RADAR BASED DISCRETE AND CONTINUOUS ACTIVITY RECOGNITION FOR ASSISTED LIVING

Aman Shrestha

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

2021

Abstract

In an era of digital transformation, there is an appetite for automating the monitoring process of motions and actions by individuals who are part of a society increasingly getting older on average. "Activity recognition" is where sensors use motion information from participants who are wearing a wearable sensor or are in the field of view of a remote sensor which, coupled with machine learning algorithms, can automatically identify the movement or action the person is undertaking.
Radar is a nascent sensor for this application, having been proposed in literature as an effective privacy-compliant sensor that can track movements of the body effectively.
The methods of recording movements are separated into two types where 'Discrete' movements provide an overview of a single activity within a fixed interval of time, while 'Continuous' activities present sequences of activities performed in a series with variable duration and uncertain transitions, making these a challenging and yet much more realistic classification problem.
In this thesis, first an overview of the technology of continuous wave (CW) and frequency modulated continuous wave (FMCW) radars and the machine learning algorithms and classification concepts is provided. Following this, state of the art for activity recognition with radar is presented and the key papers and significant works are discussed. The remaining chapters of this thesis discuss the research topics where contributions were made. This is commenced through analysing the effect of the physiology of the subject under test, to show that age can have an effect on the radar readings on the target. This is followed by porting existing radar recognition technologies and presenting novel use of radar based gait recognition to detect lameness in animals. Reverting to the human-centric application, improvements to activity recognition on humans and its accuracy was demonstrated by utilising features from different domains with feature selection and using different sensing technologies cooperatively. Finally, using a Bi-long short term memory (LSTM) based network, improved recognition of continuous activities and activity transitions without human-dependent feature extraction was demonstrated.
Accuracy rate of 97% was achieved through sensor fusion and feature selection for discrete activities and for continuous activities, the Bi-LSTM achieved 92% accuracy with a sole radar sensor.

Dedication

To my family.

Acknowledgements

I am truly indebted to my supervisors Dr. Francesco Fioranelli and Dr. Julien Le Kernec for their help and support throughout this endeavour. Without their wisdom and guidance the research in this thesis and the thesis itself would not have been possible. I am especially thankful to Francesco for taking a chance on me and to Julien for checking up on me during certain arduous moments in this journey. I am thankful also to Prof. Muhammad Imran for his guidance and help with timekeeping. To Dr. Sevgi Zubeyde Gurbuz I am thankful for opening my eyes to the potential to the capabilities of radar, to not be intimidated by challenging problems, and for the research experience in University of Alabama.

My colleagues at Glasgow were also of great help to me, with special mentions for Bruno Citoni, Jarez Patel, Dr. Paulo Valente Klaine, Dr. Joao Pedro Batistella Nadas, Dr. Abed Poursharab, Matteo Pepa, Dr. Qiheng Yuan, and my partner at work Dr. Haobo Li.

Dr. Yusuf Sambo and Giancarlo Patané were excellent company in the CSI lab room and Dr. Lun Ma gave me the gift of photography that I cannot thank enough and for rekindling my love of cycling I would like to thank Glasgow University Triathlon Club (2019-2020).

Thanks to Yingke and Joe for their help in my time of need as well.

Due to this PhD, I'm glad I was fortunate enough to meet Simon, Anna, Silvi, Peter, and Adrien .

My friends from home were also a constant source of support so I'd like to thank Mhairi, Alasdair, Zoe, Ryan and Aimee.

From my home away from home I'd like to thank Jamie, Nathan, and Dwayne for their company, intellect, and coffee.

Finally, I'd like to thank my mum, my dad, and my little brother for their love and support.

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.

Aman Shrestha

Contents

List of Tables

List of Figures

Acronyms and Symbols

AI artificial intelligence.

ANN artificial neural network.

CAE convolutional auto encoder.

CVD cadence velocity diagram.

CW continuous wave.

DNN deep neural network.

FMCW frequency modulated continuous wave.

GAN generative adversarial network.

ICA independent component analysis.

IMU inertial measurement unit.

LSTM long short term memory.

mD micro Doppler.

ML machine learning.

MLP multilayer perceptron.

PCA principle component analysis.

RNN recurrent neural network.

SFS sequential feature selection.

SPHERE sensor platform for healthcare in a residential environment.

STFT short-time Fourier transform.

SVD singular value decomposition.

SVM support vector machine.

TCN temporal convolutional network.

UWB ultra-wideband.

Symbols

f_r radial Doppler frequency

c speed of electromagnetic waves in free space

v_r radial target velocity

v_s radial and source velocity

f_d radial Doppler velocity

f_0 centre frequency

λ_0 wavelength of the transmitted wave

$s(t)$ frequency modulated signal

$s(f)$ beat-note

$f(y)$ instantaneous frequency

k chirp rate

B bandwidth

T chirp period

(ϕ_0) induced phase shift

r radar range between transmitter and target

$s_r(t)$ received modulated signal

$I(t)$ in-phase component

$Q(t)$ quadrature component

r_{max}The maximum unambiguous range

n slow time

$\Psi(t)$ phase component after target movement

$\bar{\phi}$ collection of all terms from the linear phase

δ Dirac delta function

τ_m maximum travel time from the target

r range

v velocity

a_w scale parameter

b_w translation parameter

f_m mother wavelet

$f_c(j)$ Doppler centroid

$B_c(j)$ Doppler bandwidth

$p(x)$ probability mass function of the histogram

b bias

α hidden weight

β weight

f_{ac} activation function

e^T exponential positive estimates

ξ_i slack variable

C penalty parameter

W weight

h_t hidden weight

R recurrent weight

o unnormalized log probability

U weight matrix connecting the input to the hidden weight

w weight matrix connecting the hidden to hidden connections forward in time

V hidden to output connection weight matrix

σ_t sigmoid activation function

c_t cell state

$S_n(c)$ loss value for a class

μ mean

σ^2 variance

Chapter 1

Introduction

The world health organisation expects that between 2015 and 2050, the proportion of people over 60 years will double from 12% to 22% [1]. This growing age group that continues to live to a more extended age is facing challenges unforeseen in the previous generations at a grander scale.

By 2020 it was expected that the number of over sixty-year older adults will outnumber children under five years [2]. Although due to recent world events, this is less true, there should still be a serious focus on assisting systems and support mechanisms on this age group. Supporting this demographic shift through health and social systems is a significant challenge that many societies globally will face, for which they are unprepared.

These challenges can be summarised as multimorbidity conditions that correlate with age and present severe consequences to the affected individual and societal and economic systems which support the individual [3].

To assist ageing individuals in their private homes, a multitude of sensing methods have been suggested in the literature and utilised in the industry [4][5]. The most common and widely used sensors are wearable devices, which have reached market maturity along with broad acceptance. With these devices providing information such as heart rate, blood oxygen level and electrocardiogram of the heart, wearable devices on the market are feature-rich. Ambient sensors, contrarily, provide other advantages such as privacy, physical comfort, and the ease of compliance, as many wearable sensors can

be and are left behind by users.

Ambient sensing has its own share of preexisting technologies such as video monitoring with depth cameras, which use image based techniques and give high fidelity information about the environment while maintaining privacy through image filters. On the contrary, pressure pads or fall detecting floors take information directly from the user for their specific use case scenarios.

Doppler radar continues to be a challenger to the traditional sensing methods as it combines the perceived benefits of both systems. It is another example of an ambient sensor, therefore it is conveniently stored and operated without user interaction, and it is also easy to implement in different home environments. Also while depth camera image has visual information relating to the user which can give hints about them, for example their physical size or height, the data generated by the radar is more abstract, providing no clear information about the user. Therefore it can be considered to be more secure in terms of privacy. Nevertheless, there are certain drawbacks to using radar for activity recognition too, as it is not portable; it requires a co-located processor, and it is limited to indoor environments with a narrow field of view, which is susceptible to occlusion.

Besides healthcare applications for humans, there is also significant interest for remote sensing in the animal welfare and agricultural sectors. Musculoskeletal degeneration is not confined to humans, and its impact on the physical well-being of the animals can have a financial and emotional strain on owners and the animal themselves. The benefits of ambient sensing are again useful as setting physical sensors on animals can be a challenge; furthermore, it can affect their behaviour during diagnostic testing.

Lameness is a significant problem for farmed animals and performance horses and it has a negative impact on animal welfare and economically, both in terms of lost production and treatment costs. In dairy cattle, lameness is widely regarded as a major welfare problem. Difficulties with early identification of lameness in dairy cattle are a well recognized issue [6].

There is a financial incentive to focus on this specific application as overall economic losses resulting from lameness have been estimated to be around U.S. $75 per cow per year [7]. The true extent of lameness in the UK dairy herds is unknown, but the herd

level incidence has been estimated at 50 limb cases per 100 cow years [8]. Sheep farmers are faced with a similar problem, with prevalence as high as 10% of the flock [9] and an estimated cost to the UK sheep industry of £24 million, for the most common cause of lameness in sheep [10]. For horses, the most frequent disease syndrome recorded in the UK in 2016 was lameness, accounting for 33% of the reported issues [11].

The most common form of lameness identification is by a subjective visual scoring method. While subjective gait assessment methods provide an immediate, on-site recognition and require non-technical equipment, they show a high degree of variation in reliability and repeatability of and between observers [12], [13]. More objective kinetic and kinematic methods to identify both lameness and limbic abnormalities have been studied, such as force plate systems, 3-D-accelerometers, infrared thermography, and tracking mixed with modelling from vision-based and optoelectronic systems; these show promise when compared with more traditional methods [14], [15]. In this context, radar sensing can potentially enable contactless and automatic detection of lameness, due to its previously discussed benefits such as no need of additional sensors attached to the body of the animals under test, and sensing capabilities provided in any weather or lighting conditions, including outdoor farm environments while mitigating the risks associated with manually attaching sensors on these beings.

1.1 Context

In the start of this project, as it is today, there was no general consensus on the ideal sensing technology to approach these problems. Although the use of wearable sensors made the bulk of the literature for classifying activities and monitoring movement, the use of privacy-oriented ambient sensors were mostly absent.

Between 2005 and 2016, the use of radar was proposed in literature but it remained a nascent technology for the specific purpose of detecting activities or monitoring people. Over the last five years however, it has grown as a research topic as interest in this specific application has increased. The main focus of this research interest can be summarized in a two-fold manner. Firstly the new research has focused on expanding the use of radar for activity monitoring by introducing it for novel applications and

secondly, by introducing methods to improve the accuracy of recognition so the radar performs comparatively to other competing technologies currently being utilized.

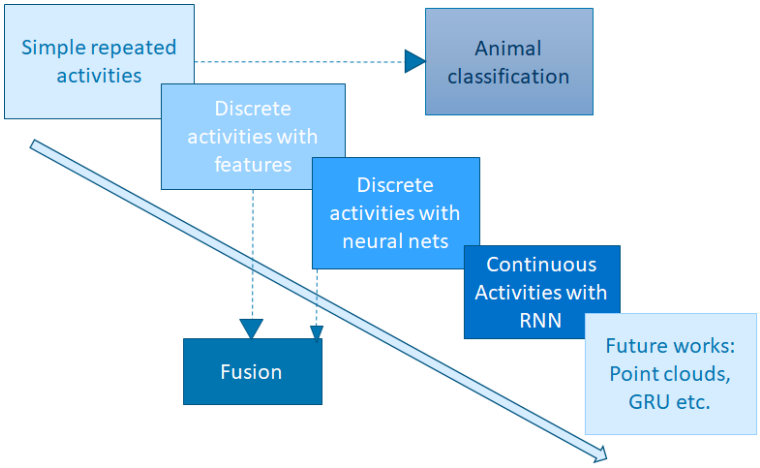

Figure 1.1: Evolution of the state of the art throughout this project.

The state of the art and the contributions provided to it is presented in Fig. 1.1 and there it can be seen how radar for assisted living has expanded from detecting simple activities at low rates to complex sequences of activities. There is a contribution here in made through innovation occurring laterally with new applications like animals and multi-sensor fusion. While ultimately the focus is innovation for single radar systems with state-of-the-art machine learning algorithms and complex realistic scenarios.

1.2 Contributions

Based on the state of the art at the commencement and during the project the following research questions were generated, inspired by papers such as Kim and Ling in [16]. The research providing the answers to these research questions are included below and are the main contributions of this thesis.

How does the target population and handcrafted feature extraction influence the accuracy of the classification algorithm?

1) By demonstrating how the variety of subject physiology affects the classification rates.

1) Through novel use of radar-based gait recognition to detect lameness in animals which are physiologically different from humans.

How can the sensing methods be improved to optimize the classification accuracy in the context of assisted living?

2) Proving that utilizing features from different domains with feature selection increases activity recognition accuracy.
2) Using different sensing technologies cooperatively to further increase accuracy.

How advances/evolution in machine learning can be leveraged to automatically segment and classify human activity in a continuous data stream with a single radar sensor?

3) Demonstrating improved recognition of continuous activities and activity transitions without feature extraction.

The literature has also shown a large variety in the collection of data where in certain cases the data collected could be considered almost excessive and the element of privacy would again be critiqued, while in others there is not enough information and therefore it is difficult to extract meaning or dependency of the participants. This links together with the research question about **how sensing methods and experimental techniques can be improved to optimise the classification accuracy in the context of assisted living?** Chapter 5 outlines our methodology to develop a growing data set containing information from multiple sensors and for a diverse range of participants to help address such question.

A smaller trend noted is the desire to expand the current applications of activity recognition with radar. Although classifying presence of animals in signatures has been shown in the literature, we introduced new applications by detecting impaired gait in domestic animals in [17]. This links together with the research question about **radar based gait recognition to detect lameness in animals which are physiologically different from humans**. The contribution here was showing that the techniques are not limited to human mobility and they can be expanded to other animals.

As for the overarching trends, the majority of the works generated new methods of

extracting salient information from the highly abstract radar data, showing **feature extraction influences the accuracy of the classification**. The contribution in this area is outlined in [18, 19, 20], where the variety of sensed data was utilized to select optimal features from a wide breadth of sensing methods. These works are covered in chapters 6 and 7.

There is also a visible trend about utilizing **advances/evolution in machine learning to automatically segment and classify human activity in a continuous data stream with a single radar sensor**. There is definite interest in moving to automatic methods of feature generation and an end-to-end classification method. The contribution to this is the use of recurrent networks to classify true continuous sequence of activities [21] in chapter 8.

1.2.1 Statement of novelty

In this thesis, the micro-Doppler signatures of animals are presented and utilised in the context of gait monitoring. Furthermore, signature of subjects with different physiology are presented to show the effect of the physiology of the subject under test. Then by using homogeneous fusion with different radar systems and heterogeneous sensor fusion between inertial measurement units and radar sensors, improvements to activity recognition with humans with radar fusion is demonstrated. Finally,a recurrent network is designed using Bi-LSTM layers and a spectrogram sequence of continuous activity signature with transitions, is used to validate and improved recognition of continuous activities and activity transitions without human-dependent feature extraction is presented.

1.3 Elaboration of contributions

This thesis mirrors the focus of the evolving research interest of using micro-Doppler radar for activity recognition and monitoring. **The scope covered by this research is discussed in the state of the art section and it encompasses hardware, software and experimental techniques to leverage radar technology with**

micro-Doppler and beyond for healthcare applications. In this rapidly developing domain, there are central questions this thesis attempts to answer. Included here-forth are the questions derived from the literature review and contributions made by this thesis for improving activity recognition through the use of radar sensors.

The first inquires about an issue which is common query in many machine learning (ML) applications for activity recognition, regardless of the sensor : **How does the target population and handcrafted feature extraction influence the accuracy of the classification algorithm?** This is answered through contributions 1 and 2 which are explained in detail in chapter 6. Specifically this is done:

1. By demonstrating how the variety of subject physiology affects the classification rates.

In the early stages, the PhD was aligned to the state of the art by using methods which were common in the literature at the time, so solving simpler activity classification problems. In this stage, the subject of interest was remarkably the actual test subjects as around 2016 the general machine learning community were starting to realise that the input, alongside the algorithm and hyper-parameters controlling them, were as important for the final outcome. Therefore the effect of variety in age, gender and physiology of the subjects on the accuracy of activity detection was evaluated, this was an early contribution as the majority of the research prior had a recognisable pattern in the demographics of the test subjects.

2. Through novel use of radar based gait recognition to detect lameness in animals which are physiologically different from humans.

Following that, the use-cases of applying micro Doppler (mD) based radar classification to problems such as gait monitoring in animals was introduced. This represented a shift in interest of the applicability of this technology as the bulk of the research focused on human-centric applications even though mammalian animals exhibited similar problems which required the inception of activity recognition in the first place. By demonstrating the ability to use mD with equines, we showed that mobility testing could be automated with results approaching to the level of human experts in detecting lameness or gait problems.

The second question asks, **How can the sensing methods be improved to op-**

timise the classification accuracy in the context of assisted living?. This is answered through contributions 3 and 4, explained in chapter 7, and it is done by:

3. Proving that utilizing features from different domains with feature selection increases activity recognition accuracy.
By this point in the literature there were a significant number of features generated using different properties of the input signal; therefore, the next goal was to utilise these different methods of generating features for automated activity recognition. By using them together with feature selection we were able to demonstrate higher accuracy in activity recognition.

4. Using different sensing technologies cooperatively to further increase accuracy.
In parallel, the question of using different sensors together was arising as the benefit of increased variance through multi-domain inputs were being discussed. Through heterogeneous sensor fusion, where sensors were measuring different physical signature, and homogeneous sensor fusion, where sensors were measuring the same physical signature but at different frequencies, we demonstrated again that the accuracy of activity recognition could be increased.

The final question inquires about **how advances/evolution in machine learning can be leveraged to automatically segment and classify human activity in a continuous data stream with a single radar sensor.**

This is answered through the final contribution 5 in chapter 8 by:
5. Demonstrating improved recognition of continuous activities and activity transitions without feature extraction.
Towards the end of the thesis, the research interest had been driven to using edge artificial intelligence (AI) techniques with transfer learning and deep neural networks being in the spotlight. In this moment we presented a time-dependent machine learning (ML) algorithm used together with a whole spectrogram to show the best case single sensor performance for radar. This was also done with an input spectrogram without specified feature generation with the key challenge of transitions present between activities mimicking realistic behaviour.

