel Recognition and Gait Classification Based on Multistatic Micro-Doppler Signatures Using Deep Convolutional Neural Networks', *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 669–673, May 2018, doi: 10.1109/LGRS.2018.2806940.

[155] Y. Kim and T. Moon, 'Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural Networks', *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 8–12, Jan. 2016, doi: 10.1109/LGRS.2015.2491329.

[156] J. Zhu, H. Chen, and W. Ye, 'A Hybrid CNN–LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar', *IEEE Access*, vol. 8, pp. 24713–24720, 2020, doi: 10.1109/ACCESS.2020.2971064.

[157] L. Cao, S. Liang, Z. Zhao, D. Wang, C. Fu, and K. Du, 'Human Activity Recognition Method Based on FMCW Radar Sensor with Multi-Domain Feature Attention Fusion Network', *Sensors*, vol. 23, no. 11, p. 5100, May 2023, doi: 10.3390/s23115100.

[158] S. Huan *et al.*, 'A lightweight hybrid vision transformer network for radar-based human activity recognition', *Sci Rep*, vol. 13, no. 1, p. 17996, Oct. 2023, doi: 10.1038/s41598-023-45149-5.

[159] C. Loukas, F. Fioranelli, J. Le Kernec, and S. Yang, 'Activity Classification Using Raw Range and I & Q Radar Data with Long Short Term Memory Layers', in *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, IEEE, Aug. 2018, pp. 441–445. doi: 10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00088.

[160] B. Jokanovic, M. Amin, and B. Erol, 'Multiple joint-variable domains recognition of human motion', in *2017 IEEE Radar Conference (RadarConf)*, IEEE, May 2017, pp. 0948–0952. doi: 10.1109/RADAR.2017.7944340.

[161] Y. Li, Z. Peng, R. Pal, and C. Li, 'Potential Active Shooter Detection Based on Radar Micro-Doppler and Range-Doppler Analysis Using Artificial Neural Network', *IEEE Sens J*, vol. 19, no. 3, pp. 1052–1063, Feb. 2019, doi: 10.1109/JSEN.2018.2879223.

[162] B. Erol and M. G. Amin, 'Radar Data Cube Processing for Human Activity Recognition Using Multisubspace Learning', *IEEE Trans Aerosp Electron Syst*, vol. 55, no. 6, pp. 3617–3628, Dec. 2019, doi: 10.1109/TAES.2019.2910980.

[163] S. Björklund, H. Petersson, and G. Hendeby, 'Features for micro-Doppler based activity classification', *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1181–1187, Dec. 2015, doi: 10.1049/iet-rsn.2015.0084.

[164] M. Li, T. Chen, and H. Du, 'Human Behavior Recognition Using Range-Velocity-Time Points', *IEEE Access*, vol. 8, pp. 37914–37925, 2020, doi: 10.1109/ACCESS.2020.2975676.

[165] V. Chen and H. Ling, *Time-Frequency Transforms for Radar Imaging and Signal Analysis*. 2001.

[166] M. S. Seyfioglu, S. Z. Gurbuz, A. M. Ozbayoglu, and M. Yuksel, 'Deep learning of micro-Doppler features for aided and unaided gait recognition', in *2017 IEEE Radar Conference (RadarConf)*, IEEE, May 2017, pp. 1125–1130. doi: 10.1109/RADAR.2017.7944373.

[167] A.-K. Seifert, M. G. Amin, and A. M. Zoubir, 'Toward Unobtrusive In-Home Gait Analysis Based on Radar Micro-Doppler Signatures', *IEEE Trans Biomed Eng*, vol. 66, no. 9, pp. 2629–2640, Sep. 2019, doi: 10.1109/TBME.2019.2893528.

[168] E. L. McGough, R. G. Logsdon, V. E. Kelly, and L. Teri, 'Functional Mobility Limitations and Falls in Assisted Living Residents With Dementia', *Journal of Geriatric Physical Therapy*, vol. 36, no. 2, pp. 78–86, Apr. 2013, doi: 10.1519/JPT.0b013e318268de7f.

[169] A. A. Hulleck, D. Menoth Mohan, N. Abdallah, M. El Rich, and K. Khalaf, 'Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies', *Front Med Technol*, vol. 4, Dec. 2022, doi: 10.3389/fmedt.2022.901331.