1.3.1 Published Articles

[21] A. Shrestha, H. Li, J. L. Kernec, and F. Fioranelli, "Continuous human activity classification from fmcw radar with bi-lstm networks," *IEEE Sensors Journal*, vol. 20, pp. 13607–13619, 2020

[22] H. Li, X. Liang, A. Shrestha, Y. Liu, H. Heidari, J. L. Kernec, and F. Fioranelli, "Hierarchical sensor fusion for micro-gesture recognition with pressure sensor array and radar," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 4, pp. 225–232, 2020

[23] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "Bi-lstm network for multimodal continuous human activity recognition and fall detection," *IEEE Sensors Journal*, vol. 20, pp. 1191–1201, 2020

[19] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "Magnetic and radar sensing for multimodal remote health monitoring," *IEEE Sensors Journal*, vol. 19, pp. 8979–8989, 2019

[24] A. Shrestha, H. Li, F. Fioranelli, and J. L. Kernec, "Activity recognition with cooperative radar systems at c and k band," *The Journal of Engineering*, vol. 2019, pp. 7100–7104, 2019

[25] A. Shrestha, J. L. Kernec, F. Fioranelli, Y. Lin, Q. He, J. Lorandel, and O. Romain, "Elderly care: activities of daily living classification with an s band radar," *The Journal of Engineering*, vol. 2019, pp. 7601–7606, 2019

[17] A. Shrestha, C. Loukas, J. L. Kernec, F. Fioranelli, V. Busin, N. Jonsson, G. King, M. Tomlinson, L. Viora, and L. Voûte, "Animal lameness detection with radar sensing," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, pp. 1189–1193, 2018

[26] H. Li, A. Shrestha, F. Fioranelli, J. L. Kernec, and H. Heidari, "Hierarchical classification on multimodal sensing for human activity recogintion and fall detection," *2018 IEEE SENSORS*, pp. 1–4, 2018

[27] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "A multisensory approach for remote health monitoring of older people," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 2, no. 2, pp. 102–108, 2018

1.3.2 Book chapter

[20] A. Shrestha, H. Li, F. Fioranelli, and J. L. Kernec, *Multimodal sensing for assisted living using radar*. Micro-Doppler Radar and Its Applications, Institution of Engineering and Technology, 2020

1.3.3 Published Conference proceedings

[28] A. Shrestha, C. Murphy, I. Johnson, A. Anbulselvam, F. Fioranelli, J. L. Kernec, and S. Gurbuz, "Cross-frequency classification of indoor activities with dnn transfer learning," *2019 IEEE Radar Conference (RadarConf)*, pp. 1–6, 2019

[29] A. Shrestha, J. L. Kernec, F. Fioranelli, E. Cippitelli, E. Gambi, and S. Spinsante, "Feature diversity for fall detection and human indoor activities classification using radar systems," in *International Conference on Radar Systems (Radar 2017)*, 2017

[30] A. Shrestha, J. L. Kernec, F. Fioranelli, J. Marshall, and L. Voûte, "Gait analysis of horses for lameness detection with radar sensors," in *International Conference on Radar Systems (Radar 2017)*, 2017

[18] H. Li, A. Shrestha, F. Fioranelli, J. L. Kernec, H. Heidari, M. Pepa, E. Cippitelli, E. Gambi, and S. Spinsante, "Multisensor data fusion for human activities classification and fall detection," *2017 IEEE SENSORS*, pp. 1–3, 2017

1.3.4 Open-access Research Database

[31] F. Fioranelli, S. A. Shah, H. Li, A. Shrestha, S. Yang, and J. L. Kernec, "Radar sensing for healthcare," *Electronics Letters*, vol. 55, pp. 1022–1024, 2019

1.4 Thesis Structure

The thesis is organised as follows:

Chapter 2 and 3 introduce the working principles of frequency modulated continuous wave (FMCW) radar, its frontend and pre-processing together with the ML backend required to recognise activities automatically. Note that feature extraction will not be

covered in the background as this is explained throughout the thesis as part of the research development. **Chapter 2** serves the basic ideas behind the operation of a frequency modulated continuous wave (FMCW) Radar and in a step-by-step format, overviews the hardware and signal processing required to generate a mD signature from an activity performed by a target. **Chapter 3** overviews the machine learning aspect of this research by discussing the fundamental versions of algorithms and machine learning architectures used for this thesis. Furthermore, it explains how the input signature becomes a categorical decision by using the classifier. **The genesis of features for the classification process is discussed in this chapter.**

Chapter 4 provides an overview of radar research related to human healthcare until recently. It covers the initial experiments where radars were used on humans for diagnostic purposes, continuing to the early 2005s where it was first used for activity recognition and 2015-2020 where the research interest and the innovation of machine learning techniques together with novel sensing took it to unforeseen levels of interest. Furthermore this chapter highlights the gaps in the literature which provided opportunities for the research presented in this thesis to be conducted, answering the question on what techniques can improve activity recognition performance using radar.

Chapter 5 discusses the setup of the various experiments conducted for this thesis. It clearly provides the reasoning behind the sensor and environmental setup whilst explaining why it is necessary to have a demographic of the participants that is diverse.

Chapter 6 explains the initial use of handcrafted features for activity recognition with a single sensor. It specifically mentions two applications one relating to animal welfare [30, 17] and the other regarding dependency of variety in human subjects [29, 18] for activity recognition applications.

Chapter 7 shows the strength of leveraging different types of information from different sources through sensor fusion in activity recognition. This can be done in the form of homogeneous sensor fusion where the both sensors are ambient radar sensors [24] or heterogeneous sensor fusion [19, 20, 32] where one or many sensors measuring different information to the radar can be applied. Both of these techniques are explored in this chapter while the effect of classification models having prior information of the target on the recognition accuracy is evaluated.

Chapter 8 specifies how automatic feature generation methods are utilized for activity recognition by generating salient features for classification of complex and continuous activity sequences. Using a time dependent neural network design, a radar signature of sequences of five activities are classified without feature selection or handcrafted feature extraction [21] seamlessly.

Chapter 9, lastly, presents the concluding remarks of this thesis and discusses future ideas and perspectives for activity recognition with radar.

Chapter 2

Radar background

The detection of human motion with radar has two fundamental technical components: the radar subsystems and the cognitive machine learning systems. As the physical hardware comes first in the overall detection scheme, we will look at this first in this thesis as well.

This chapter serves the basic principles behind the operation of radar and, in a step by step format, overviews the hardware and signal processing required to generate signature from an activity performed by a target. Firstly, a brief overview of the history of radar is given and the operating of radar presented. Following on, the extraction of the range and Doppler information, the structures and types of radar systems, micro-Doppler and, the radar hardware used in this thesis is discussed.

2.1 A brief history of radar

The foundation of Radio detection and ranging (Radar) was set when James Clark Maxwell revealed his theory on electromagnetism; then Heinrich Hertz, using this theory, experimented in 1888 to produce radio waves and demonstrate how metallic bodies could reflect and collate them.

As a first in the direct application for this technology, it was applied by Christian Huelsmeyer [33] for detecting ships in fog. Furthermore, Dr A. Hoyt Taylor et al [34]. from the United States Navy noticed that radio signals were interrupted when

large metallic ships crossed the antennas. Furthermore, at this time, Marconi endorsed the use of radio waves for detection. It can be seen that even in the early days, the application of radar was the main driver of this technology. As with other technologies ahead of their time, radar was unused. It changed in the 1930s where it had a resurgence due to the looming threat of war. In the arms race before World War Two, Sir Robert Watson-Watt and his scientific assistant Arnold Wilkins theorized the use of radio interruptions based on his experience of Radio based meteorology. They added a key innovation and using BBC radio antennas, demonstrated that aeroplanes could be detected at up to 13 km. Known as the Daventry experiment, it was the preliminary demonstration which resulted in the mainstream adaption of radar for defence and an appropriate application to show how useful the technology can be.

The innovation Watson-Watt and Wilkins made is the use of pulse modulation, where short pulses were transmitted and received, and this information could then be used to detect the angular velocity of a target.

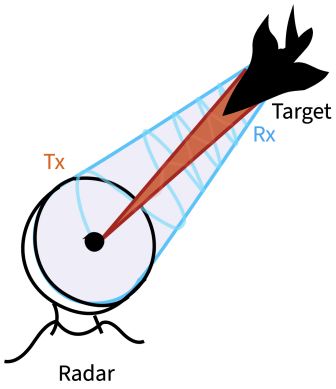

Figure 2.1: The basic operation of a radar detector. The transmitted wave is denoted as Tx and the received wave is Rx. In this example the target is a airborne target with high reflectivity.

2.2 Radar basic principles

In this section, the basic principles behind the operation of radar will be over-viewed with focus on aspects such as modulation, range extraction and micro Doppler extrac-

tion. This will provide a basis for the underlying processes in 2.1 and the composite technologies and methods required for target detection such as the frequency modulation, range and Doppler estimation, and the micro-Doppler effect in radar.

2.2.1 Continuous wave

Taking the examples of the pioneers of radar, by using targets in fig.2.2, the basic radar equation is used to deduce the radial velocity.

$$f_r = \left(\frac{c + v_r}{c + v_s}\right) f_0 \tag{2.1}$$

Eq. 2.1 is the Doppler equation as popularised by Christian Doppler, which applies to all waves. Here, f_r is the radial Doppler frequency, c is the speed of electromagnetic waves in free space, while v_r and v_s are the radial target and source velocities, respectively [35]. In the case of radar targets, however, since their velocity is much smaller compared to electromagnetic waves, the target Doppler can be defined by:

$$f_d = \frac{-f_0(v_s - v_r)}{c} \tag{2.2}$$

f_d in Eq. 2.2 shows the radial Doppler velocity of a target and f_0 denotes the centre frequency.

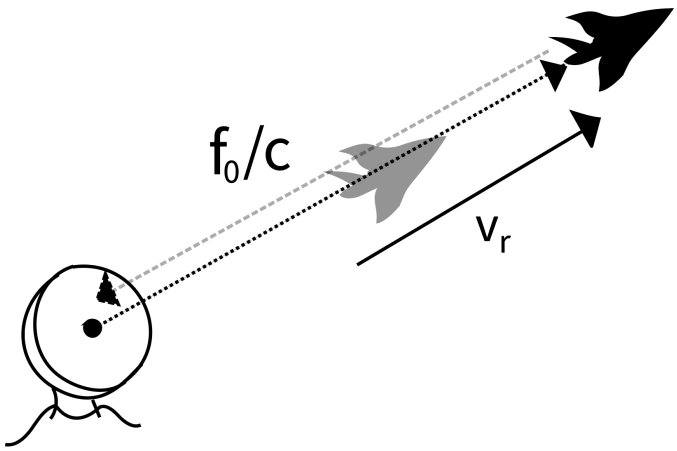

Figure 2.2: Detecting the radar target's radial velocity

Since the electromagnetic wave travels twice, once from the transmitter to the target and then from the target to the receiver, therefore the relationship becomes:

$$f_d = \frac{-2f_0(v_s - v_r)}{c} \tag{2.3}$$

Finally, if the radar is considered to be stationary then the following relationship occurs:

$$f_d = \frac{-2f_0(v_r)}{c} = -\frac{2v_r}{\lambda_0} \tag{2.4}$$

where the λ_0 is the wavelength of the radar. Eq. 2.2-2.4.

2.2.1.1 Frequency modulated continuous wave

With advancements in computing and processors becoming faster and smaller, new modulation methods compared to the previous pulse-Doppler methods could now be employed. Furthermore, an alternative modulation pattern, which does not require the transceiver to operate in a duty cycle, unlike pulse modulation was introduced. This alternative modulation method was frequency modulated continuous wave (FMCW).

In addition to radial velocity, similar to pulse-Doppler radar, the FMCW radar can also give the range of the target.

The frequency modulated signal is given by eq. 2.5, where $f(y)$ is an instantaneous frequency:

$$s(t) = Acos\left(2\pi \int_0^t f(y)dy\right) \tag{2.5}$$

For the commonly used linear modulation, considering k to be chirp rate while B being bandwidth and T being the chirp period, the equation becomes:

$$s(t) = Acos\left(2\pi(f_0 t + \frac{k}{2t^2})\right) \tag{2.6}$$

where:

$$k = \frac{B}{T} \tag{2.7}$$

Assuming a phase shift (ϕ_0) is introduced during transmission, the relationship becomes:

$$s_t(t) = A_t cos\bigg(2\pi(f_0 t + \frac{k}{2}t^2) + \phi_0\bigg). \tag{2.8}$$

For the received wave, assuming the target is at a distance there will be a time delay before the signal is received which can be characterised as:

$$\tau = \frac{2r}{c} \tag{2.9}$$

where, r is the range between the radar and the target and c is the speed of electromagnetic wave in a vacuum.

Considering eq. 2.9, the received signal is represented by:

$$s_r(t) = A_r cos\bigg(2\pi(f_0(t-\tau) + \frac{k}{2}(t-\tau)^2) + \phi_r\bigg) \tag{2.10}$$

To demodulate it, the signal is duplexed to first give the in-phase (I) component:

$$I(t) = s_t s_r = A_b(\bigg(cos(2\pi(f_0(t-\tau) + \frac{k}{2}(t-\tau)^2) + \phi_r\bigg) \tag{2.11}$$

where the total amplitude A_b is:

$$A_b = \frac{A_t A_r}{2} \tag{2.12}$$

while the phases ϕ_1 and ϕ_2 are:

$$\begin{aligned} \phi_1 &= \phi_r + \phi_0 \\ \phi_2 &= \phi_r - \phi_0 \end{aligned} \tag{2.13}$$

$I(t)$ is then low pass filtered to suppress the first harmonic close to the double of the centre frequency, finally giving:

$$I(t) = A_b\bigg(cos(\pi k\tau^2 - 2\pi kt\tau - 2\pi f_0\tau + \phi_2\bigg) \tag{2.14}$$

The quadrature component(Q) is given by mixing the received signal shifted by 90° which results in:

$$Q(t) = -A_b\Big(sin(2\pi t(2f_0 - k\tau) + 2\pi kt^2 + \phi_1 - 2\pi f_0\tau) + sin(\pi k\tau^2 - 2\pi kt\tau - 2\pi f_0\tau + \phi_2)\Big) \tag{2.15}$$

After filtering it in a similar manner to I, Q becomes:

$$Q(t) = A_b sin(-\pi k\tau^2 + 2\pi kt\tau + 2\pi f_0\tau - \phi 0) \tag{2.16}$$

Mixing the I and Q components leads to the generation of the beat note or the beat frequency which is given as:

$$s(t) = I(t) + jQ(t) = A_b e^{2\pi kt\tau - \pi k\tau^2 + 2\pi f_0\tau - \phi_2} = Ae^{\Psi(t)}.$$

The maximum unambiguous range can then be estimated using the derivative of $\Psi(t)$:

$$r_{max} = \frac{ck\tau}{2k} = \frac{c\tau}{2}, \tag{2.17}$$

due to :

$$f_b = \frac{1}{2pi}\frac{d\Psi(t)}{dt} = k\tau. \tag{2.18}$$

For a moving target, the range changes over time:

$$\tau(t) = \frac{2(\tau + vnTcos\theta)}{c}, \tag{2.19}$$

where θ is the aspect angle between the radar and the target while the v is the velocity. Considering a target moving away between consecutive chirps, where the period of each chirp is denoted as T, τ can now be given as:

$$\tau(n) = \frac{2(\tau + vnTcos\theta)}{c} = \frac{2vnTcos\theta}{c} + \tau_0 \tag{2.20}$$

Here, n is the slow time, which is time between the chirps and t is the fast time over which the sweep of the chirp is completed.

With the centering in the middle of the pulse, to offset the linear component ($t = T/2nt = L\Delta T/2$), the phase $\Psi(t)$ from 2.2.1.1 therefore changes to:

$$\Psi(t) = (2\pi k\tau_0 + \frac{2\pi kvcos\theta}{c}LT) - (2\pi f_d)nT - \bar{\phi}, \tag{2.21}$$

where $\bar{\phi}$ is the collection of all terms from the linear phase.

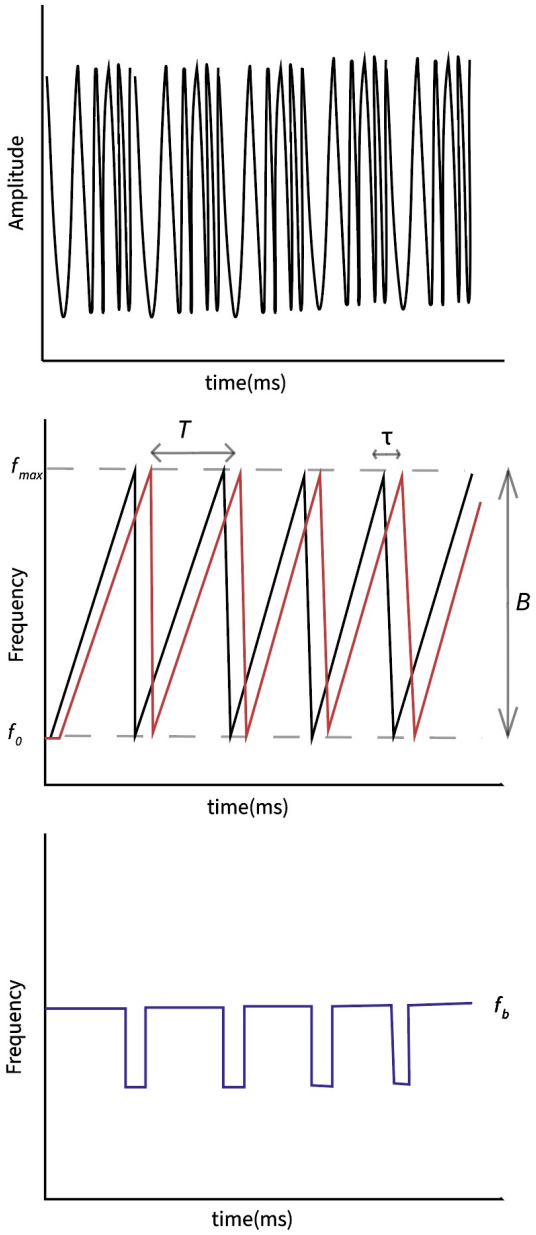

Figure 2.3: FMCW signals in time domain, frequency domain and the beat-note.

2.2.1.2 Double Fourier transform processing for range and Doppler estimation

There are now two frequency elements in the phase term in eq.2.21. They are:

$$\begin{aligned} w_f &= 2\pi k\tau 0 + \frac{2\pi kvcos\theta}{c}LT \\ w_s &= 2\pi f_d \end{aligned} \tag{2.22}$$

Applying the 2-D fourier transform to eq. 2.21 gives:

$$S(p,q) = \sum_{q=1}^{L}\left(\sum_{t=\tau_m}^{T} s(t)e^{-j2\pi pt/(T-\tau_m)}\right)e^{-j2\pi qt/L} = \hat{A}\delta(p-w_f, q-w_s), \tag{2.23}$$

where δ is the Dirac delta function, τ_m is the maximum travel time from the target. From here, the Doppler frequency and velocity can be estimated from $S(p,q)$ and w_f can be used to find the range.

For slow moving targets, the second frequency component of w_f may be ignored therefore:

$$\begin{aligned} w_f &= 2\pi f_f = 2\pi k\tau_0 \\ w_s &= 2\pi f_s = 2_\pi f_d \end{aligned} \tag{2.24}$$

The range and velocity can now be derived with:

$$\begin{aligned} r &= \frac{cf_f}{2k} \\ v &= \frac{f_s\lambda}{2} \end{aligned} \tag{2.25}$$

To summarise, in a linear modulated FMCW radar, the chirped signal is transmitted then received, then received at which point it is demodulated and stacked in the representation shown in eq.2.23. A 2-D Fourier transform can then be applied to estimate the range and velocity cells.

2.2.2 Radar system structure

A CW radar is composed of one or more transmitter and receiver antennae to transmit and receive the radar signal, an oscillator to generate the base signal and a demodulator to separate I and Q data. Low pass filters are present in the chain to remove the centre and first harmonics, amplifiers to increase the baseband information and finally, an analog to digital converter to convert the baseband information into data which can be processed further digitally.

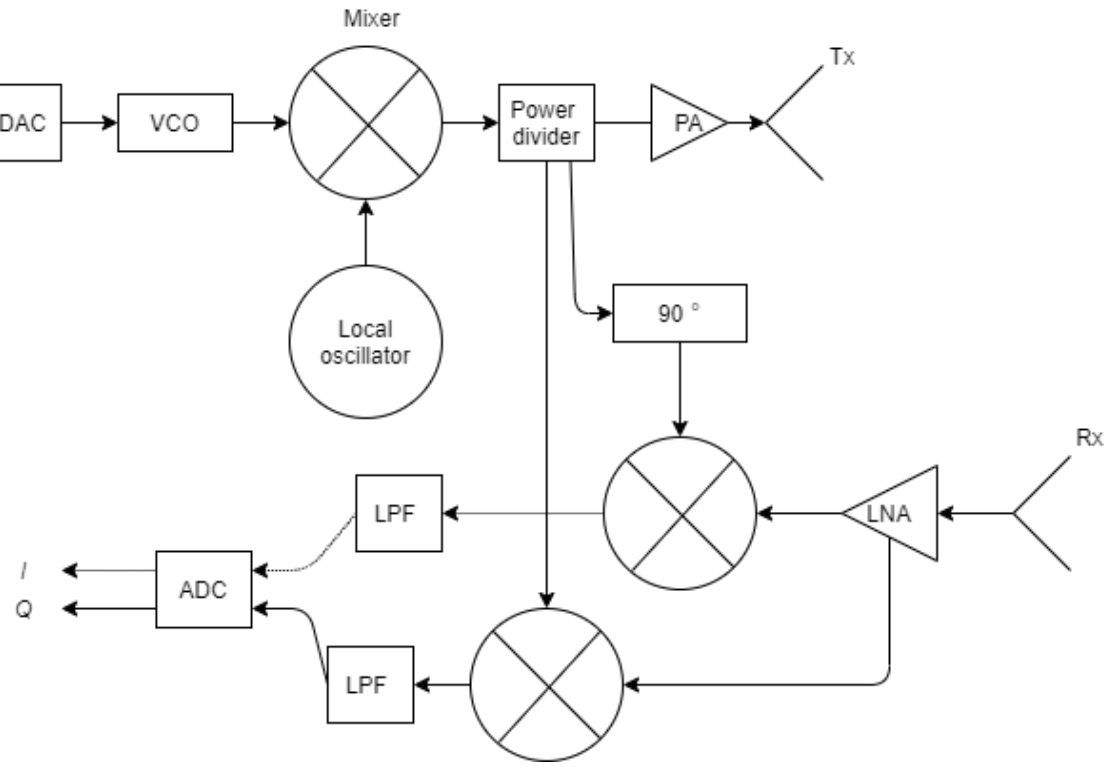

Figure 2.4: Block diagram of a typical FMCW radar system.

An FMCW radar differs from a CW radar due to the presence of a ramp generator in the system, usually in the form of a voltage controlled oscillator. Since the chirps are required to be mixed with the centre frequency generated by the signal generating oscillator, a mixer is also required. Therefore, these additional parts can make FMCW radar systems more expensive commercially, however due to their ubiquitous nature for different radar applications, economies of scale mean this issue is not of great concern.

2.2.3 Micro-Doppler

In a high frequency system, small vibrations can induce a large phase change, which induce detectable Doppler shifts. Mechanical systems, biological or otherwise, often have components with oscillatory movement which is broadly referred to as micro-motion. As opposed to the bulk or macro movement of a system, micro-motion refers to smaller oscillatory movements which would normally be occluded with a radar detector. In aircraft, this could be the fixed rotor or an engine turbine, while in humans, this would be swinging and rotating arms and limbs.

Micro Doppler therefore, is the small scale Doppler shifts generated by micro motions. mD appears in radar signatures as sidebands, and with these modulations, aforementioned movements such as rotation of rotors and swinging of limbs can be detected, and the corresponding kinematic properties can be identified.

The unambiguous detection of a Doppler shift is related to the centre frequency and mD too has this fundamental property. [36] discusses the displacement change required for an X-band and an L-band radar. A motion with a vibration of 15 Hz being detected by an X-band radar requires a 0.3 cm displacement. For an X-band system (3cm wavelength), it induces a 18.8 Hz m-D shift. However, for an L-band system, a 1 cm movement is required to induce the same 18.8 Hz shift. This means that radars with a higher carrier frequency can detect smaller m-D signatures and that the centre frequency is an important property to consider for m-D applications.

Chen in [36] further characterises the mD shift as a time-varying frequency shift which can be extracted from the complex output of a quadrature detector. However for a monotonic radar system where a single component is present, the Hilbert transform can be used to reconstruct the real and imaginary part of the signal. This allows any radar system to be used to generate a mD signature, through the joint time frequency analysis method.

2.2.4 Joint time frequency representation

As mD shifts are aperiodic and non-stationary effects, they require joint time-frequency analysis due to the inability of the Fourier transform to convey time-varying spectral

information. Due to computational advances again, the possibility of performing large numbers of successive Fourier transforms became real which in turn allowed the expansion of mD research.

The short-time Fourier transform (STFT) is the prominent joint time frequency method used due to its simplicity and efficiency.

$$STFT(t,\omega) = \int x(\tau)w(\tau - t)e^{-j\omega t}d\tau, \tag{2.26}$$

where $\tau - t$ is a time lattice.

As shown by equation 2.26 it is in essence, a rolling windowed Fourier transform meaning it can now show time-varying information.

The spectrogram is a method to visualise the coefficients of the STFT, it is given by:

$$Spectrogram(t,\omega) = |STFT(t,\omega)|^2 \tag{2.27}$$

and it is the squared magnitude of the STFT without the phase information of the signal [36].

Due to the time based limitation of the window function, it means there is a trade off between the time and the frequency resolution. A longer time window means increased Doppler resolution while a smaller integration time means increased time resolution. Overlapping windows can help with edge discontinuities, to generate a smooth signature at the cost of increased computational load.

For radars capable of ranging, the joint time-frequency method can be performed for individual range bins. As range profiles will have less sampling points than a raw quadrature signal, zero-padding can be necessary to increase the resolution. Using zero padding has two benefits, the resulting spectrogram will have increased resolution, and subjects or targets can be isolated by selecting specific range bins.

Wavelet transform (WT) has been proposed. Unlike STFT, which has a fixed frequency resolution, WT has multi-resolution capability. A 'mother' wavelet is used and its scaled and translated versions in order to detect frequencies at multiple resolutions. There is a variety of choices including, but not limited to: Haar, Coiflet, Mexican Hat,

Morlet, Symlet and Daubechies etc. Its expression is shown below, a_w is the scale parameter, b_w is the translation parameter and f_m is the mother wavelet.

$$WT(a_w, b_w) = \frac{1}{\sqrt{a_w}} x(t) f_m^* \left(\frac{t - b_w}{a_w} \right) dt \tag{2.28}$$

While WT overcomes the shortcomings of the STFT, the frequency bands for human micro-Doppler are limited to a 150 Hz window where the simplicity and the computational efficiency of the STFT is more desirable. The kernel based nature of the WT also provides further complexity (i.e. which wavelet to choose etc.). Both WT and STFT are examples of linear transforms. In addition to these, bilinear and quadratic transforms have also been used in literature.

A high resolution TF technique is the Wigner-Ville Distribution (WVD) which requires a time dependent autocorrelation function for signal $s(t)$ and a Fourier transform to be performed on it to give:

$$WVD(t, f) = \int s(t + \frac{t'}{2}) s^*(t - \frac{t'}{2}) exp(-j2\pi f t') dt' \tag{2.29}$$

It has a better joint time frequency resolution than any linear transforms but it has a weakness which appears in the form of cross term interference which means the sum of two signals ($s = s1 + s2$) does not equal the sum of their WVD. A residual component would be present and to reduce this, filtered WVD have been used. This reduces the time-frequency resolution but also reduces, significantly, the cross term components. With a linear low pass filter, the WVD belongs to the Cohen's class (Cohen's class being generalised forms of time frequency distributions):

$$C(t, f) = \int \int s(t + \tau/2) s^*(t - \tau/2) \phi(t - u, \tau) exp(-j2\pi f \tau) du d\tau \tag{2.30}$$

$\phi(t, \tau)$ is the Fourier transform of the filter and is called the kernel function. Selection of different kernels can be used to reduce cross terms with a variety of choices which changes the properties of the transform. Other non-linear TF methods have been applied: Hilbert Huang Transform (HHT) have also been used. Firstly, Empirical mode decomposition (EMD) is performed on a signal to decompose it into components.

Hilbert transform is then utilised to analyse the instantaneous frequency of each of these components. The methods have been listed in the order of complexity, with a trade-off between resolution and processing present and given the multiple drawbacks of these methods the simplest and the fastest method of STFT was used in this research.

2.2.5 Human micro Doppler

A person performing motions in front of radar can be illustrated through the spectrogram and this is shown in fig. 2.5 and 2.6 where examples of mD signatures are provided for human activities for walk and a punch-retract action while the subject remains stationary. The difference in the dynamic range of the modulations from the torso (normalised at 0 dB) and the limbs (around -23 dB) are identifiable, and these motions are central behind the idea of using mD signatures for classifying different activities. The dynamic range was clipped at -40 dB.

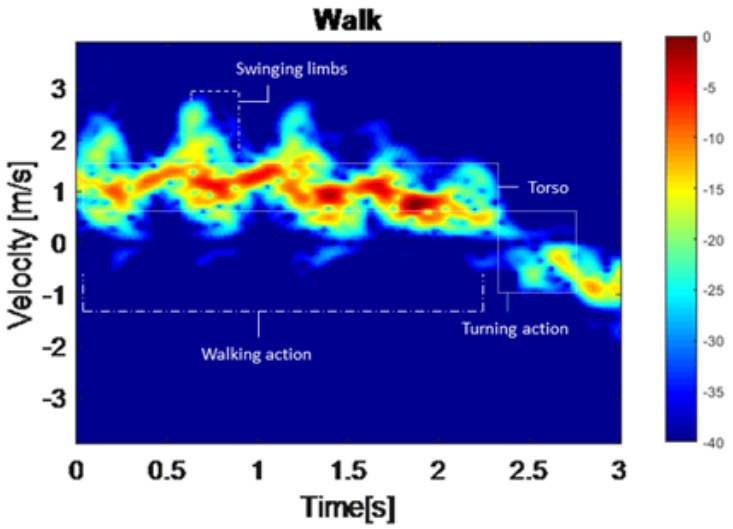

Figure 2.5: Micro Doppler signature of a person walking forwards and backwards. The signatures from the torso and the limbs are annotated describing the movement the signature is derived from. The intensity of the signature around the torso movement gives a stronger reflection than the limbs.

Fig. 2.5 shows a mD signature for a person walking, where they are moving towards the radar (positive Doppler), turning (zero-Doppler crossing), and walking away from

the radar (Negative Doppler). From 0 to 2.5 s, the subject moves towards the radar. The deep red trace represents the mean speed of the subject with the torso as the main reflection contributor and surrounding the red trace yellow-green fluctuations that represent the Doppler of limbs relative to the torso main Doppler component. From 1.5 to 2.5 s, the subject decelerates; therefore the main Doppler component and relative-Doppler components swing range decreases too. The person reaches a stand-still position at approximately 2.5 s, where they turn away from the radar and start walking in the other direction. mD signatures have visually recognisable properties that can help identify different states of the human body. It can show if a person is walking or moving slowly, if they are turning or moving their limbs in a certain manner, and together with sequences of these smaller motions if they are performing specific activities.

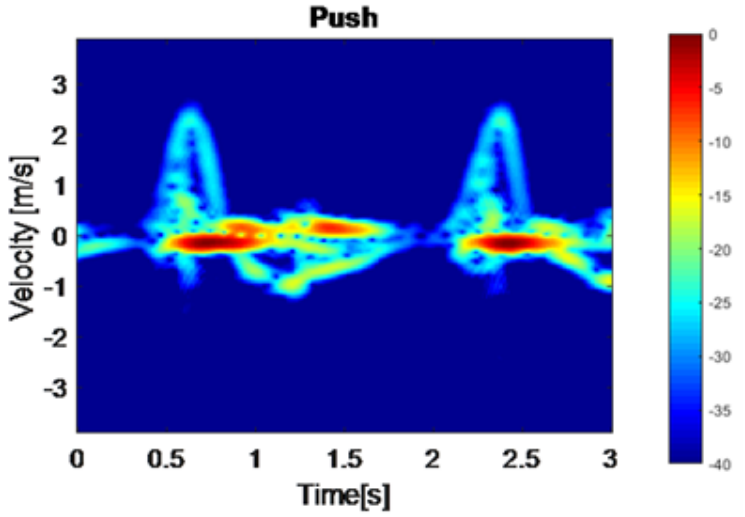

Figure 2.6: Micro Doppler Spectrogram of a person punching/retracting while staying stationary normalised to 0dB.

Fig. 2.6 shows the mD signature of a subject punching/retracting spectrogram while standing still. Initially, at time 0.5 seconds a sinusoidal-shaped mD signature starts appearing around the torso; this is due to the movement of the arm in the radial direction of the radar, which is accelerating until approximately 1s where it is brought to rest causing the mD curve to return to 0 m/s. Then negative Doppler is observed between 1 and 1.5 s as the arm is slowly retracted toward the main body, after which the

same activity is repeated at time 2 s. In this activity, the torso movement is restrained compared to walking, and the shape of the mD curves of the arm are not occluded by the motion of the torso or legs as in the walking action. The lower intensity of the limb motion in the mD signature correlates correctly with the relative size of the arm which has a lower radar cross-section than the legs or torso. It is also explained by the nature of the motion, which presents a lower reflective surface compared to the walking action where the arms move along the body.

Therefore, to summarise these simple examples, differences in micro-motions of the body parts can be identified through mD signatures, and the different signatures for different activities shown are perceptibly different. The aspect angle, i.e. the angle between the radar line-of-sight and the trajectory of the target's movement, can also be a significant parameter, affecting how the mD signatures appear and the accuracy obtainable for classification problems.

2.3 Summary

As privacy oriented sensors radar is becoming an emerging technology choice and gaining mainstream focus as a solution. The basic functionality of radar has not changed significantly and it is described by the radar equations. Signal processing advancements, such as the use of micro-Doppler effect and time frequency transforms has allowed analysis of micro-motion of targets with increased fidelity.

This high degree of fidelity with Doppler data has allowed micro-motions of humans and animals to be observed in the radio-frequency electromagnetic spectrum. To realise potential real-life uses of this system, a few off the shelf radar systems are readily available in the market, such as: the Xethru UWB system and the RFbeam portfolio of radar sensors. Together with these hardware and signal processing solutions, radar can be considered a strong challenger for motion sensing for human and animals and in the next chapter the other component of this system, machine learning, will be explored.

Chapter 3

Machine Learning Background

Radar sensors can give rich information from a target, and through joint time-frequency methods, they can deliver a visual description of the time varying motions the target performs or undergoes. However, without ML, the optimal use of this information is not possible. To meaningfully identify the movements from the radar, automatic classification algorithms are essential.

As the current advances in computing continue, with speed, memory capacities and parallelism previously unheard of, increasing applications with classification are utilising ML algorithms adapted specifically to these types of challenges. With this increasing computational prowess, there is an opportunity to make operator-centric applications automatic, thereby reducing human reliance and errors in a cognitive system capable of detecting the motions performed in front of the sensor.

In the context of radar for humans, the ulterior goal is a system capable of distinguishing targets and their movements automatically and using or adapting ML algorithms for radar is seen to be a solution.

Hill [37] states the diverse applications ML covers, which everyone encounters on a daily basis. Learning tasks, such as spam detection and image recognition, together with control tasks such as operating "intelligent" machinery are given as ML examples.

To explicitly describe a machine learning problem, it could be stated as the problem of improving a measure of performance P when executing a task T, through a type of training experience E. Specifically, in learning for an email spam filter, the task T

is to learn a function, which maps any given input email to an output label of spam or regular email. The performance metric P to be improved would be the accuracy of this spam classifier, and the training experience E would be composed of a collection of emails with labels. Alternatively, penalties and weightings could be introduced to obtain performance metric P, for example an algorithm that assigns a higher penalty when non-spam is labelled spam, than when spam is labelled regular. One might also define a different type of training experience, for example by including unlabelled emails along with those labelled as spam and not-spam [37]. In the context of the research conducted in this thesis, the experience would be labelled movement data collected through sensors, the performance metric P to be improved would still remain the accuracy of the classifier and the ultimate goal would be the categorisation of the different activities performed by a human or an animal, automatically.

This chapter overviews the machine learning aspect of this research by framing the activity recognition as an ML problem, then discussing the fundamental versions of algorithms and machine learning architectures used for the solution. By doing so, it explains how the input signature becomes a categorical decision by using the classifier.

3.1 An overview of Machine Learning

To take the description above and generalise it, for any system where an ML algorithm is applied, the sensor information, which can be a sample denoted as x_i and the observation is a denoted as y. x_i is a member of a real set, which referred to as inputs. ML tasks often require an input and an observation and the ML algorithm processes a library of these samples/observation to find a relationship between them.

3.1.1 Disciplines

The dichotomy of machine learning algorithms are defined by the presence and use of prior labels and they are presented as useful approaches for different cognitive applications.

3.1.1.1 Supervised Learning

If an ML algorithm uses available labelled data during the training process then these types of algorithms are considered to be supervised learners. Supervised learners use the labelled data to attempt to minimise the error between the self-generated prior generated, and the metric derived from the ground truth/observation. Simply speaking, it is learning by example [38].

A general element of supervised learning algorithms are that they are often used to predict or identify objects using past patterns such as using historical weather data to predict solar energy usage is an example of regression. This is an example of a supervised learner and similarly so are the classification of activities using prior knowledge, which is the main focus of this thesis.

3.1.1.2 Unsupervised Learning

In instances where labelled data is not available and the observation is simply the distribution of the data, clustering algorithms can be used to separate them. These algorithms, which show the distribution or cluster data are considered to be unsupervised learners [38].

With unsupervised learning the main benefit is that labelled data, which often has to be labelled and verified with significant effort, is not required. This means the data does not need to be overseen meaning the process can be completely automatic. Clusters are not limited to a specific number of classes for example, as in this case the symbolic relationships and the distribution of the data dictate their grouping. However, this can be a problem too as outlying data or data distributions with broad spread can mean a high degree of error for these algorithms.

As there is a focus on high classification accuracy with this work, these algorithms were not utilised but they have promise in future research, particularly with the open class / open set problem.

3.2 Review of physical features from spectrogram

The physical features from the spectrogram are traits extracted from the visible spectrogram image. These are often identifiable with the human eye and have origins in the computer vision or audio recognition literature and communities.

3.2.1 Doppler centre of mass

The Doppler centre of mass, referred to as centroid is the intense peak point of the spectrogram map expressed in frequency native to the radar. It translates to the sum of movements made by the torso and limbs of bodies under the observation of the radar and is expressed as:

$$f_c(j) = \frac{\sum_i f(i)S(i,j)}{\sum_j S(i,j)}, \tag{3.1}$$

3.2.2 Doppler bandwidth

Linked with the centre of mass, the **bandwidth** represents the range of movements made by the limbs. It is defined as :

$$B_c(j) = \sqrt{\frac{\sum_i (f(i) - f_c(j))^2 S(i,j)}{\sum_i S(i,j)}} \tag{3.2}$$

Decomposition based features

Decomposition based features reduce the overall data by mimicking a compression. These methods often extract a large proportion of the input variance using a small proportion of the decomposed elements to form a low rank approximation.

3.2.3 Singular value decomposition

Singular value decomposition has also been applied [39] in efforts to minimise feature selection. SVD decomposes the spectrogram into spectral and temporal content, with

the left and the right singular vectors [40][41]. The first vectors carry the data that contain salient information, but here the extracted subspace has time and Doppler frequency information [41]. This means, depending on the specific nature of the data analysed, which may have greater temporal or spectral variety, one subspace can influence the inputs to the classification model. In the current applications, this approach has been used to increase feature space diversity.

$$M = USV^T \tag{3.3}$$

The first left and right singular vectors contain the spectral projection of the frequency and time axes respectively. The vectors on higher orders contain information on the spectrogram: its shape and possible symmetries.

3.2.4 Features from classic image recognition

As the spectrogram can be considered as a bitmap array, classical image-based feature extraction can provide useful features. Treating the spectrogram as a greyscale image, the different dynamic ranges could be interpreted as different shades of grey. This means the intensities of the torso and limbs now translate to different segments of the grey spectrum and that a histogram can accurately map the amplitude of the movement by those parts of the body.

3.2.5 Image entropy

Closely linked to Shannon's entropy, **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:

$$H = -\sum p(x)logp(x) \tag{3.4}$$

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

3.2.6 Image skewness

Skewness is also a parameter that could be derived from histograms. Statistical moments (mean, standard deviation and higher-order moments) for images have a standard form given as:

$$fm = mean[p(r)] \tag{3.5}$$

$$M_n = \sum_{L-1}^{i=0} (r_i - fm)^n p(r_i) \tag{3.6}$$

3.2.7 Transform based features

Transform based features require a further Fourier transform in the time domain to extract more detailed temporal information. These features utilize the transformed domain to easily extract and assess properties such as repetition and frequency of motions within the original spectrogram signature.

3.2.8 Cadence velocity diagram

By taking the Fourier transform of the spectrogram in time, the repetition or the cadence of the activity is exposed [42][43]. The cadence velocity diagram (CVD) gives the frequencies at which the different velocities repeat and, by using this, the information about the shape, size and frequency of the curves and shapes in the spectrogram created by the moving parts of the target. The features extracted from the CVD are largely peak based. The **step repetition frequency** is derived from finding the local maximum of the CVD, then ranking them in the order of prominence and using the cadence frequency of the prime peak, and **the energy curve** is extracted by approximating the cumulative area under the curve of the CVD.

3.3 Legacy Classification algorithms

As with the features, a diverse plethora of options are present for classification. However, since human m-D provides non-linear distribution of features, the classifiers would

need to be able to utilise this property of the input data. The classifiers in this section have been largely superseded by other more advanced and resource intensive algorithms. However they remain relevant due to their smaller footprint and compatibility with smaller training sets, which is the case with radar datasets.

3.3.1 K nearest neighbour

A simple option is the K- nearest neighbour classifier. It is defined as:

$$\hat{Y}(x) = \frac{1}{k} \sum_{x_i \subset N_k(x)} y_i \tag{3.7}$$

Based on the k closest points, $N_k(x)$, a subset of the training matrix is formed. In order to classify, k observations are taken with the closest neighbours in Euclidean or Mahalanobis distance taken into account [44]. Using these neighbours margins can be generated to segment the data points into different classes.

3.3.2 Linear Discriminant

Diagonal linear discriminant analysis assumes that the samples of classes fit a multivariate Gaussian distribution. This is built on the original Fisher's discriminant rule and in modern applications, during the initial training phase, the mean and the covariance matrix of the distribution is estimated [45]. To expand on the binary theorem, for a multiclass problem a single covariance matrix Σk is calculated instead and an assumption is made that only the mean values alter between classes [46]. The feature space is then split into segments depending on the expected classification cost C, which is linked to the posterior probability $\hat{P}$:

$$\hat{y} = argmin \sum_{k=1}^{K} \hat{P}(k|x) C(y|k) \tag{3.8}$$

3.3.3 Naive Bayes

Naive Bayes classifier also expects a Gaussian distribution for each class [46]. It models posterior probabilities with the Bayes rule:

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)} \tag{3.9}$$

where $p(C_k|x)$ is the probability of class k given x; $p(C_k)$ is the probability of classes and $p(x)$ is for the samples; $p(x|C_k)$ is the probability of samples given the class.