[170] A.-K. Seifert, M. G. Amin, and A. M. Zoubir, 'New analysis of radar micro-Doppler gait signatures for rehabilitation and assisted living', in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Mar. 2017, pp. 4004–4008. doi: 10.1109/ICASSP.2017.7952908.

[171] V. C. Chen, Fayin Li, Shen-Shyang Ho, and H. Wechsler, 'Micro-doppler effect in radar: phenomenon, model, and simulation study', *IEEE Trans Aerosp Electron Syst*, vol. 42, no. 1, pp. 2–21, Jan. 2006, doi: 10.1109/TAES.2006.1603402.

[172] F. Wang, M. Skubic, M. Rantz, and P. E. Cuddihy, 'Quantitative Gait Measurement With Pulse-Doppler Radar for Passive In-Home Gait Assessment', *IEEE Trans Biomed Eng*, vol. 61, no. 9, pp. 2434–2443, Sep. 2014, doi: 10.1109/TBME.2014.2319333.

[173] G. Klarenbeek, R. I. A. Harmanny, and L. Cifola, 'Multi-target human gait classification using LSTM recurrent neural networks applied to micro-Doppler', in *2017 European Radar Conference (EURAD)*, IEEE, Oct. 2017, pp. 167–170. doi: 10.23919/EURAD.2017.8249173.

[174] K. Saho, K. Uemura, K. Sugano, and M. Matsumoto, 'Using Micro-Doppler Radar to Measure Gait Features Associated With Cognitive Functions in Elderly Adults', *IEEE Access*, vol. 7, pp. 24122–24131, 2019, doi: 10.1109/ACCESS.2019.2900303.

[175] H. T. Le, S. L. Phung, and A. Bouzerdoum, 'Human Gait Recognition with Micro-Doppler Radar and Deep Autoencoder', in *2018 24th International Conference on Pattern Recognition (ICPR)*, IEEE, Aug. 2018, pp. 3347–3352. doi: 10.1109/ICPR.2018.8546044.

[176] B. Erol and M. G. Amin, 'Fall motion detection using combined range and Doppler features', in *2016 24th European Signal Processing Conference (EUSIPCO)*, IEEE, Aug. 2016, pp. 2075–2080. doi: 10.1109/EUSIPCO.2016.7760614.

[177] B. Erol and M. G. Amin, 'Radar Data Cube Processing for Human Activity Recognition Using Multisubspace Learning', *IEEE Trans Aerosp Electron Syst*, vol. 55, no. 6, pp. 3617–3628, Dec. 2019, doi: 10.1109/TAES.2019.2910980.

[178] J. Bai, L. Zheng, S. Li, B. Tan, S. Chen, and L. Huang, 'Radar Transformer: An Object Classification Network Based on 4D MMW Imaging Radar', *Sensors*, vol. 21, no. 11, p. 3854, Jun. 2021, doi: 10.3390/s21113854.

[179] B. Erol, S. Z. Gurbuz, and M. G. Amin, 'GAN-based Synthetic Radar Micro-Doppler Augmentations for Improved Human Activity Recognition', in *2019 IEEE Radar Conference (RadarConf)*, IEEE, Apr. 2019, pp. 1–5. doi: 10.1109/RADAR.2019.8835589.

[180] C. Ding, Y. Zou, L. Sun, H. Hong, X. Zhu, and C. Li, 'Fall detection with multi-domain features by a portable FMCW radar', in *2019 IEEE MTT-S International Wireless Symposium (IWS)*, IEEE, May 2019, pp. 1–3. doi: 10.1109/IEEE-IWS.2019.8804036.

[181] J. Guo *et al.*, 'Complex field-based fusion network for human activities classification with radar', in *IET International Radar Conference (IET IRC 2020)*, Institution of Engineering and Technology, 2021, pp. 68–73. doi: 10.1049/icp.2021.0572.

[182] M. M. Rahman, S. Z. Gurbuz, and M. G. Amin, 'Physics-Aware Design of Multi-Branch GAN for Human RF Micro-Doppler Signature Synthesis', in *2021 IEEE Radar Conference (RadarConf21)*, IEEE, May 2021, pp. 1–6. doi: 10.1109/RadarConf2147009.2021.9455194.