3.3.4 Support Vector Machine

Another more modifiable classification method is the Support Vector Machine (SVM). SVM can be used for separating data into two different classes with a vector containing specific points close to the separating hyperplane. This hyperplane could be defined as:

$$f = x'\beta + b = 0 \tag{3.10}$$

and the consequent minimisation function could be expressed as:

$$min_{\beta,\beta_0}\frac{1}{2}||\beta||^2 + C\sum_{i=1}^{N}\xi_i \tag{3.11}$$

ξ_i is the slack variable and C is the penalty parameter.

If a separation margin, which separates the classes in the given space is not present, a "kernel trick" can be utilised to map features to a higher dimension (where a linear boundary is present). Choices for the kernel are based on the desired hyperplane: gaussian, cubic, quadratic and Radial Basis Functions (RBF) are common choices.

SVM is used mostly in binary classification, as it only allows a single hyperplane within each problem set. To classify multiple activities, error coding needs to implemented. In the one-vs-one coding utilised, each class is comparatively classified with another resulting in an outcome with a certain vote of confidence. The outcome with the

biggest overall vote is deemed the predicted class. In short, an ensemble of support vector machines are required to solve a multi-class classification problem.

3.4 Neural Networks

In a biological brain, within the neurons, the ions from the membrane permeates through an action potential to induce a response. This is thought to be the backbone behind intelligent thinking and it led to early efforts in the 1950s to mimic this process electronically. Similar to how a brain process inputs from different senses, the artificial neural network (ANN) was intended to use salient information from different inputs to make decisions.

This section overviews the different approaches to achieve that end goal with specific focus on the workings of the artificial neuron, the different types of networks and current methods utilised together with radar [47].

3.4.1 Artificial Neuron

An ANN consists of a cascade of layers composed of artificial neurons, which are the fundamental building blocks of an ANN, and connections between these layers. The artificial neuron can be defined as:

$$y = f\left(\sum_i w_i x_i + b\right), \tag{3.12}$$

where w is the weight for inputs, b is the bias, and f is the activation function while x and y are the gate input and output, respectively.

Classically, sigmoid and tanh have been used as the primary non-linear activation functions but since 2010, rectified linear unit (ReLU) have become more prominent. This is due to the sigmoid saturating between [0,1] and and tanh, between [-1,1], meaning they both encounter vanishing gradient problems. Using an unlimited activation function means this issue isn't encountered by ReLU. The activation functions are defined as

below and shown in figure 3.1:

$$sigmoid : f(x) = \frac{1}{1 + e^{-x}} \tag{3.13}$$

$$tanh : f(x) = \frac{e^{2x} - 1}{e^{2x} + 1} \tag{3.14}$$

$$ReLU : f(x) = max(0, x) \tag{3.15}$$

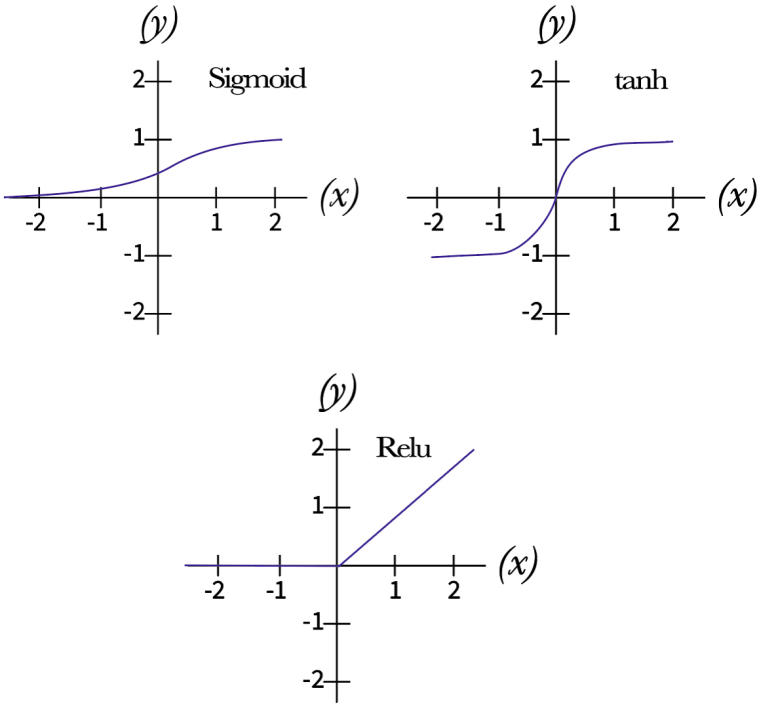

Figure 3.1: Activation functions

To mimic neural behaviour, these non-linear activation functions are necessary as they allow non-linear relationships in the input domain to be mapped.

3.4.2 Multilayer perceptron and Neural Network Training

To learn even more complex non-linear relations, the stacking of neural network layers to create a type of feedforward neural network called a multilayer perceptron is possible.

As the name suggests, a multilayer perceptron (MLP) is composed of a sequence of layers of artificial neurons, as shown in Fig. 3.2. The initial layer is a linearly activated input layer, which is connected to a 'hidden' layer, composed of neurons with non-linear activation functions. This network is concluded with an output layer, which has a linear activation similar to the input layer. The layers themselves are 'fully connected' meaning each neuron is mapped to every other neuron in the subsequent layer. This property allows the MLP to approximate shapes in the input domain making it a suitable architecture for image recognition.

The general form of the network remains as approximated in eq. 3.12, however if the specific classification application is considered, as opposed to regression, which the MLP can be used also be used for, then for k classes the following relation occurs:

$$\begin{aligned} Z_m &= f_{ac}(\alpha_{0m} + \alpha_m^T x) \\ T_k &= \beta_{0k} + \beta_k^T Z, \end{aligned} \tag{3.16}$$

Z is the hidden layer, f_{ac} is the activation function, and α is the hidden weight. For the output layer T, β is weight while m is the number of hidden layers.

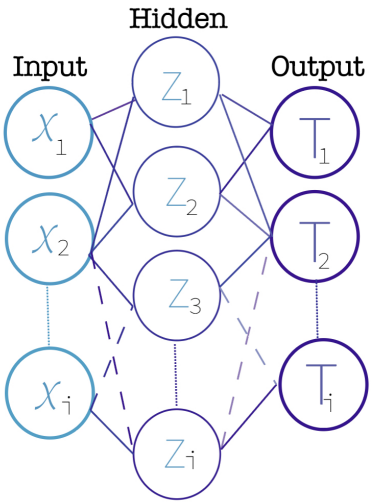

Figure 3.2: Network architecture of an MLP neural network.

The output T of eq. 3.16 is a target measurement which is passed through a softmax layer which takes a vector of T and produces a vector of positive estimates (e^T) summing

to 1, it is represented as:

$$g_k(T) = \frac{e_k^T}{\sum_l^{max(k)} e^{T_l}}, \tag{3.17}$$

With g_k the class identity can then be extracted from the network.

3.4.2.1 Training neural networks with back-propagation

Reviewing eq. 3.12, the weight parameter(w) is the critical element which requires optimal values to fit the training data well.

From eq. 3.16 it can be inferred that the individual layers of the neural network will have sets of weight, α and β and these are linked to the number of hidden neurons M and output classes k. To measure the fit of data, the sum of squared error loss can be calculated using these components:

$$R(w) = \sum_{i=1}^{N} \sum_{k=1}^{max(k)} (y_{ik} - f_k(x_i))^2, \tag{3.18}$$

where y_{ik} is the observation and $f_k(x_i)$ is the classifier prediction. minimisation is the goal here with a desire to avoid the global minimum to prevent over-fitting of the training data. This minimisation of $R(w)$ is commonly performed through gradient descent, also referred to as back-propagation.

For back-propagation, the derivatives of the subset weights α and β are calculated and used to find the current model "errors" δ_{ki} and s_{mi}:

$$\begin{aligned} \frac{\partial R_i}{\partial \beta_{km}^{(r)}} &= -2(y_{ik} - f_k(x_i))g_k'(\beta_k^T z_i)z_{mi} = \delta_{ki} z_{mi} \\ \frac{\partial R_i}{\partial \alpha_{km}^{(r)}} &= -\sum_{k=1}^{max(k)} 2(y_{ik} - f_k(x_i))g_k'(\beta_k^T z_i)\beta_{km} f'(\alpha_m^T x_i)x_{il} = s_{mi} x_{il}, \end{aligned} \tag{3.19}$$

then these errors:

$$s_m i = f'(\alpha_m^T x_i) \sum_{k=1}^{max(k)} \beta_{km} \delta_{ki} z_{mi} \tag{3.20}$$

are used to update the model at steps $(r+1)$ together with the learning rate γ:

$$\begin{aligned} \beta_{km}^{(r+1)} &= \beta_{km}^{(r)} - \gamma_r \sum_{i=1}^{N} \frac{\partial R_i}{\partial \beta_{km}^{(r)}}, \\ \alpha_{km}^{(r+1)} &= \alpha_{km}^{(r)} - \gamma_r \sum_{i=1}^{N} \frac{\partial R_i}{\partial \alpha_{km}^{(r)}}, \end{aligned} \tag{3.21}$$

Eq. 3.21 used to satisy eq. 3.20. This is performed as a two pass process, firstly, the current weights are fixed and the output from the network is computed. Then, during the backward pass phase, δ_{ki} is computed and back-propagated to calculate the error $s_m i$ then these errors are used to calculate the gradient.

3.4.3 Convolutional Neural Network

Although MLPs were initially popular for image classification, the ubiquitous image classifier currently is the Convolutional Neural Network (CNN).

As opposed to the vectorisation of the image which MLPs require, the CNN has a filter which sweeps the image, taking a moving average as it moves across the input image grid. The filter has a stride length S which corresponds to the number of pixels covered by the window during each convolution. This layer of the CNN is called the convolutional layer and it outputs the activation map. A sketch of this layer's behaviour is shown in fig. 3.3.

This process is defined in [38] as:

$$\begin{aligned} s(t) = (x * w)(t) &= \sum_{\inf}^{\inf} x(a) w(t-a), \\ S(i,j) = (x * w)(i,j) &= \sum_{m} \sum_{n} x(m,n) w(i-m, j-n). \end{aligned} \tag{3.22}$$

for the one dimenional and two dimensional cases, where x is the input, and $s(t)$ or $S(i,j)$ are the activation maps.

CNNs have further layers which make the architecture more complex. Specifically, the pooling and fully connected layer are present in CNNs to downsample and vectorise the image data. These layers are notably absent from primitive neural networks.

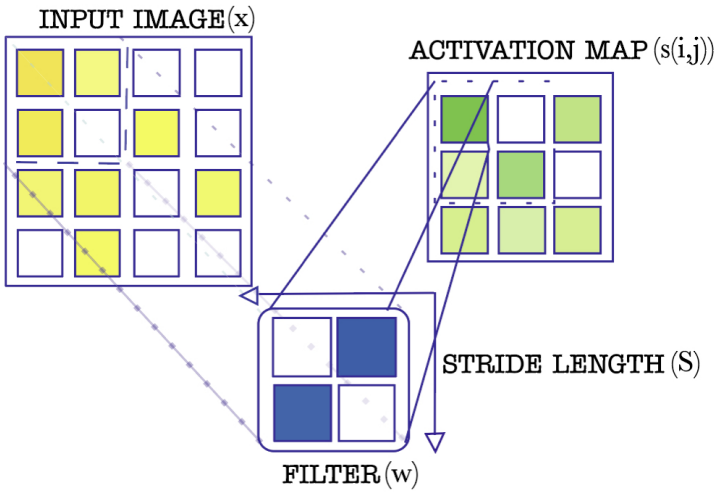

Figure 3.3: Overview of Convolutional layer.

The max pooling layer performs a similar scan of the data at the input however, instead of performing convolutions, this layer takes the maximum value of each segment of the activation map. The output of the pooling layer is converted into a vector by the fully connected layer, flattening the input image into a a single dimensional highly salient information.

The remainder of the network performs similarly to a MLP however there are other architectures utilising further layers and regularization methods such as Dropout.

3.4.4 Regularization and Dropout

Earlier, the importance of optimally getting weights and the updating processed was mentioned and this section will review approaches to ensure the model does not over-fit to the training data. In other words, this means for good general performance, additional steps are required otherwise there is a risk of the algorithm working with high rates with the small training set however with new general data it will not work well.

The simplest methods used in literature are the L1 and L2 regularization [38]. These are introduced as penalties to the cost function when updating the gradient so that with every update the learning algorithm the penalties of certain weights become zero.

To formulate this relationship, take the loss equation for updating a weight β (eq. 3.18) and add the L1 and L2 regularization:

$$\beta_{km}^{(r+1)} = \beta_{km}^{(r)} - \gamma_r \sum_{i=1}^{N} \frac{\partial R_i}{\partial \beta_{km}^{(r)}} + \lambda \sum_{i=1}^{N} \beta_{km}^{(r)}, \tag{3.23}$$

$$\beta_{km}^{(r+1)} = \beta_{km}^{(r)} - \gamma_r \sum_{i=1}^{N} \frac{\partial R_i}{\partial \beta_{km}^{(r)}} + \lambda \sum_{i=1}^{N} (\beta_{km}^{(r)})^2, \tag{3.24}$$

where λ is the strength of the regularization being applied.

Another method to achieve this push away from the global minimum is Dropout.

As opposed to penalizing the loss function with a direct term, Dropout stochastically eliminates neurons from the various layers in the network. This means the updates are 'harder' as the loss increases in noise due to the absence of reliable weights. To represent this, the set $i = 1 : N$ can be taken from (eq. 3.21) then remove certain elements so the set itself is smaller. This reduced set acts as a penalty again as the negative update term ($\gamma_r \sum_{i=1}^{N} \frac{\partial R_i}{\partial \beta_{km}^{(r)}}$) is minimised.

3.4.5 Recurrent Neural Network and Long Short Term Memory cell

For time varying data, although image MLP are utilised, there are other networks that exploit this temporally variant nature better. Recurrent neural networks retain repeating information in the data and updates when a change is detected meaning temporal dependencies between the input and the output can be used for improved classification when the data supplied to it changes in time [38].

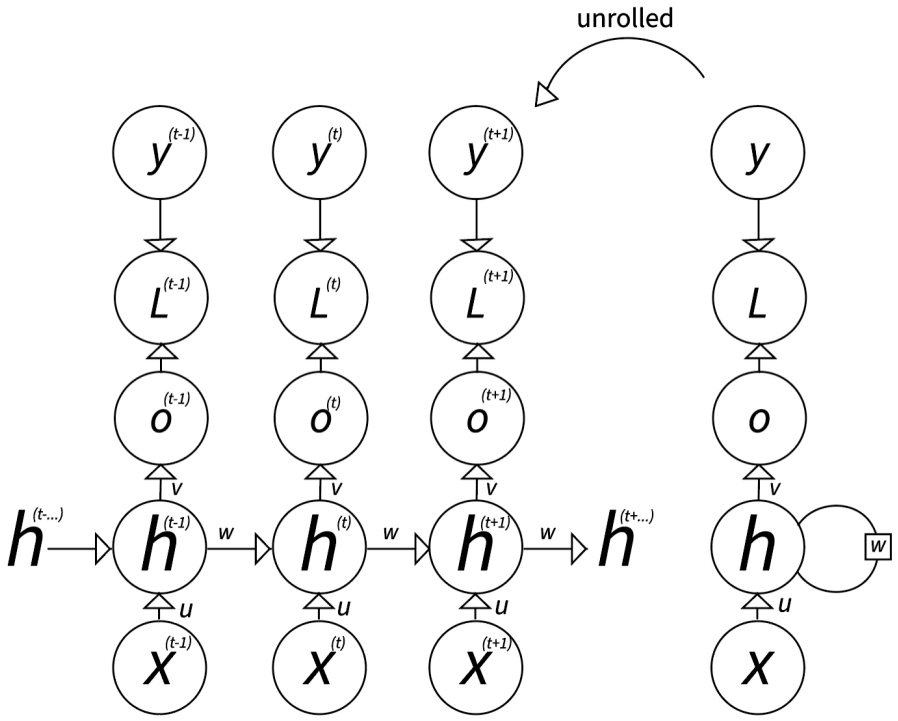

Figure 3.4: Computational graph behind the Recurrent neural network. The unrolling shows the updates as they occur in time.

The RNN, shown in Fig. 3.4 relies on the updates of the y, L, O, and h states. The graph form provides a view of how the input sequence(x) is mapped to the output sequence(y) in this network architecture. The loss(L) maps the difference between the expected output for a given time and it implicitly calculates the softmax loss internally. o is the unnormalised log probability which is the target for the softmax function. U is the weight matrix connecting the input to the hidden weight, whereas w is the weight matrix connecting the hidden to hidden connections forward in time, and finally V is the hidden to output connection weight matrix. When unfolded each of these nodes constitutes as a observation in one time instance.

The forward propagation for this network occurs with the following equations:

$$a(t) = b + Wh(t - 1 + Ux(t), \tag{3.25}$$

$$h(t) = tanh(a(t)), \tag{3.26}$$

$$o(t) = c_t + Vh(t), \tag{3.27}$$

$$\hat{y}(t) = softmax(o(t)), \tag{3.28}$$

Assuming the hyperbolic tangent activation function, as one is not specified in the diagram, for each time step from $t = 1$ to $t = \tau$, the equations above are applied with every update. The softmax operation is used as a post-processing step to generate a vector $\hat{y}$ of normalised probabilities over the output. Forward propagation $h(0)$ is specified.

The implementation of this theoretical process differs from the ones used in various libraries and iterative advancements have been made to the basic RNN to improve its functionality. One of these networks is LSTM.

The LSTM network only implements a forward based dependency in analyzing the data from the sequential timesteps. 3.29-3.32[48].

$$f_t = \sigma_g(W_f x_t + R_f h_{t-1} + b_f) \tag{3.29}$$

$$i_t = \sigma_g(W_i x_t + R_i h_{t-1} + b_i) \tag{3.30}$$

$$g_t = \sigma_c(W_g x_t + R_g h_{t-1} + b_g) \tag{3.31}$$

$$o_t = \sigma_g(W_o x_t + R_o h_{t-1} + b_o) \tag{3.32}$$

The inner process of the gates in an LSTM layer are given by equations above. Equation 3.29 shows the operation of the forget gate and is based on an activation function applied on the sum of the weighted input (with weight W and input x) with the product of recurrent weights R and the hidden states h from the previous iteration, and a

bias term b. The other gates perform in a similar manner, with differences occurring from the input and recurrent weights, together with the bias being unique to each gate. The sigmoid activation function, cell state output, and hidden states output are represented by equations:

$$\sigma(x) = (1 + e^{-x})^{-1} \tag{3.33}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{3.34}$$

$$h_t = o_t \odot \sigma_c(c_t) \tag{3.35}$$

To understand how the use of LSTM based networks for time-dependent data is beneficial, the LSTM cell behavior can be empirically described. Fig. 8.1 presents a sketch of the LSTM cell showing the two outputs at the current timestep t: the hidden state h_t and the cell state C_t. h_{t-1} is the hidden state at the previous timestep while C_{t-1} the cell state at the previous timestep.

These two signals, together with the input data at the current timestep X_t, are the input signals to the LSTM cell. Meaning, the outputs at the current timestep depend on the hidden state and cell state from previous timesteps, therefore, the memory of the network is utilised. This process mirrors the transition which are presented in the previous set of equations.
The original recurrent neural network architectures, before the development of LSTM, did not have states. Therefore temporally pertinent information across many timesteps was not retained; the cell state changed this, as longer time-based dependencies could now be memorized.

Four components control the two outputs:

- f is the forget gate which resets the state of the cell making it forget prior information from the previous cell state.
- g is the cell candidate which provides input to the cell state keeping memorable or recurrent information and providing it to the cell state.

- i is the input gate which co-ordinates with g.
- o is the output gate to control the addition of the cell state to the hidden state.

In terms of radar data, this means that the information on human movements can be memorized and correlated over a relatively long time. In the Doppler-time representation (spectrograms), an activity in a sequence of movements is perceived by the radar as a specific pattern of active Doppler bins over time. The network can learn this pattern in its internal parameters to recognize this activity even when it has different lengths of 'activation' or delays.