[183] A.-K. Seifert, M. G. Amin, and A. M. Zoubir, 'Toward Unobtrusive In-Home Gait Analysis Based on Radar Micro-Doppler Signatures', *IEEE Trans Biomed Eng*, vol. 66, no. 9, pp. 2629–2640, Sep. 2019, doi: 10.1109/TBME.2019.2893528.

[184] F. Fioranelli, M. Ritchie, S. Z. Gurbuz, and H. Griffiths, 'Feature Diversity for Optimized Human Micro-Doppler Classification Using Multistatic Radar', *IEEE Trans Aerosp Electron Syst*, vol. 53, no. 2, pp. 640–654, Apr. 2017, doi: 10.1109/TAES.2017.2651678.

[185] S. Visalakshi and V. Radha, 'A literature review of feature selection techniques and applications: Review of feature selection in data mining', in *2014 IEEE International Conference on Computational Intelligence and Computing Research*, IEEE, Dec. 2014, pp. 1–6. doi: 10.1109/ICCIC.2014.7238499.

[186] F. Castanedo, 'A Review of Data Fusion Techniques', *The Scientific World Journal*, vol. 2013, pp. 1–19, 2013, doi: 10.1155/2013/704504.

[187] Centre National de la Recherche Scientifique, 'METHOD AND DEVICE FOR HUMAN ACTIVITY CLASSIFICATION USING RADAR MICRO DOPPLER AND PHASE', EP21306742, 2022

[188] K. Simonyan and A. Zisserman, 'Very Deep Convolutional Networks for Large-Scale Image Recognition', Sep. 2014.

[189] K. He, X. Zhang, S. Ren, and J. Sun, 'Deep Residual Learning for Image Recognition', in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[190] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, 'Learning Transferable Architectures for Scalable Image Recognition', in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 8697–8710. doi: 10.1109/CVPR.2018.00907.

[191] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, 'Densely Connected Convolutional Networks', in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

[192] X. Zhang, X. Zhou, M. Lin, and J. Sun, 'ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices', in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 6848–6856. doi: 10.1109/CVPR.2018.00716.

[193] 'Histogram', in *The Concise Encyclopedia of Statistics*, New York, NY: Springer New York, pp. 242–244. doi: 10.1007/978-0-387-32833-1_178.

[194] Y. Wang, U. Ahsan, H. Li, and M. Hagen, 'A Comprehensive Review of Modern Object Segmentation Approaches', Jan. 2023, doi: 10.1561/0600000097.

[195] K. Qin, K. Xu, F. Liu, and D. Li, 'Image segmentation based on histogram analysis utilizing the cloud model', *Computers & Mathematics with Applications*, vol. 62, no. 7, pp. 2824–2833, Oct. 2011, doi: 10.1016/j.camwa.2011.07.048.

[196] R. Storn and K. Price, 'Differential evolution--a simple and efficient heuristic for global optimization over continuous spaces', *Journal of global optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[197] J. Kennedy and R. Eberhart, 'Particle swarm optimization', in *Proceedings of ICNN'95-international conference on neural networks*, 1995, pp. 1942–1948.

[198] D. Whitley, 'A genetic algorithm tutorial', *Stat Comput*, vol. 4, no. 2, pp. 65–85, 1994.

[199] B. Liu, H. Aliakbarian, Z. Ma, G. A. E. Vandenbosch, G. Gielen, and P. Excell, 'An efficient method for antenna design optimization based on evolutionary computation and machine learning techniques', *IEEE Trans Antennas Propag*, vol. 62, no. 1, pp. 7–18, 2013.

[200] B. Liu, S. Koziel, and N. Ali, 'SADEA-II: A generalized method for efficient global optimization of antenna design', *J Comput Des Eng*, vol. 4, no. 2, pp. 86–97, 2017.

[201] T. J. Santner, B. J. Williams, W. I. Notz, and B. J. Williams, *The design and analysis of computer experiments*, vol. 1. Springer, 2003.

[202] J. E. Dennis and V. Torczon, 'Managing approximation models in optimization', *Multidisciplinary design optimization: State-of-the-art*, vol. 5, pp. 330–347, 1997.

[203] M. T. M. Emmerich, K. C. Giannakoglou, and B. Naujoks, 'Single-and multiobjective evolutionary optimization assisted by Gaussian random field metamodels', *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 4, pp. 421–439, 2006.

[204] B. Liu, Q. Zhang, and G. G. E. Gielen, 'A Gaussian process surrogate model assisted evolutionary algorithm for medium scale expensive optimization problems', *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 2, pp. 180–192, 2013.