3.5 Machine Learning Concepts

Machine learning related techniques include the algorithms discussed in the previous section, with many more available in the literature. However, they all follow concepts which are uniformly present for all types of classifiers and estimators, and this section delves into some of these shared ideas that are essential in machine learning.

3.5.1 Fitting

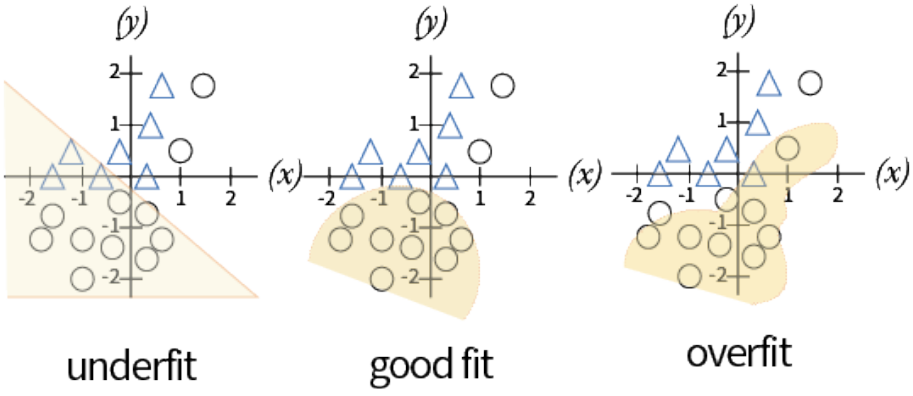

Figure 3.5: Intuitive overview of underfitting, good fit and overfitting.

The concept of fitting derives from the weights and biases discussed in this chapter. Excessive bias can change the margin where two independent classes are detected. In

fig. 3.5 a two dimensional example is given where a detector for the circle class draws the separating margin for three potential outcomes of the fitting process.

When the classifier **underfits**, the margin will exclude observations which are outlying the primary cluster of points. This means that some observations will not accurately be detected while observations of the other class will be incorrectly identified as belonging to the class of interest. This occurs due to a relaxed bias parameter and it can lead to a low classification accuracy with high false positives.

Instead, when the classifier has a **good fit** although the primary cluster will remain the main segment of interest for the separation margin, the margin itself will be tighter around the class of interest. This means points which are outwith the primary cluster will not be detected similar to the case when it is underfit, however false positives will occur less due to the new tighter margin.

In cases where the classifier is overzealous with the coverage of the separation margin and it attempts to cover outlying points, this is called **overfitting**. Overfitting is special case which is of concern as it initially often leads to high rates of detection for the class of interest. However, there is an increase in the number of false positives as the margin is wider and overlaps with the cluster of the other class. Notably, the model will not necessarily cover outliers.

In machine learning applications, during the initial part of the training process the classifier usually is either overfitting or underfitting with a large number of training iterations required to converge to a good fit. This iterative fitting process, together with the gradient update process is what is understood as training the classifier.

3.5.2 Validation

The fitting process requires a test with a-priori data to establish how successfully the data can be classified and there are two methods of validation, or testing, which fitting is performed through: k-fold validation and holdout validation.

Holdout validation is the most common method of validation in literature where a ratio of the total data is selected to be the test set. The indices of this test group is selected randomly and over the entirety of the data set, meaning all of the classes will have

a similar number of observations in the testing set which corresponds to their overall numbers in the dataset. This is performed randomly to avoid bias and it provides representative results when the number of samples in the data set is large.

In K-fold validation however, the data-set is split into K groups, an ensemble of learners are used to train on $K-1$ groups, with the last group used for testing, as shown in Fig. 3.6. This is useful in cases where the total dataset has observations in the terms of hundreds (hence small, like in many radar-based classification problems), instead of the thousands which is the volume commonly encountered in classical machine learning applications.

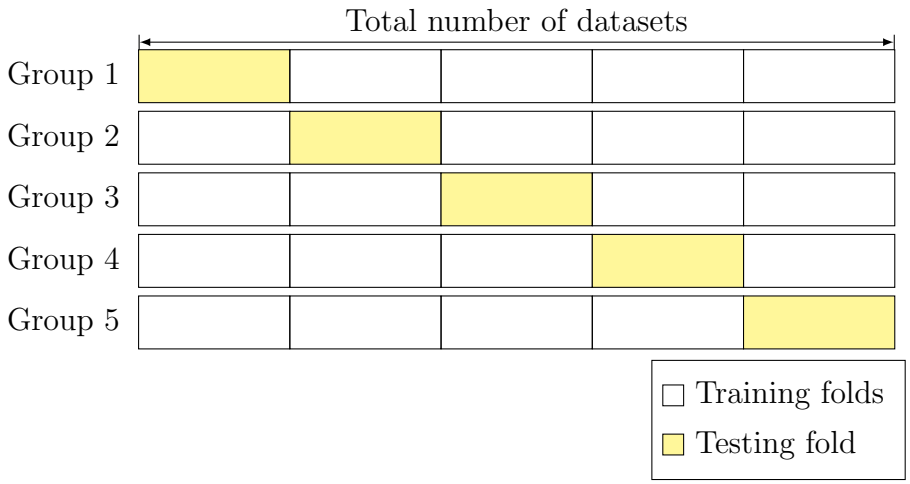

Figure 3.6: K-fold grouping illustrated. At every step the active testing fold is changed while the remaining folds are used for training.

The methods of validation is decided by the abundance of data and in many of the forthcoming sections there will be different approaches of validation used and an insight will be provided behind their use.

3.5.3 Measures of success

To evaluate the performance of the classifiers, metrics can be derived from the correct classification and misclassifications events in confusion matrices, as shown in the simple binary example in table 3.1. The diagonal elements indicate the correct identification for a given class of interest (e.g. A). This is represented as a True Positive (TP) and

when the other class which is not of interest is classified correctly (e.g. B or C), whereas a false negative (FN) indicates a 'missed detection' and a false positive a 'false alarm'(FP).

Table 3.1: Multi-class classification confusion matrix example

Class	A	B	C
A	TP	FN	FN
B	FP	TN	TN
C	FP	TN	TN

In addition to the rate at which the classifier accurately detects an activity or label, there are further metrics which are used to assess the quality of the classification.

$$Precision = \frac{TP}{TP + FP} \tag{3.36}$$

$$Recall/Sensitivity = \frac{TP}{TP + FN} \tag{3.37}$$

$$Specificity = \frac{FP}{TP + TN} \tag{3.38}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.39}$$

Precision (eq. 3.36) denotes the correct classification rate with respect to false positives. Recall (eq. 3.37), also known as sensitivity, is related to the ratio of correct class detection with respect to false negatives and specificity (eq. 3.38) is calculated to measure the 'false alarm rate' of the classifier for that specific class. Accuracy (eq. 3.39) is the balanced correct classification rate considering all classes.

3.6 Summary

ML is necessary to infer information from radar sensors for automatic recognition as it removes the central component of radar operation in the processing chain, the operator. In the classical interpretation of machine learning problems, a radar classification could be defined as the task; the rate of classification as the performance and the human operator having the training experience. When it is applied for activity or condition recognition, it is an example of supervised learning as the abstract detail, be it movement such as walking or condition such as gait, is a categorical concept which requires past labelled data and the experiments and research done for this thesis consist of this discipline of machine learning.

Over the course of this research, the state of the art has rapidly changed from algorithms such as linear discriminant and SVM, algorithms which could be considered to be legacy, to holistic algorithms such as artificial neural networks, to intelligent state tracking algorithms as LSTMs.

Although the algorithms and architectures have changed dramatically, with newer algorithms requiring increased amount of computational power and long duration of training time, the basis for performance has remained constant. This is usually done through the bulk accuracy of classes identified although there is a renewed focus on the resulting metrics such as precision, recall, and specificity.

As with the general machine learning community however, recently the focus is shifting on from the architectures to the type and quality of input data which will be the focus of the upcoming chapters.

Chapter 4

State of the Art: Monitoring activity using Radar and Micro Doppler

Since the early days of radar, initial experiments were performed to assess if radar could be used on humans for diagnostic purposes. However, half a century later in the early 2005s, radar began to be used for activity recognition and other human position and movement-related applications. Since then, between 2015 and 2020, due to advances in sensing, machine learning techniques and experimental design, activity recognition with radar has attracted significant interest. This chapter provides an overview of the literature in this timeline and it focuses on techniques to improve activity recognition performance using radar. It also shows the gaps in the current knowledge, where opportunities to contribute to the overall activity recognition with radar are present and open.

4.1 Radar's place in the activity detection paradigm

In Chapter 1, we discussed the maturity of established sensors in the space of activity recognition, which then led to the question: why use radar?

In general the activity recognition hardware is categorised into two groups: wearable and non-wearable sensors [49]. The former are either affixed to the body parts of the monitored subject, or they are worn and carried in pockets. These sensors take fine

resolution data from the inertia of the human torso and limbs, characterised through their namesake sub-sensor: acceleration with the accelerometer and magnetic field with the magnetometer, with the addition of gyroscopes for angular movements. They are also capable of direct measurement of the surface temperature, arterial movement for vital signs.

Non-wearables on the contrary are at a distance from the targets, being less invasive regarding interaction and requiring less management and effort by the end-users and deployers, who may often be older people with impaired cognition [49, 50, 51]. Furthermore, they are also less reliant on temporary power sources such as batteries, which is an added benefit management-wise.

Among non-wearable sensors, radar has had increased attention as a possible alternative to established sensors such as video-cameras, due to its insensitivity to light conditions and easy integration into the end-users' home environment, as modern radar systems can be inconspicuous in the modern smart home (for example, just like a common Wi-Fi router).

Furthermore, radar may present less privacy issues than cameras, as plain images or videos of the end-users and their private environments are not collected [52, 53, 54, 55, 16].

In summary, radar is a viable home monitoring device for activity recognition with a long list of benefits. However, the question remains about **how to leverage radar technology with micro-Doppler and beyond for healthcare applications, specifically activity recognition?**

This chapter encompasses the state of the art attempts at hardware, software and experimental research and innovations to answer this question.

4.2 Early experiments involving Humans and animals with radar

Since V. Chen coined the term mD [36] , the study of using radar with micro-Doppler for detecting, tracking and monitoring activities from living beings has steadily grown

over the recent years. However, there was interest in human detection with radar as far as 1958 [56] where Schultz et. al. identified the radar cross section of a "man" with both the monostatic and the bistatic radars at different frequencies ranging from 410 to 937 MHz, and both horizontal and vertical polarisation. In this experiment, a CW radar was used to detect the cross section of a man who weighted 200 pounds and was six feet tall. Then other men were 'measured' and ultimately it was found that the man's cross section was approximately proportional to the weight of the target with polarization having an effect on the measurements. This experiment was the earliest indicator of physiology affecting the radar signature of a subject setting out a research theme which would be influential in the far future: **how the subject physiology affects the detection of the target,** which we attempted to address in [29].

In [57], utilising the principals used in eq. 2.1, a radiometer was used to detect the reflectivity of different targets. Specifically, a pedestrian, a cyclist, and a car were measured to have a varying level of returns as targets.

Until the millennium, experiments such as these were the limited examples where radar was used with humans however this changed with the advances in computation which allowed processing required for time frequency transforms, the mD effect and more accessible commercial off the shelf radar systems.

One of the earliest examples of using time frequency with a radar system was done in [58] in 2002, where Geisheimer et. al. used a CW radar to record the signature human gait. They recorded the legs, arms, and torso which motions have different relative velocities through the gait cycle and most notably, used STFT to parameterise the human gait. This was seminal in the development future research in that field. Other works focused on the angle of arrival [59] and through the wall detection of individuals [60].

Between 2005 and 2012, the focus was mainly on detecting the target but it is also a significant period of time where interest in radar for human applications were beginning to attract interest in the wider radar community. An important advent during this period was the introduction of the mD effect [61].

Otero assessed the presence of participants using a radar in [62]. In this instance it was a simple classification problem of identifying the presence of a male subject and a

female subject and their overall absence from the line of sight of the homebrew radar system. Notably, the mD signature of a dog was compared with in this paper but not utilised for experimentation or classification. This shows that there were early interest in the applications of radar based activity recognition for domestic animals.

The next notable contribution was by Kim and Ling in [16] where experimental and analysis techniques presented were adopted by a large number of future works. In this paper the analysis ranged from feature extraction, non-direct line of sight and through the wall performance to difficulty in classifying sequential data and cross validation options.

The experimental makeup contained the use of a single CW radar at 2.4 GHz with which seven activities were classified with the data from 12 participants.

In certain ways, this paper highlighted the potential research avenues for mD activity recognition. These formative topics ended up being the focus of research over the next decade, including this work.

4.3 Specialisation into classification and discriminative applications

After these clear opportunities were outlined in literature, the research focus began to diverge and propel into the newer realms of discriminative applications such as activity, person and gait recognition.

4.3.1 Feature extraction for recognition

Feature extraction was an important focus during this period and methods of extracting salient features were explored in the literature. Starting from the generation of fine handcrafted feature which extracted direct movements, the process was refined over time so features were generated automatically and in this segment overviews this iterative process which was ultimately superseded by deep learning methods.

One of the main features utilised for activity classification is the Doppler centroid which was generated in [63] to identify armed and unarmed personnel. Similarly, in

Table 4.1: Table of formative experiments of radar use on humans.

Paper	Year	Radar	Purpose	Participants
[56]	1958	CW	Measuring the RCS of different participants with varied body and sizes. Ultimately stating RCS is proportional to body size and weight.	5 males
[57]	1987	Radiometer	Radiometers can be used to discern between cyclists/pedestrians and cars	2 unknown
[58]	2002	CW	legs, arms, and torso which motions have different relative velocities through the gait cycle which is visible in a spectrogram	1 unknown
[59]	2007	CW	Demonstrates the TF diagrams of human gait and provides functions for I-Q imbalance.	10 unknown
[60]	2005	CW	Presented through-the-wall micro-Doppler signature of a person walking	1 unknown
[61]	2002	CW	explored the micro-Doppler effect and provided examples of human movements observable with the radar.	unknown
[62]	2005	CW	first use of classification algorithm with gait data from radar, also shows the gait of a dog.	2 one male and one female
[16]	2009	CW	demonstrate signature of multiple human activities and outline challenges for radar based activity recognition with using support vector machine. Continuous activity challenge also proposed.	12 unknown

[43] Seifert et.al. used features generated from [62] to classify three different gaits to check presence of assistance from a cane. These two works were an early example of not only generating fine handcrafted features but also using them together with component analysis or decomposition based feature generation methods.

An early example of these methods was from Chen, who used independent component analysis (ICA) to extract features in [64]. Using another component analysis technique, in [65], Clemente et. al. used robust principle component analysis (PCA) to improve classification accuracy and as a tool for dimensionality reduction. The singular value decomposition (SVD) of the spectrogram was also an innovative method to generate features and reduce the components of the spectrogram to isolate salient ones. Earlier examples of this method were synthesised and applied by De. Wit in [66] and Fioranelli in [63].

None of these aforementioned methods, i.e. ICA, PCA and SVD, require precise derivation of information and modifications. Furthermore, they do not require fine-tuning of the algorithm to specifically suit a set of radar data. They are resilient to various levels of signal-to-noise ratio and produce salient features in most cases. It is notable that these works innovated or generated new methods of extracting useful information from the highly abstract radar data, showing that **feature extraction influences the accuracy of the classification**. It becomes clear that there are many methods of generating features which are distinctly different from each other. This thinking was behind [18, 19, 20], where the variety of sourced data was utilised to select optimal features from these vastly different sources.

4.3.2 Gait analysis

In 2014, Wang et. al. in [67] conducted 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. This is one of the few examples where the location of the test is considered as although this is an important factor, it is not highlighted upon. In this experiment, they generated gait parameters from radar spectrogram signatures and classified four types of gaits from 13 participants. In [43], as mentioned before,

three gaits were assisted, and similarly in [68], the authors considered five different classes of gaits.

In [69], pseudo-Zernike moments derived from spectrograms were used to classify firstly, between gaits of individuals and groups then secondly, between horse riders and individuals.

The topics of detecting maligned gait and classifying presence of animals in signatures showed how in this period, new applications were continually introduced. Using these two overarching topics in [17], we utilised **radar based gait recognition to detect lameness in animals which are physiologically different from humans.**

Gait monitoring and activity recognition with radar both started to catch up to the state of the art methods like deep learning around in this time and considerations such as number of participants began to be important as researchers pushed to increase the significance of their experimental results.

One of these examples is in [70], *Le et. al* used autoencoders together with spectrograms and scalograms to identify the gait of participants who were walking. While in another example [71], an LSTM was used to classify the walking gait of small groups of people vs individual persons in an outdoor scenario. As for experimental improvements, [72] who validated mD gait signatures as precursors of cognitive ability using data from 74 people. This was one of the larger scale evaluations which truly tested the aspect of participant physiology for gait recognition.

4.3.3 Activity recognition/ gait monitoring with deep learning

After the initial focus on feature oriented classification the development of deep learning [73] and related classification methods based on neural networks, attracted significant interest for their application to radar-based monitoring of human activities [74]. Their main advantage was the ability to extract salient features automatically within the network, without explicit inputs or fine-tuning of parameters preventing features which may be prone to overfitting.

There are numerous contributions in the literature using deep neural network (DNN) based architectures to process the radar data as images. The work in [75] used DNN

where they classified specific individuals and groups of individuals based on their walking gait. They also provided comparison with conventional supervised-learning classifiers such as Naïve Bayes and SVM were provided, demonstrating better performances when using the deep networks. A DNN was also used in [55] for human gait recognition, exploiting a dual-channel architecture where the network had two separate branches at the input, in order to accept spectrograms calculated with different temporal resolutions. A specifically designed DNN was also used in [52] to identify specific individuals in different rooms based on their walking gait, with the additional complexity of the subjects following free-form, unconstrained trajectories.

Then on, even more complex DNN architectures began seeing use in literature and [32] provides examples of these works.

In [76], DNNs were used to classify human activities from their spectrograms, and in [77], a novel DNN based architecture was proposed to specifically account for the diversity induced by the different aspect angles on the radar signatures of human movements, especially with respect to their Doppler signature. Modifications to the conventional architectures of DNNs were proposed in [74, 78, 79], exploiting convolutional auto encoder (CAE) to perform unsupervised pre-training of the weights of the network.CAE and DNN were also combined with a novel technique to augment the amount of available data in the training set by using Kinect-based motion caption simulations, enhanced by a diversification technique to improve the life-likeness of the simulated synthetic data.

Our application of deep learning was focused on the unique challenge of using data from a variety of radar sensors. In [28], we used transfer learning to classify activities but also showed that transplanting weights generated with a radar with one centre frequency can be used to classify data from another independent radar system operating at a different frequency.

4.3.4 Multi-sensor fusion with radar

Due to the maturity perhaps, the other sensors have seen the adoption of sensor fusion in numerous cases. One of the factors behind this is that inertial sensors, which are

often commensurate and co-sampled at the same interval, has multiple co-sensors which are weak on their own but strong when combined.

Particularly the combination of ambient and wearable sensing is not present in the literature. There were initiatives such as the sensor platform for healthcare in a residential environment (SPHERE) project which utilised a sensing environment containing sensors measuring traditional activity monitoring [80] but also unique information such as: ambient temperature, ambient kitchen cooking temperature and network allocation used by the participants living in the SPHERE house. In these data-sets the individual sensors are the mature conventional heavyweights such as stereo cameras and wearable sensors. This combination of data collected could be considered almost excessive and the element of privacy would again be critiqued. A common question asked is **How can the sensing methods be improved to optimise the classification accuracy in the context of assisted living?** However, in this question is not limited to the literal sensing methods but also the optimal number of them and the type of data recorded. It means to ask what type of data and how much of it should be collected while maintaining the privacy of the test subject.

In [20], we give an overview of the different opportunities for data fusion and showing improvements for our set of daily activities. Although radar performance at different frequencies has been thoroughly evaluated by [74] the co-operation of radar at different frequencies did not appear in literature. We also addressed this secondary issue in [20] by demonstrating co-radar fusion.

4.3.5 Continuous activity sequences

It is apparent that activity classification and gait recognition was a popular topic in literature, the use and classification of lifelike sequences or activities with transitions were missing from the literature.

Notably during this time, LSTM was minimally discussed in the literature as a standalone tool for radar-based human activities classification. Therefore it represented an under-explored approach compared to the established deep neural network based approaches.

In [81, 82], recurrent networks had been used to classify six different human activities. An LSTM was used in [82] and a stacked gated recurrent unit network, based on a simplified architecture of the LSTM cell, was used in [81]. Although these works addressed the absence of the LSTM cells, for these experiments the data was collected as separated individual discrete recordings which were then concatenated to generate a sequence. This conventional snapshot data collection was also applied in [83] and [84], where LSTM networks were applied respectively to raw IQ radar data and to range profiles to classify separated human activities.

In short there was a gap in the literature where attempts at utilising **advances/evolution in machine learning to automatically segment and classify human activity in a continuous data stream with a single radar sensor** had been performed. However they had the shortcoming of not having genuine sequential data, and in certain cases not addressing activity recognition.

In [21], we address this by taking genuine sequence of activities then using a novel bidirectional LSTM based network to **demonstrate improved recognition of continuous activities and activity transitions without feature extraction** using a single radar sensor.

Table 4.2: Previous works for activity recognition and gait monitoring with radar.

Paper	Year	Finding	Radar	Classifier	Accuracy(%)	Participants	No. of classes	cross-validation
[16]	2008	Demonstration of activity classification using ensemble of SVM	CW	SVM	90	12 unknown	7	k-fold: 4
[63]	2016	Classification of armed personnel using mD signature's centroid, bandwidth and SVD.	FMCW	NN, DA, decision tree	98	3 unknown	3	80% holdout
[43]	2017	Classification of active use of cane for walking assistance	FMCW	NN	84	3 unknown	3	30% holdout
[64]	2004	Discussing theory and application of using PCA, ICA for feature extraction and SVM, LD for classification.	-	-	-	-	-	-

[65]	2013	Demonstration of a robust PCA method for feature extraction	CW	SVM	93	>3 unknown	5	30% holdout
[66]	2014	Use of singular vector for feature extraction	CW	-	-	-	-	-
[32]	2019	A state of the art survey of deep learning methods for activity recognition with radar. It highlights the current challenges such as continuous activity recognition.	FMCW, CW, UWB	CNN, LSTM, CAE	-	-	-	-
[70]	2018	Using deep autoencoders to classify three types of gaits and use of arm swing.	CW	CAE	96.2	20: 12 males 6 females	3	k-fold: 6
[67]	2014	Extraction of gait parameters from micro-Doppler signature and comparison with a motion tracking camera.	Pulse	-	-	13: 7 males 6 females	-	-

[68]	2017	Comparison between different types of sensors for walking aid gait detection	CW, Pulse	SVM, NB	80	unknown: males females present	5	30% holdout
[69]	2015	presented pseudo-Zernike moments as a method of generating features from humans and horses and classification between them	Pulse	SVM	95	>3 unknown	5	30% holdout
[71]	2005	Demonstration of through the wall angle of arrival and micro-Doppler signatures	CW	-	-	unknown	-	-

[72]	2019	Extracts the gait information of a large number of elderly participants above the age of 75 and correlates gait parameters with cognitive ability.	CW	-		74: 29 males 45 females	-	-
[74]	2019	A case study of different deep learning classifiers for activity recognition. It compares convolutional networks, classic methods and transfer learning.	CW	CAE, DNN, SVM	95	11 unknown	12	k-fold: 10
[75]	2018	Individual identification in a group of people varying between4 and 24 individuals.	CW	DNN	63-95	24: 12 males 12 females	4-24	-

[55]	2019	Classification of 3 gait patterns generated through a motion capture database.	Mo-cap	DNN	96	simulated	3	30% holdout
[52]	2018	demonstrates the use of DNN to identify gait of different participants in a group	FMCW	DNN	79	5 unknown	5	20% holdout
[76]	2016	A deep network based approach to identify human participants from other targets and activity classification.	CW	DNN	90.9	12 unknown	7	k-fold: 4
[77]	2020	Difference in aspect angle dependency between simulated and real movement observed by radar classified using a DNN.	UWB	DNN, SVM, KNN.	98	7 unknown	6	33% holdout

[78]	2018	Introduction of convolutional autoencoder as a feature extractor and classification with a CNN, and comparisons with an SVM	CW	CAE, SVM	94.2	11 unknown	12	10% holdout
[79]	2018	Demonstration of transfer learning for activity recognition supplanted with kinect generated augmented micro-doppler signatures.	CW	DNN	96	unknown	11	20% holdout
[80]	2016	Dataset for heterogeneous fusion utilising wearable and remote. sensors	-	-	-	-	-	-
[81]	2018	stacked gated recurrent units for activity recognition	CW	RNN	98	2 unknown	6	k-fold: 4
[82]	2019	Classification of synthetic sequence of activities using a stacked LSTM based approach	CW	RNN	92	unknown	6	k-fold: 4

[83]	2019	Use of radar IQ data and LSTM network for activity classification.	FMCW	RNN	93	7 unknown	6	25% holdout
[84]	2019	Use of radar range profile and bidirectional LSTM network for activity classification.	Mo-cap	RNN	90.3	unknown	6	k-fold:5

4.4 Summary

This chapter discussed the state of the art regarding how to leverage radar technology with micro-Doppler and beyond for healthcare applications, specifically activity recognition.

In table 4.3.5 previous works discussed in the chapters are identified and summarised. There are a series of trends which appear in this table with the first being the predominant use of the CW radar with other systems either used for comparison or as alternative. Secondly there is a lack of information about the participants for the majority of the works with age and physical information being the most absent traits in the papers. Finally, there is no clear correlation between the number of classes and the type of classifiers used and the set of activities are different depending on the authors. A positive trend is that the current methods already show exceptional performance with the specific subset of activities they are applied to. There is also a general increase in the number of classes which means the classification challenge is progressively increasing. However, besides these observations, the cumulative works up to now have generated a few research questions of contention.

Chapter 5

Methodology and datasets

Given the breadth of experimental variations in the activity recognition with radar literature, there are many potential activities which could be the focus of any study. However, certain types of activities are more challenging to solve at a distance with an ambient sensor. This chapter will look at the initial selection of these activities and the iterative changes of the discrete activity set to reach the maximum complexity in the form of continuous activities analysed in this thesis.

Additionally, this chapter will also overview the setups of the experiments conducted for this thesis and discuss the differences between previous experiments in the literature. It will also explain the operational similarity between these independent experiments while addressing their evolution towards lifelike environments.

5.1 Activity decomposition of initial set of Activities

The categorisation of the activities had a specific purpose, as in most activity detection system, there are certain actions which are more important to be classified. The initial set of activities was inspired by this line of thinking together with previous data sets such as the ones presented in [16]. The set should represent activities, which use specific parts of the body as well as also include movements from the whole body. This means commonly made movements would be included in the set together with rare but dangerous activities such as falls. False alarms are a key concern in fall detection

systems as valuable resources could be unnecessarily used up following a false fall identification, therefore the broadness in the number and types of activities was a key desire.

There was also a desire to keep consistency between the activities; therefore similarity between sets was ideally maintained. Although the initial focus was on activity recognition of discrete type, the focus at the end changed to continuous activities which was a shift that was present in the literature too towards more realism.

5.2 Experimental design decisions

Although there is a difference in the sensors, participants and at times, activities between the activities, the experimental setup remains uniform between the experiments considered in this thesis.

The experiments involve a single participant performing activities in the field of view of a radar sensor located at a close distance. In all the experiments, a time window was provided where the participant could decide to start and finish their activity, meaning there was an intrinsic variation between the recordings. Furthermore, the participants performed the activities with repetitions to ensure multiple recording samples were present from one individual. Finally, no specific instructions were provided to the participants, meaning the movement made was based on their interpretation of the activity stated.

The reasoning behind these decision comes from factors which activity recognition applications would face in a real world scenario. For example, the sensors were located in a close distance from the participants as the application of ambient sensors are predominantly indoors. Similarly, for the open time window, different individuals depending on age, capacity, and physical state perform movements and activities at different speeds. In addition, as the activities are often a sequence of smaller motions, individuals can have "quirks" or unique patterns which could be hard to replicate, yet an activity recognition system would need to identify this in the real world. Finally, the lack of prescribed movements for the activities would further add to this variety.

5.3 Experiment details

This section will detail the variables in the experiments such as sensors, number, and gender of participants and operational parameters of the radar systems.

5.3.1 Radar sensors setup

The main radar system used for the experiments was an FMCW radar system from Ancortek operating at 5.8 GHz, an instantaneous bandwidth of 400 MHz and a pulse repetition frequency of 1 kHz. This radar was connected to separate transmit and receive Yagi antennas with transmitted power of 100 mW, with an antenna gain equal to 17 dBi approximately and beam width of 24° in azimuth and elevation.

A colocated radar was also used in certain experiments, and this was the RF-beam st100 with K-MC1 transceiver unit that transmits approximately 18 dBm EIRP (Effective Isotropic Radiated Power) at 24 GHz. The transmitter and receiver antennas were micro-strip patch antennas (gain 18.5 dBi) with beam aperture 25° in elevation and 12° in azimuth. This system is a CW radar therefore it has no instantaneous bandwidth.

5.3.2 Inertial sensors setup

Other sensors were also used cooperatively as these they either measured parameters which were different from the ambient sensors such as inertial sensors, or they measured higher fidelity information such as depth cameras. The main counterpart were wearable sensor-based devices. Compared to contactless sensing methods, wearable devices use the fact of being directly attached to the user to map the motions by recording the variation in physical characteristics of the body under movement. These characteristics are quantified as acceleration, angular speed and magnetic field strength by recording through an Inertial Measurement Unit (IMU). Classically, inertial and magnetic sensors are placed on the wrist of the body to measure the acceleration, angular speed and magnetic field change in the different axes with respect to different human activities. However there are instances where they have also been placed along the waist or aside

the legs. This is a benefit that radar systems do not have as there is a potential to be selective about sensor placement.

The accelerometer acts like a displaced mass on a string, where upon acceleration of the monitored body it experiences a fall and the corresponding acceleration recorded is estimated by the displacement of the string [85] and proportional to the overall movement. In the commercial market, piezoresistive, piezoelectric, and capacitive components are used to convert mechanical displacement into an electric voltage [86]. Piezoresistive materials are ideal at measuring sudden changes of high acceleration, whereas piezoelectric materials are sensitive to the upper frequency range and are temperature tolerant. On the contrary, capacitor-based accelerometers are sensitive to low frequencies.

Gyroscopes estimate the angular speed and maintain the direction to the poles [87][88]. The gyroscope is typically used together with the accelerometer, constructing the inertial navigation system. The main gyroscope frame consists of a gimbal and a rotor, where the spin axis is free to represent any orientations without interference from tilt and rotation. Modern gyroscopes are made of Micro-electromechanical systems technology, which allows packaging multiple gyroscopes for different axes in a single physical chip.

Magnetic sensor or magnetometer, often categorized into magnetic Hall Effect sensor and magnetoresistance sensor, can detect weak bio-fields inside the human body. It can be further categorised into anisotropic magnetoresistance , giant magnetoresistance and tunnel magnetoresistance [89].

Hall sensor is widely used for human activity recognition due to its sensitivity range, whereas magnetoresistance sensors can capture subtle variation of magnetic field (10-6- 10-12 Tesla) via an array structure. The magnetometer moves across the earth magnetic field when part of the human body is moving in a 3-D space; different voltages are produced from the conductor according to the amplitude of the motion and aspect angle to the earth magnetic field [19]. This is known as Hall-effect, where the applied magnetic field can be expressed through applying a floating electric current on the conductor to produce a hall voltage. This allows the hall-effect sensor to convert different human activities into difference in output voltage and turn it to the magnetic information accordingly. The magnetometer is a single Bosch BMM150 Hall sensor

sampling at 20 Hz with a ±1300 µT range and resolution of approximately 0.3 µT. It is a sub-component of the IMU produced by X-IO technologies containing other inertial sensors.

The use of these sensors in the experiments was due to the desire for uniformity between the experiments. To a certain extent, this meant that the main radar sensor was a control variable central to this research, while the other sensors were explored as avenues to assist the main radar in improving the activity recognition rates in a multimodal fashion.

5.3.3 Hardware considerations

Power requirement is an important parameter for the different sensors. As the wearable sensors are located away from the processor, they require a battery-based supply to operate the sensors and transmit the data. However, the radar system does not have this detriment as it is connected to a computer for operation and recording, therefore the power source and consumption is less important compared to the inertial sensors. For regulatory compliance, radars are powered by USB operating at 2.5 Watt maximum. The operational transmit and receive power are in the order of +10dBm which is within the range of wireless electromagnetic networks such as WIFI.

5.3.4 Environmental, participant and activity setup

This section discusses the composition of the experiments and how one set of experiments influenced the next.

Dataset 1

The first datasets for this thesis were generated in 2016 in two different locations. Dataset 1 was performed in an indoor meeting room at the School of Engineering at the University of Glasgow.

The environmental setup mimicked a normal living area containing several pieces of office furniture such as desks, chairs, cupboards, computers while maintaining radar

line-of-sight to the targets. The radar itself was located at a height of approximately 1.2 m pointing at the torso of the subjects.

The makeup of the participants had six different volunteers who took part to the data collection. 3 were males and 3 were females, with their age between 20 and 30 years.

For this collection there were seven different actions recorded:

- walking towards and away from radar,
- sitting on and standing up from a chair,
- bending to pick up an object from the floor and standing up,
- circling one arm while standing,
- clapping while standing,
- pushing: extending an arm towards the radar fast, and then slowly retracting,
- pulling: retracting arm quickly then slowly extending.

In this experiment the actions were performed repeatedly over 60 s long recordings for each activity. They were repeated once for each subject meaning a large number of repetitions of the particular movement was recorded under test.

Furthermore, additional data were collected with two of the six subjects facing different aspect angles, namely 30°, 45° and 60° away from the line-of-sight of the radar. This was done to test the effect of the aspect angle parameter on the signatures and on the classification algorithms.

Dataset 2a/2b

The recording of the second dataset, Dataset 2a was performed in the laboratory of the Telecommunication Systems Group at the Università Politecnica delle Marche, Ancona, Italy.

The height of the sensor in this case was approximately 1-1.1 m from the floor and targets were located approximately 2 to 4 m away from the radar depending on the activity.

For this set, ten different actions were recorded:

- walking (A1),
- walking while carrying an object (A2),
- sitting down on a chair (A3),
- standing up from a chair (A4),
- bending to pick up a pen (A5),
- bending to tie shoelaces (A6),
- drinking multiple sips from a glass while standing (A7),
- take out a mobile phone from pockets and pick up a call (A8),
- simulated tripping with frontal fall (A9),
- and crouching to check underneath then standing (A10).

Three different recordings were collected per person for each of the 10 activities. In each recording only one repetition of the particular movement considered was collected meaning compared to the previous dataset the actions had a finite start and stop time. The recordings had different durations depending on the activity, from 5 s to 10 s.

The participants in this experiment were seven individuals between the ages of 23 and 40 years old.

For these data, simultaneous recordings of the activities were also collected using the RGB-D sensor Kinect, located in frontal position with respect to the subjects. The layout of these sensors are shown in 5.1.

Figure 5.1: Laboratory setup for dataset 2.

Additional data in the form of Dataset 2b was then recorded for the same activities in a residential environment with two subjects: one 62 years old male whose signature is presented with younger participants 5.2 and another 58 years old female. The purpose here was to investigate differences between signatures of younger and older subjects, in order to assess the robustness of classification approaches, which are usually developed on data from younger and able bodied subjects in laboratory environments. The distance from sensor and the height of the radar were maintained from Dataset 2a, making this a counterpart set for this experiment.

The patterns and use of dataset 1 and 2a/2b are discussed in chapter 6, but the classification results of these sets highlighted the increased difficulty of classification of activities presented in the format of dataset 2a. Therefore, the future datasets from this point used it as a basis.

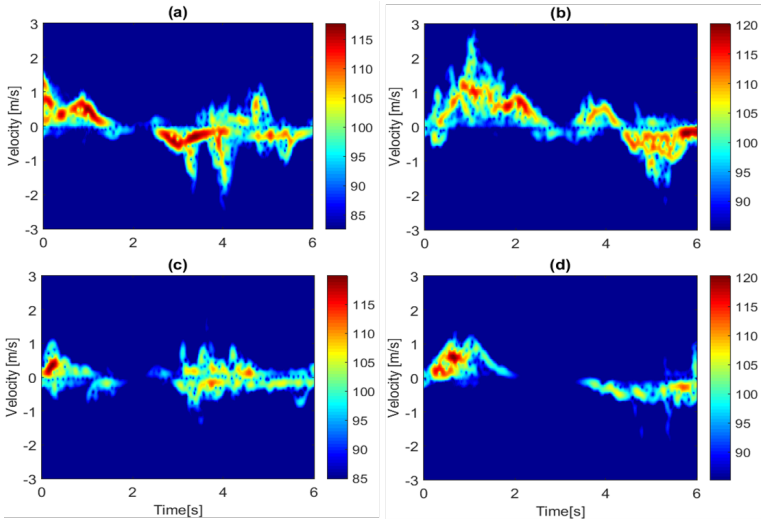

Figure 5.2: Activitiy A10: "crouching to check underneath then standing " as performed by four participants of different ages for dataset 2a.

Dataset 3

For Dataset 3, after deciding to maintain the activities in Dataset 2a (5.3.4) measurements were collected at the James Watt South building, University of Glasgow in 2017. In this experiment, in addition to the main FMCW radar, the colocated CW radar was used simultaneously. The antennas were set up at 0.6 m, in line with the torso of the participants.

Both systems were located close to each other and they were set up as shown in Fig 5.3. The distance between the antennas and the participant in this set of recordings was 1 m.

The activities are categorised and ordered into three central movements: dynamic movement, torso traversal, and limb based activity.

Dynamic movements have a wide range of motions involving translation of the torso and all limbs as seen in activities A1 and A2. While torso traversal activities have a central component of movement of torso, as seen in activities A3 to A8. Forelimb based activities mainly involve the movement of arms interaction with secondary objects such as A9 and A10. The categorisation here also helps in specifying activities similar to high risk activities such as A8: fall. These are similar to the main signature of interest, and referred to as confusers. A5: bending to pick up objects, A6: checking under bed

Figure 5.3: Sensor setup for dataset 2.

and A7: tying shoelaces, are considered as confusers for fall.

The participants in this case were twenty males between the ages of 21 and 34 years. This was also a litmus test to understand the case where variety in gender/sex was absent in the data, while physiological differences were present as variety in the form of body height and body shape.

As for the additional sensor in this experiment, a magnetometer was also recording data in parallel. As for the list of activities, they were held consistent with Dataset 2 with the main difference being the inclusion of an extra sensor.

Dataset 4

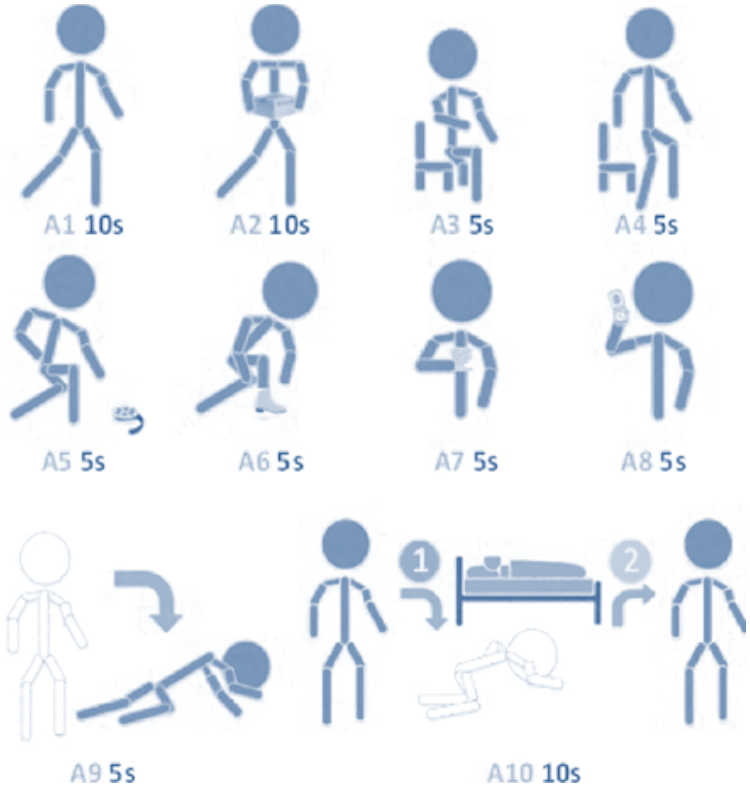

Figure 5.4: Pictograms of activities performed by the participants.