[205] M. O. Akinsolu, B. Liu, V. Grout, P. I. Lazaridis, M. E. Mognaschi, and P. Di Barba, 'A Parallel Surrogate Model Assisted Evolutionary Algorithm for Electromagnetic Design Optimization', *IEEE Trans Emerg Top Comput Intell*, vol. 3, no. 2, pp. 93–105, 2019, doi: 10.1109/TETCI.2018.2864747.

[206] Z. Li, J. Le Kernec, F. Fioranelli, Q. Abbasi, S. Yang, and O. Romain, 'Human Activity Classification with Adaptive Thresholding using Radar Micro-Doppler', in *2021 CIE International Conference on Radar (CIE Radar 2021)*, Haikou, Hainan, China, 2021.

[207] Z. Li, J. Le Kernec, Q. Abbasi, F. Fioranelli, S. Yang, and O. Romain, 'Radar-based human activity recognition with adaptive thresholding towards resource constrained platforms', *Sci Rep*, vol. 13, no. 1, p. 3473, Mar. 2023, doi: 10.1038/s41598-023-30631-x.

[208] X. Li, Y. He, and X. Jing, 'A Survey of Deep Learning-Based Human Activity Recognition in Radar', *Remote Sens (Basel)*, vol. 11, no. 9, p. 1068, May 2019, doi: 10.3390/rs11091068.

[209] L. Alzubaidi *et al.*, 'Review of deep learning: concepts, CNN architectures, challenges, applications, future directions', *J Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[210] S. Z. Gurbuz and M. G. Amin, 'Radar-Based Human-Motion Recognition With Deep Learning: Promising Applications for Indoor Monitoring', *IEEE Signal Process Mag*, vol. 36, no. 4, pp. 16–28, Jul. 2019, doi: 10.1109/MSP.2018.2890128.

[211] K. K. Verma, B. M. Singh, and A. Dixit, 'A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system', *International Journal of Information Technology*, vol. 14, no. 1, pp. 397–410, Feb. 2022, doi: 10.1007/s41870-019-00364-0.

[212] C. Cortes and V. Vapnik, 'Support-vector networks', *Mach Learn*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[213] S. Visalakshi and V. Radha, 'A literature review of feature selection techniques and applications: Review of feature selection in data mining', in *2014 IEEE International Conference on Computational Intelligence and Computing Research*, IEEE, Dec. 2014, pp. 1–6. doi: 10.1109/ICCIC.2014.7238499.

[214] S. Z. Gürbüz, B. Erol, B. Çağlıyan, and B. Tekeli, 'Operational assessment and adaptive selection of micro-Doppler features', *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1196–1204, Dec. 2015, doi: 10.1049/iet-rsn.2015.0144.

[215] D. Nguyen, H. Hong, K. Kim, and K. Park, 'Person Recognition System Based on a Combination of Body Images from Visible Light and Thermal Cameras', *Sensors*, vol. 17, no. 3, p. 605, Mar. 2017, doi: 10.3390/s17030605.

[216] S. Yang, S. Hoque, and F. Deravi, 'Improved Time-Frequency Features and Electrode Placement for EEG-Based Biometric Person Recognition', *IEEE Access*, vol. 7, pp. 49604–49613, 2019, doi: 10.1109/ACCESS.2019.2910752.

[217] L. Yang, G. Li, M. Ritchie, F. Fioranelli, and H. Griffiths, 'Gait classification based on micro-Doppler features', in *2016 CIE International Conference on Radar (RADAR)*, IEEE, Oct. 2016, pp. 1–4. doi: 10.1109/RADAR.2016.8059301.

[218] Y. Yang, C. Hou, Y. Lang, G. Yue, Y. He, and W. Xiang, 'Person Identification Using Micro-Doppler Signatures of Human Motions and UWB Radar', *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 5, pp. 366–368, May 2019, doi: 10.1109/LMWC.2019.2907547.

[219] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[220] M. Wang, Y. D. Zhang, and G. Cui, 'Human motion recognition exploiting radar with stacked recurrent neural network', *Digit Signal Process*, vol. 87, pp. 125–131, Apr. 2019, doi: 10.1016/j.dsp.2019.01.013.