Radar sensors were collecting data simultaneously with the wearable sensor. The system was used with one transmit and one receive yagi antennas with gain of 17db, with transmit power of 100 mW. The antennae have a 48-degree beam-width in azimuth; they were 0.5m apart therefore operating in a monostatic configuration. 80 cm height means the radar's main lobe would encompass the human body with the increased backscatter coming from the torso of the participant.

Three repetitions for each of the 10 activities were taken for each volunteer with 600 samples in total taken for each sensor.

Inertial measurement unit was the co-opted sensor in this experiment. The data analysed in this work were collected with a group of 20 volunteers aged between 22 and 32 years. A pictorial representation of these activities is given in Fig 5.4.

Dataset 3 and 4 are part of a bigger database which is publicly available [31].

Figure 5.5: Environment of the experimental setup with radar antennas mounted on tripods and sources of static clutter (furniture) shown.

Dataset: Continuous

In this dataset, the challenge evolved from preventing confusers to not only correctly identifying activities within a sequence, but also transitions between activities and the whole sequence of activities.

The participants in this experiment were 14 male and 1 female aged between 21 and 35 years, collected at the University of Glasgow in July 2018.

The participants recreated daily life activities/movements in this room, where an area under observation was set for them to perform their movements. This area and the radar setup, together with the environmental setup with the clutter and furniture which are visible in Fig. 5.5. The height of the radar in this case was 0.8 m and the distance between the sensor and the area where the subjects were performing the activities was between 1 and 2 m. One male participant provided data twice, and these data from the repeated recordings have been used as the validation set for the networks for improved generalization of the classifier.

The data include six human activities:

Figure 5.6: A pictorial list of activities; these six activities were performed in a different order in three different continuous sequences

- walking (A1),
- sitting on a chair (A2),
- standing up (A3),
- bending to pick up an object (A4),
- drinking a glass of water (A5)
- simulating a frontal fall (A6).

These activities are shown in Fig. 5.6 where the individual activities shown as discrete actions, however they were performed sequentially, with each action performed in succession, with varying duration and unconstrained transitions between them. The time length of each sequence was 35 seconds, and three different sequences were recorded for each participant. They were:

- A1: A2: A3: A4: A5: A6
- A5: A4: A2: A3: A1: A6
- A4: A5: A1: A2: A3: A6

For the 15 participants, the three different sequences of continuous activities culminated in 45 different recordings for the main dataset, with an additional 3 recordings for the validation set with a repeated participant.

Table 5.1: Summary of the datasets generated from experiments used in this thesis.

Dataset	Type	No. of Participants	No. of Actions	age	Sensors used
Dataset 1	discrete	6, 3f 3m	7	20-30	FMCW
Dataset 2a	discrete	7	10	23-40	FMCW, RGBD
Dataset 2b	discrete	2, 1f 1m	10	62,58	FMCW
Dataset 3	discrete	20	10	21-34	FMCW, Magnetometer
Dataset 4	discrete	20	10	22-32	FMCW, CW, Inertial
Dataset 5	continuous	15, 14m 1f	6	21-35	FMCW, CW, Inertial

5.4 Summary

This chapter overviewed the procedure of the experimental setup and the different datasets that were generated and used in this thesis. Taking inspiration from the previous works in the literature, a set of activities which represented common actions performed daily while making the body use a variety of motions. Notably, although the initial set of activities composed of repeated simple motions, the subsequent activities were a more refined set encompassing the complex problem of continuous action sequences. The intention behind this was to reflect the progression of the experimental design towards life-like or realistic scenarios.

Aside from the physical design, there was also a careful setup of the background. Different to earlier experiments where reflection and back scatter was attempted to be mitigated using anechoic chambers or materials, the experiments were conducted in a room full of static clutter such as furniture. This made the scenario more similar to an ordinary home environment which these types of experiments in the literature attempt to replicate. Finally, this collection of data was designed from the beginning to be a growing dataset with the intention of making it publicly available. A subset of these data are part of a bigger database which is publicly available [31]. Using this data-set few exploratory works have already been generated thanks to the Radar Challenge "Human Activity Classification with Radar" which was hosted by 2020 IET International Radar Conference in Chongqing, China, with [90, 91, 92, 93] as some examples among others.

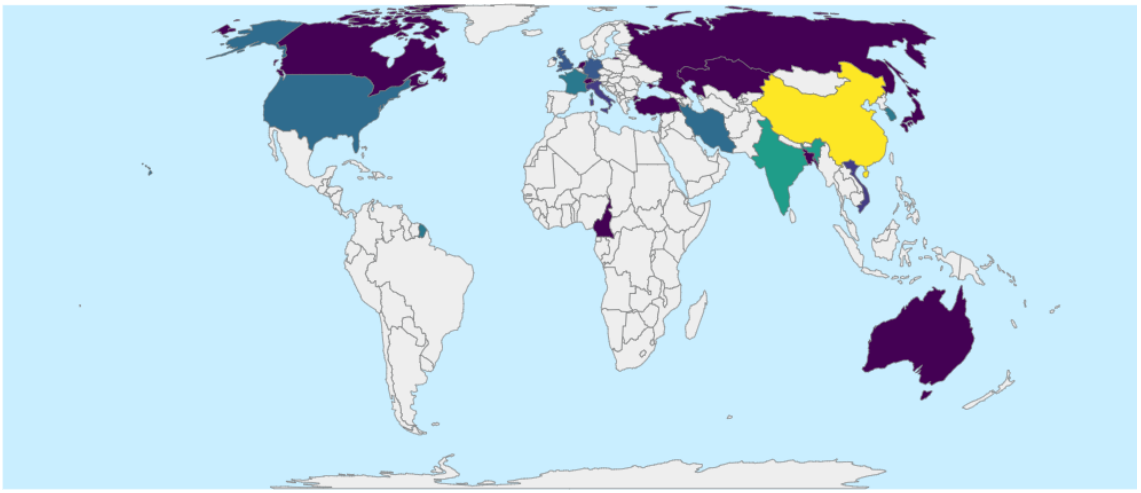

Figure 5.7: Map of location of the 147 researchers who entered the radar challenge available at : https://humanactivitiyclassificationwithradar.grand-challenge.org/

For many researchers there is a barrier that is present regarding the subject of activity recognition with radar. In the past it required bespoke hardware and time and effort to formulate experiments and find willing participants. By releasing this data-set we removed this barrier while also providing additional data to researchers in our field. Furthermore, there are other events expected where this dataset will be used for outreach and to increase the attention to this topic. One of these events will be a hackathon at the 2021 Electromagnetics Special Interest Group (EMSIG) available at :https://www.emsig.org.uk/emsigradarchallenge. Together with the hackathon and the previous challenge the goal was to make research in activity recognition with radar a global topic. Fig. 5.7 shows our current progress in this regard with a map of the countries from which researchers used our dataset for the radar challenge.

Chapter 6

Features extraction for activity and animal lameness recognition

Any cognition from intelligent beings requires identification of special properties of the object or action that is observed. In nature this is visible with animals where they rely on the transmitted signal for survival, for example using the sense of smell to track prey and measuring darkness to optimise timing for hunts. However on the primitive level, finding notable elements in the object, in other words its defining traits, is the key task behind many of these advanced activities. Therefore for computer cognition, the question about how things can be identified arises. The answer however is the same as observed in nature, through features.

Unlike biological features however, the features found in radar signals are abstract; therefore derivation and processing of the received signals is necessary to obtain them. Furthermore, referring back to nature, the properties observed by animals can be rather specific, such as the colour of the animal and the gait, and often they are used to identify if the target observed is a prey, predator or competitor. Although the exact categories may be different, this concept is fundamental in ML when feature based classification is performed.

This chapter explains the initial use of handcrafted features for activity recognition with a single sensor, specifically mentioning two applications one relating to animal welfare [30, 17] and the other regarding dependency on variety in human subjects [29, 18] for activity recognition applications. First the effect of variety in age, gender and

physiology of the subjects on the accuracy of activity detection is discussed, followed by the novel use of radar based gait recognition to detect lameness in animals which are physiologically different from humans.

First, the radar-based activity recognition is performed to assess the effectiveness of the then-current methods on a broad range of participants of different genders and age. Following this, the use-cases of applying micro Doppler (mD) based radar classification to problems such as gait monitoring in animals was explored. By porting the features used for the first application and demonstrating the ability to use mD with animals, mobility testing could be automated with results approaching to the level of human experts in detecting lameness or gait problems.

In the absence of end to end classification which are more commonplace currently, the initial efforts required extraction of interesting properties of the spectrogram or other input data from the radar as described in Chapter 3. These initial set of features are a few amongst many in the radar literature and are categorised into three groups: physical features, decomposition based features and transform based features. In this chapter their application for classification of biological movements will be discussed.

6.1 Activity classification for broad test subjects

For the first experiment, the activity recognition for a diverse set of participants is performed to address the initial topic of interest, the ambient monitoring of humans at risk of falling. Using the various feature extraction methods described then assessed their effectiveness on a variety of subjects of different ages and physical build with the dataset described in 5 as dataset 1, 2a and 2b. This experiment was also an opportunity to observe the differences between a radar spectrogram in an elderly participant compared to an able bodied participant.

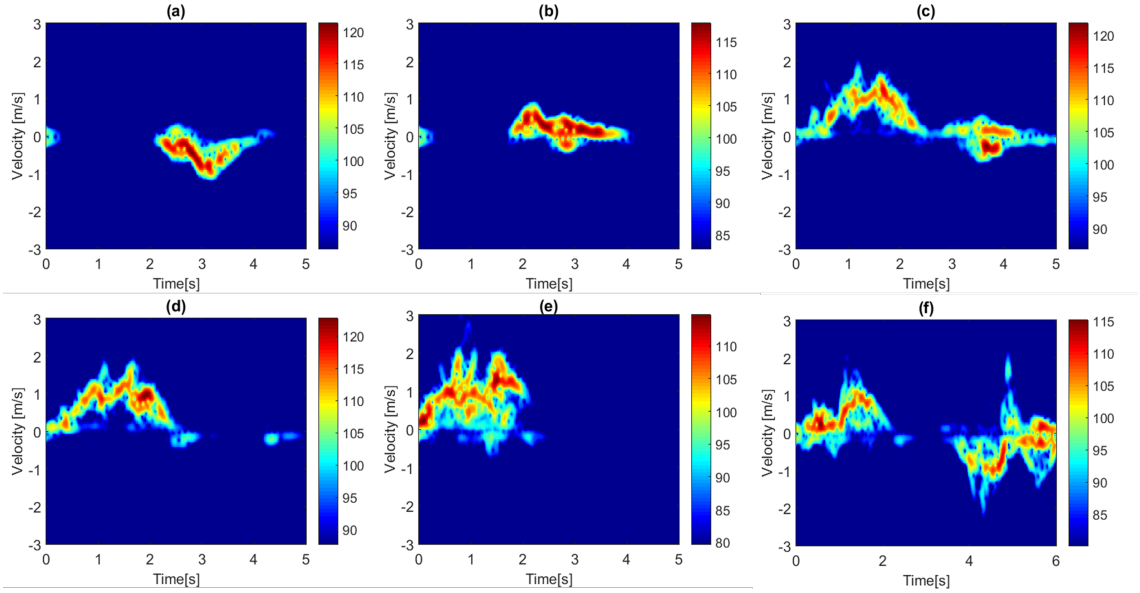

Figure 6.1: Spectrograms for 6 activities performed by the same subject: (a) sitting on a chair, (b) standing up from a chair, (c) bending and picking up a pen, (d) bending and staying low to tie shoelaces, (e) frontal fall, and (f) crouching to look below a piece of furniture and standing back up.

Figure 6.1 shows an example of spectrograms for six different actions performed by the same subject from dataset 2a. The temporal duration of the signature and the change and extension in positive/negative Doppler appear to differentiate the actions, with the challenge of finding suitable features that can capture these differences effectively and be robust to the variability from one subject to another. Some actions are more similar than others, e.g. the frontal fall in Figure 6.1e is very similar to the bending action in 6.1d, presenting a challenge for false alarms in fall detection. Furthermore, the signatures in Figure 6.1c and 6.1f could be confused with the actual fall (6.1e) as well, if the classification algorithm only considers the initial part of the signature.

Figure 5.2 in the previous chapter shows four spectrograms for the same action (crouching to look below a piece of furniture and coming back up) performed by four different subjects, one of which (Figure 5.2d) was significantly older than the other 3. One can also observe in Figure 5.2 how the same action produces rather different signatures for different subjects, and the fact that the signature for the older subject (Figure 5.2d) appears to be more limited in Doppler frequency extent than for the younger subjects.

This may highlight the importance of collecting data from actual older subjects for effective development and validation of classification techniques. This needs to be validated through the collection of a large number of signatures, including older volunteers, to validate the statistical significance.

The feature samples were then processed using different classifiers implemented in MATLAB. These were: Naïve Bayes, diagonal-linear version of the discriminant analysis, Nearest-Neighbour with 7 neighbours, binary classification tree, support vector machine with radial basis functions, and ensemble method based on random forest/bagged tree.

Table 6.1: Confusion matrix for SVM classifier (cubic kernel) for dataset 1

Accuracy [%]	Walk	Push	Pick up item	Pull	Circle arms	Clap	Sit/Stand
Walk	95.9	0.5	0.2	0.7	1.3	0.5	0.7
Push	0.2	94.7	0.3	1.8	1.2	0.3	1.1
Pick up item	0.4	0.6	94.2	1.5	1.9	0.2	0.9
Pull	0.3	4.2	0.3	91.9	1.6	0.2	1.2
Circle arms	0.3	1.7	1.6	2	91.4	1.2	1.3
Clap	0.3	0.6	0.1	0.1	1.7	96.5	0.3
Sit/Stand	0.3	1.5	0.3	1.6	1.6	0.3	94

When analysing dataset 1, the 60 seconds long spectrograms were partitioned in 3 seconds long segments and one feature per data sample was extracted from each individual segment. For this dataset, ten features were considered as input to the classifier: mean and standard deviation of the centroid and bandwidth of the signature, entropy and histogram skewness, and the six other features based on SVD. This generated a total of 2460 feature samples and from the samples set, the data was randomly partitioned in two equal subsets for training and samples; this process was then repeated 50 times to test the validity of the classification approach. After this, the average accuracy was calculated and the results per class were reported as shown in the confusion matrix in Table 6.1. The classifier used to generate this confusion matrix is an SVM with cubic

Table 6.2: Confusion matrix for SVM classifier (RBF kernel) for dataset 2

[%]	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	68	26.4	0	0	5.6	0	0	0	0	0
A2	29.2	70.4	0	0	0.4	0	0	0	0	0
A3	0	0	86	0	2.4	0	4.8	0	0	6.8
A4	0	0	0	80.8	4	2.4	9.6	3.2	0	0
A5	1.6	0	5.6	8.8	56	22.4	4	0	0.4	1.2
A6	5.6	0	0	0	19.6	66.4	3.2	0	0	5.2
A7	0	0	0	8.8	8.8	4.4	71.2	6.8	0	0
A8	0.8	0	0	4	3.6	0	9.6	82	0	0
A9	0	0	0	0	0	0	0	0	91.2	8.8
A10	0	0	0.4	2	0.4	2.4	0	0	2.4	92.4

kernel, implemented in MATLAB with one-vs-one approach for multiclass problems.

The average accuracy across the seven activities was approximately 94%. The classification accuracy is higher than 90% per activity, with misclassification events spread fairly consistently across the other activities, meaning there are no visible pairs of activities misclassified one with each other. This suggests that periodic motions can be classified with relative ease. For this dataset the recordings were collected at different aspect angles (0, 30, 45, and 60 degrees) and these have been used jointly for both training and testing. These results used the ten available features jointly, and this inspired future works to assess methods to explore the diversity in performance obtained with different combination of optimal features.

For dataset 2a, one feature sample was extracted from each spectrogram, generating 270 samples for each feature in total, as there were 10 activities repeated in 3 recordings from 9 volunteers. For this dataset 6 features were used as inputs to the classifier based on centroid and bandwidth and the textural features described earlier in this chapter.

To test the validity and the robustness of this approach 80% of the data were used to train the classifier and 20% for testing, repeating this process 50 times with different randomly selected samples for the training and testing process. The final accuracy shown is the average across the 50 iterations in Table 6.3 which presents the summary of the classification accuracy obtained with the different classifiers.

Table 6.2 instead shows an example of confusion matrix obtained for the SVM classi-

fier.The best classification accuracy of around 76-77% is obtained with the bagged tree and SVM classifiers, whereas simpler classifiers such as LD or KNN yield reduced accuracy. It is notable to assess where misclassification events happened in the confusion matrix, especially because the activities in dataset 2a were chosen to be confusers for other, and to test the effectiveness of false alarms detection. For example, A1 and A2 (walking and walking carrying an object) are confused one with each other often, and the fall (A8) is mostly confused with bending and sitting activities (A4,A5 and A7). This highlights the importance of developing feature extraction techniques capable of rejecting these false alarms and characterising the differences between confusers. The activities A9 and A10 performed on the spot (drinking, mobile phone call) present the highest classification result, and they are confused one with another often.

Table 6.3: Classification accuracy for dataset 2a with different classifiers

Classification accuracy	[%]
Naive Bayes	67.88
Linear Discriminant	58.28
K-Nearest Neighbors	60.4
Binary Tree	66
Bagged Tree	77.8
Support Vector Machine	76.44

6.1.1 Gait abnormality identification for animals

The second application of this initial feature oriented activity recognition was with animals. Extensive literature already exists on the use of radar signatures to analyze human gait and activities in the assisted living context and security/surveillance [94], [63]. However, there is very limited work on radar for lameness detection of animals, to the best of our knowledge, with the exception of a few papers where the signature of quadrupeds is treated as a potential "confuser" for human detection [95], [96]. This section expands the preliminary results in [30] by providing an initial validation of the use of radar sensing to detect lameness in dairy cows, sheep, and horses. Ex-

perimental data was collected at the facilities of the Veterinary School, University of Glasgow, Glasgow, UK, and analysed with techniques inspired from radar automatic target recognition. Promising results were achieved using the aforementioned radar features: mean and standard deviation of the center of mass and bandwidth of the micro-Doppler signatures, and classifiers such as SVM and KNN, again being similar to human application. The choice of features appears to have greater impact than that of the classifier, as similar accuracy is achieved with the same combinations of features for SVM and KNN. The rest of this chapter describes in more detail the experimental setup, data collection, and results.

6.1.2 Data composition and mD signature of farmed animals

The data analyzed here were collected using a commercial FMCW radar operating at 5.8 GHz [63]. The radar signal had 400-MHz bandwidth and 1-ms duration, providing an unambiguous Doppler range of ±500 Hz. This is equivalent to a maximum recordable velocity of approximately 12.9 m/s (46 km/h) which is sufficient to capture the movements of the animals. The transmitted power was approximately +19 dBm, and Yagi antennas with 17 dBi gain were used for the transmitter and receiver, in a monostatic radar configuration. The range resolution of the radar was 37.5 cm (related to 400-MHz bandwidth), and the 3-dB beamwidth of the antennas was approximately 24° in azimuth and elevation. The data were collected in three different environments at the Cochno Farm and Weipers Equine Hospital, University of Glasgow, one for each species of animals.

There were small differences between the data collection campaigns for the different animals. Dairy cows were walked individually along a narrow corridor adjacent to the milking parlor with radar recordings collected for both the anterior view and the posterior view of each cow. For sheep, individual animals were gathered in a small fenced zone near the radar, and allowed to walk away along the narrow fenced corridor to rejoin the rest of their flock while recording with the radar.