[221] F. Ordóñez and D. Roggen, 'Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition', *Sensors*, vol. 16, no. 1, p. 115, Jan. 2016, doi: 10.3390/s16010115.

[222] M. Wang, G. Cui, X. Yang, and L. Kong, 'Human body and limb motion recognition via stacked gated recurrent units network', *IET Radar, Sonar & Navigation*, vol. 12, no. 9, pp. 1046–1051, Sep. 2018, doi: 10.1049/iet-rsn.2018.5054.

[223] Z. Li, J. Le Kernec, F. Fioranelli, O. Romain, L. Zhang, and S. Yang, 'An LSTM Approach to Short-range personnel recognition using Radar Signals', in *2021 IEEE Radar Conference (RadarConf21)*, IEEE, May 2021, pp. 1–6. doi: 10.1109/RadarConf2147009.2021.9455218.

[224] A. Shrestha, H. Li, J. Le Kernec, and F. Fioranelli, 'Continuous Human Activity Classification from FMCW Radar with Bi-LSTM Networks', *IEEE Sens J*, vol. 20, no. 22, pp. 13607–13619, 2020, doi: 10.1109/JSEN.2020.3006386.

[225] Y. LeCun, Y. Bengio, and G. Hinton, 'Deep learning', *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[226] J. Liang and R. Liu, 'Stacked denoising autoencoder and dropout together to prevent overfitting in deep neural network', in *2015 8th International Congress on Image and Signal Processing (CISP)*, IEEE, Oct. 2015, pp. 697–701. doi: 10.1109/CISP.2015.7407967.

[227] H. Li, A. Shrestha, H. Heidari, J. Le Kernec, and F. Fioranelli, 'A Multisensory Approach for Remote Health Monitoring of Older People', *IEEE J Electromagn RF Microw Med Biol*, vol. 2, no. 2, pp. 102–108, Jun. 2018, doi: 10.1109/JERM.2018.2827099.

[228] M. S. Greco, J. Li, T. Long, and A. Zoubir, 'Advances in Radar Systems for Modern Civilian and Commercial Applications: Part 1 [From the Guest Editors]', *IEEE Signal Process Mag*, vol. 36, no. 4, pp. 13–15, Jul. 2019, doi: 10.1109/MSP.2019.2911108.

[229] M. S. Greco, J. Li, T. Long, and A. Zoubir, 'Advances in Radar Systems for Modern Civilian and Commercial Applications: Part 2 [From the Guest Editors]', *IEEE Signal Process Mag*, vol. 36, no. 5, pp. 16–18, Sep. 2019, doi: 10.1109/MSP.2019.2925158.

[230] C. Dhiman and D. K. Vishwakarma, 'View-Invariant Deep Architecture for Human Action Recognition Using Two-Stream Motion and Shape Temporal Dynamics', *IEEE Transactions on Image Processing*, vol. 29, pp. 3835–3844, 2020, doi: 10.1109/TIP.2020.2965299.

[231] Jiang Wang, Zicheng Liu, Ying Wu, and Junsong Yuan, 'Mining actionlet ensemble for action recognition with depth cameras', in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2012, pp. 1290–1297. doi: 10.1109/CVPR.2012.6247813.

[232] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, 'Multimodal Machine Learning: A Survey and Taxonomy', *IEEE Trans Pattern Anal Mach Intell*, vol. 41, no. 2, pp. 423–443, Feb. 2019, doi: 10.1109/TPAMI.2018.2798607.

[233] N. C. Garcia, P. Morerio, and V. Murino, 'Modality Distillation with Multiple Stream Networks for Action Recognition', 2018, pp. 106–121. doi: 10.1007/978-3-030-01237-3_7.

[234] J. Zhou and J. Le Kernec, '4D radar simulator for human activity recognition', *IET Radar, Sonar & Navigation*, Sep. 2023, doi: 10.1049/rsn2.12468.

[235] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, 'Explainable AI: A Review of Machine Learning Interpretability Methods', *Entropy*, vol. 23, no. 1, p. 18, Dec. 2020, doi: 10.3390/e23010018.

[236] J. Wang, X. Jing, Z. Yan, Y. Fu, W. Pedrycz, and L. T. Yang, 'A Survey on Trust Evaluation Based on Machine Learning', *ACM Comput Surv*, vol. 53, no. 5, pp. 1–36, Sep. 2021, doi: 10.1145/3408292.