Figure 6.2: View of the experimental setup at the test corridor at the University Weipers Equine Hospital

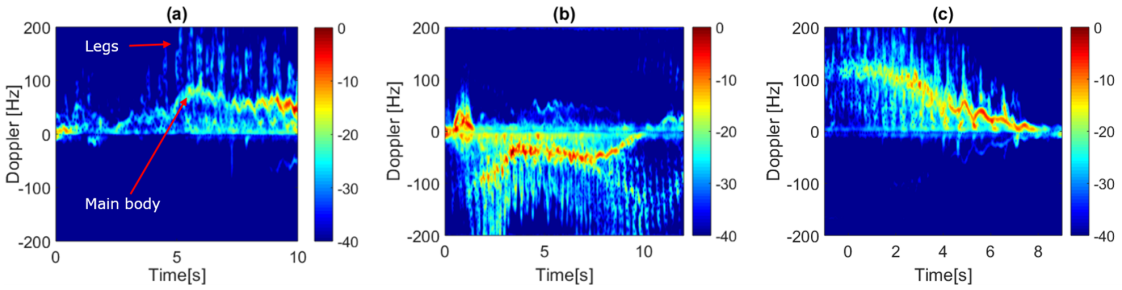

Figure 6.3: Examples of micro-Doppler signatures of farmed animals: (a) healthy dairy cow, (b) lame sheep, and (c) mildly lame horse

Horses were led by a groom back and forth along the corridor shown in Fig. 6.2 while the radar was recording, at both walking and trotting pace, and for both anterior and posterior views on the horses being tested. During each recording, the animals had their lameness assessed by veterinary clinicians to provide ground truth for comparison

with the radar data. For sheep and cows a binary scenario of detecting lame versus non-lame animals is considered here, considering mild lameness cases as part of the "non-lame class", and medium and severe lameness cases as "lame class." This is done to match the empirical assessment provided by veterinary clinicians during the data collection. For horses, a more elaborate scenario with three classes (lame, non-lame, and mildly lame) is considered, to include the borderline cases when some signs of lameness were present, but difficult to confirm definitively due to uncertainty from the experts. Micro-Doppler signatures were generated by short-time Fourier transform with 0.3s Hamming window and 95% overlap. The data were pre-filtered to remove static clutter at 0 Hz. Three examples of micro-Doppler signatures are shown in Fig. 6.3 for the animals: a healthy dairy cow walking toward the radar in Fig. 6.3(a), a lame sheep walking away from the radar in Fig. 6.3(b), and a mildly lame horse trotting in Fig. 6.3(c). The signatures appear similar to those recorded for humans [94][39]. However, the main contribution from the body of the animal occupies the middle of the signature - for example, at about 75–80 Hz at 5 s in Fig. 6.3(a) - and periodic contributions from the legs at higher Doppler/velocity around the main component, which is clearly visible above 100 Hz and up to about 200 Hz in Fig. 6.3(a). The signatures for sheep present more frequent limb movements compared with the cow signature, and this is as expected because they were moving much faster during the data collection.

Fig. 6.4 presents the signatures for two dairy cows, one which was healthy and another which was severely lame, showing both anterior view when the animals were walking towards the radar, and posterior view when the animals were walking away from the radar. Although an empirical visual comparison is not straightforward to make, the main Doppler contribution tends to have lower values for the lame animal, corresponding to reduced walking pace at approximately 30–40 Hz compared with the 70–80 Hz of the healthy cow. Furthermore, the pattern of Doppler contribution from the legs are less intense for the lame animal. Each micro-Doppler signature was divided into segments of 1.5–2 s for the time intervals when the animals were present in the radar beam, and numerical features were extracted from each segment. Four simple features previously used for human micro-Doppler analysis [63] were considered for these experiments. These are the mean and standard deviation of the Doppler centroid and

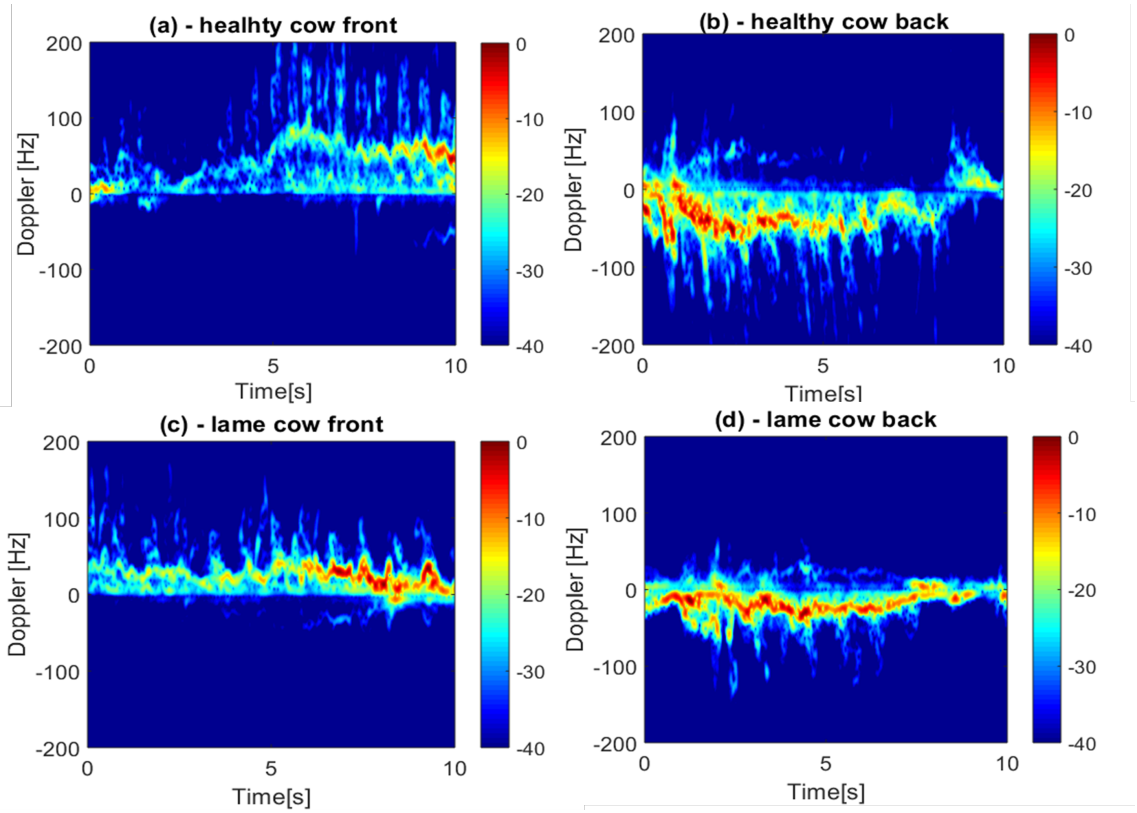

Figure 6.4: Examples of Micro-Doppler signatures of dairy cows: (a) healthy cow walking towards the radar, (b) healthy cow walking away from the radar, (c) lame cow walking towards the radar, and (d) lame cow walking away from the radar

bandwidth, as described earlier in this chapter. In practical terms, the centroid can be related to average Doppler component in the signature, meaning the bulk velocity of the animal body, and the bandwidth can be considered as a measure of the signature spread due to the patterns of leg motions.

For classification, results for SVM and KNN with three neighbors are presented in this chapter. Classification results were cross-validated using 80% of the samples for training and 20% for testing, repeating the process 20 times with randomly selected subsets for training and testing and providing the average accuracy across all the tests. This process was repeated for each combination of features considered.

6.1.3 Results for animals

Dairy Cows

Five dairy cows were considered for this test, with radar recordings taken looking at the posterior and anterior views. The cows were scored for lameness on a 0: no lameness, up to 3: severe lameness, by two veterinary clinicians while walking, independently for the anterior and posterior tests [97]. This scoring system is a standard approach used in veterinary practice in the UK, although other scoring methods there exist worldwide [98]. Three of the five cows were classified as lame with scores of 3, 2.5, and 2 present on average, and two as healthy who scored 0.5 and 1 using a binary classification. These scores were grouped from the 0–3 scoring system into a binary non-lame (score 0–1) versus lame (2–3) classification was done to match the overall empirical assessment of the veterinary clinician who assisted with the data collection, which is also a possible approach in veterinary literature [99]. A total of 53 samples were obtained, with 18 samples for the "non-lame class" and 35 for the "lame class." This class imbalance depends on the time each cow was visible to the radar sensor, as their speed cannot be controlled while recording data.

Fig. 6.5 shows the classification accuracy as a function of feature combinations for dairy cows in anterior and posterior view. These feature combinations include: 1)#1–4: Individual features, where 1 and 2 are the mean and standard deviation of the centroid, respectively, and 3 and 4 are the mean and standard deviation of the bandwidth. 2)

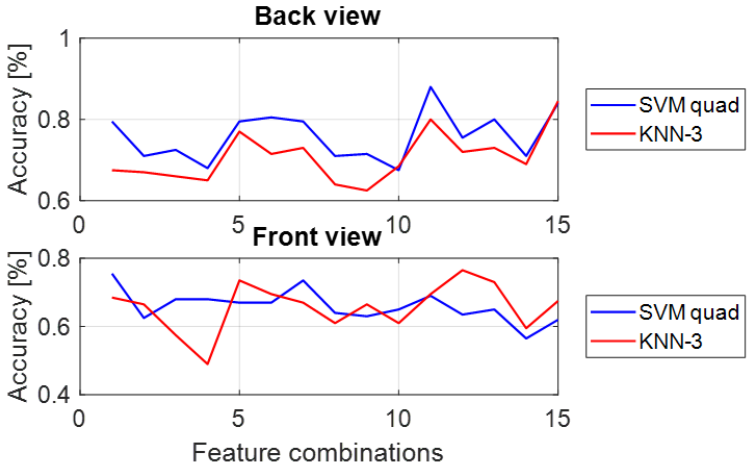

Figure 6.5: Classification accuracy as a function of feature combinations for dairy cows, front and back view

Table 6.4: Confusion matrix for best prediction for cows

Accuracy [%]	Predicted Healthy	Predicted Lame
True Healthy	70%	30%
True Lame	8.60%	91.40%

#5–10: Six possible pairs of features combining the aforementioned four individual features. 3) #11–14: Triplets of features (combinations of three features). 4) #15: All four features used jointly.

SVM-Quadratic (SVM-Q) and KNN appear to provide similar levels of accuracy as a function of features, for the posterior view at least. This suggests that the choice of features has a great impact on the final performance, far more than the type of classifier for the classification problem considered here. Accuracy of 80% is achieved with the two features, mean of centroid and bandwidth, with a maximum of 85% adding the standard deviation of the centroid as a feature. Accuracy for the anterior view is lower than that for the posterior view, and this is expected as lameness in dairy cows tend to be more significant in the hind limbs, which are less visible from an anterior perspective. An example of confusion matrix for the highest accuracy case in posterior perspective for SVM-Q classifier with three features as input is shown in Table 6.4. There are more "false positives" meaning healthy cows which were predicted as lame than "false negatives" where lame cows were predicted as healthy. This trend also

appears for the anterior perspective. Evaluating what is more penalizing in practical and economic terms between false positives and false negatives depends on the context and implications of the lameness diagnosis for the affected cows and is a future aspect of interest. For instance, predicting a healthy cow as lame can have a significant logistic/economic impact, if the animal is taken away from the herd and removed from the dairy production cycle for treatment. Furthermore, if it is assumed that a veterinary clinician will check every case flagged as lame by the automatic detection system, then false positives will increase time and cost of the procedure. As for false negatives, they are also undesirable for the long-term well-being of the animals and the overall effectiveness of the proposed system, especially if lameness cases are systematically missed until the health of the animals affected are seriously compromised.

Sheep

Measuring sheep is more challenging, as they generally dislike being separated from other sheep and tend to stick together with the herd, nose to tail with other sheep. For this test, six sheep were used, three of them presenting healthy/normal gait and three of them presenting some form of lameness. The sheep were normally marked on a binary 0/1, non-lame/lame scale by veterinary clinicians which were categorised into the different classes.

A total of 42 samples were considered for classification, 19 samples for the "non-lame class" and 23 for the "lame class" and the radar was deployed to look at the hind limbs of the sheep. Fig. 6.6 presents accuracy as a function of feature combinations, for both SVM-Q and KNN classifiers. Very high classification accuracy approaching 100% was achieved with just a single feature, mean of centroid, using both SVM-Q and KNN with a pair of features. Essentially, in this case, the classification appears to be strongly related to the mean velocity of the sheep meaning the mean of the main body velocity from the micro-Doppler signature. For sheep, the average velocity can be considered in the first instance a good proxy for lameness in sheep as lame animals cannot move as fast as healthy ones although they all try to run to rejoin their flock.

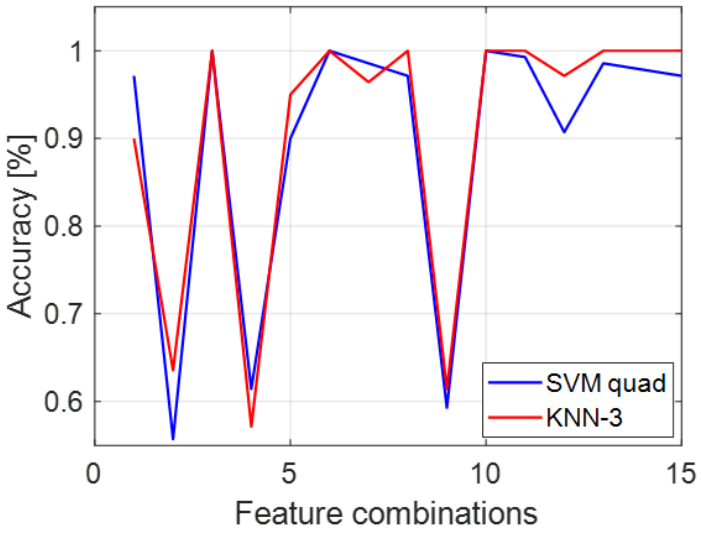

Figure 6.6: Classification accuracy as a function of feature combinations for sheep in posterior view.

Table 6.5: Confusion matrix for horses

Accuracy[%]	Predicted Healthy	Predicted Lame	Predicted Min. Lame
True Healthy	94.2	5.8	0
True Lame	11.6	87.7	0.7
True Min. Lame	1.4	3.4	95.2

Horses

For this test, four horses were recorded for both walking and trotting gait patterns and of the four horses, one had no visible lameness while another had severe lameness in the anterior, and two others had mild lameness in the proximal front limb. These horses were labeled into three groups: "healthy," "lame," and "minimal lame", and overall they generated 162 samples with 54 samples for the "healthy," 36 samples for "lame," and 72 samples for "minimal lame" classes with both walking and trotting recordings considered together. The classification results from this three-class set are shown in Fig. 6.7 and Table 6.5. The trend here is similar to the cows and the sheep meaning there is no distinct difference between the classification algorithms' performance. The only change from the previous instances was a difference with the SVM, where a cubic kernel was used, which provided improved results compared with the quadratic kernel of approximately 12% higher on average. This may be due the

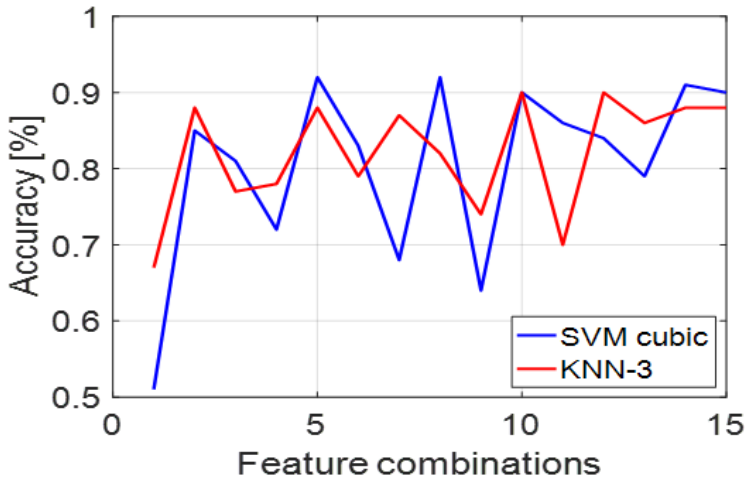

Figure 6.7: Classification accuracy as a function of feature combinations for horses

more complex distribution of features, the problem being a three-class classification. However the suitability of centroid-based features remains significant even in this case as shown by peaks of accuracy in Fig. 6.7 when centroid-based features are added to the feature set during testing all the possible combinations, similarly to what was recorded for sheep in Fig. 6.6. The most suitable combination gives classification accuracy of up to approximately 92% and although the classifier correctly identifies the "healthy" and "minimal lame" to very high rates in Table 6.5, there is a relatively high false negative for the "lame" class, as approximately one out of ten "lame" horses are incorrectly classified as "healthy."

6.2 Conclusions

In this chapter, the results from using feature extraction with human activity detection are presented. Following this 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 achieveable 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 a daily activity or a harmful activity such as a fall. By this, activities in this case could be detected at rates between 68% to 92.4%, but individual discrete activities were more challenging to classify compared to segmented of repeated discrete activities which could be identified above 91.9%.

The other part of this chapter focused on the porting of techniques and classifiers to identify gait abnormalities in animals. This was an innovation on the classic gait analysis with radar as now the domestic animals were the focus of the classification. The fact that the feature extraction and classification algorithm worked well suggested an element of ubiquity in this sensor and method, for this application even when the species of animals under test is different. This showed that lameness in dairy cows can be identified to 80%, up to 100% in sheep, and up to 92% in horses for a three-class problem of classifying severe, mild, and absence of lameness. It is interesting to note that these promising proof of concept results presented here led to a larger experiment with many more animals, specifically dairy cows and sheep, with results reported in [100].

Ultimately, by introducing this new domain of problems to be solved with radar and demonstrating the feasibility using feature based classification, this broadened the scope of what is possible with radar concerning activity classification.

Chapter 7

Sensor fusion for discrete activity recognition

This chapter introduces the second novelty of this thesis which is the use of homogeneous sensor fusion and heterogeneous sensor fusion are 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. Taking the example of nature again, the information received by a single sense can be confirmed by another. For example, noticing a fruit falling from a tree through audiovisual cues, or determining the type of fruit through visual and tactile confirmation, these all rely on a multitude of sensors. In this chapter, this concept is used to show the strength of leveraging different information from varying sources through sensor fusion for activity recognition in the form of heterogeneous sensor fusion, where one sensor is a radar sensor and the other is an inertial measurement unit [19, 20, 32], and homogeneous sensor fusion, where both sensors are ambient radar sensors [24]. Both of these techniques are explored in this chapter with an additional analysis on 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 sensing methods to optimise the classification accuracy in the context of assisted living. This was addressed in a twofold manner:

Firstly by showing the usefulness of features from different domains with feature selection to improve activity recognition accuracy. Previously, in the literature there were a number of features using different properties of the input signal being used for radar based activity monitoring. Therefore, the next goal was to unify these different methods to generate a varied feature set for automated activity recognition. By using them together with feature selection we were able to demonstrate higher accuracy in activity recognition.

Secondly, cooperative use of varying sensing technologies was found to further increase accuracy when compared to single sensor results. Although this was a lateral research direction to the aforementioned feature selection work, it was a result of the question of using different sensors together arising as the benefit of increased variance through multi-domain inputs being observed.

These two research opportunities were separated and solved through heterogeneous sensor fusion, where sensors were measuring different physical quantities and signatures, and homogeneous sensor fusion, where radar sensors were measuring the same physical signature but at different frequencies. With both these techniques, improvements to the accuracy of activity recognition were demonstrated.

7.1 Review of multi-sensor fusion

As individual sensors has certain limitations to the scope of information it can record [101]. For example, the aspect angle problem is an open challenge for healthcare applications as the returned signature may be attenuated if the sensor is not in the direct line of sight [102]. Usually there is a decrease in the returned signature proportional to the cosine angle relative to the direct line of sight, this ultimately means the returned mD signature is weaker . This means the movements derived from it will be inaccurate, therefore feature space derived from the radar in this case would therefore be erroneous. This is a scenario where a different type of sensor can be introduced to mitigate the error from the radar spectrogram and to ensure maximum variety from leveraging a different domain.

7.1.1 Sensor fusion typologies

The term information fusion is applied to combinations of any types of data or original sources. As different sensors have strengths in different domains, the main reason for sensor fusion is to effectively utilise variety and having commensurate information.

Depending on the types of information sources or sensors available, existing types fusion can be applied as sensor fusion networks. These configurations can be be either co-operative, competitive or complementary [101].

Cooperative sensor networks use data from two independent sensors receiving the same type of information from different perspectives. This type of fusion network uses spatial variety to use commensurate data from different points in space to generate new types of data [101]. An example of this would be stereoscopic depth perception through numerous cameras observing the same scene with different focal lengths.

Competitive fusion networks leverage the same data with a spatial and temporal variation to make the system more robust [101]. In this scenario, the information fused is from the same type of sensors, but the location or timing changes.

Complementary fusion networks use independent sensors to receive data with maximum variety from different domains to get a bigger overall picture [101]. An example of this would be using an electrocardiogram and electroencephalogram to detect vital signs as they are utilizing information from the same observation but two different domains, electric pulses in the chest and voltages in the brain, to come to a decision where the person is identified as alive or not.

Assuming the sensors follow the requirements of these networks and have a common goal, either to measure the different movements or to classify a certain activity, fusion can be performed at different levels with varying amounts of complexity through three methods [101][27][19]: signal level fusion, feature level fusion and decision level fusion. These are illustrated in fig. 7.1, where the different points in the processing stage fusion can be applied to are highlighted. The sensor is the initial step in the chain where the data is pre-processed into digital data. Following this, other processes including feature extraction and fusion are performed.

The point at which fusion occurs defines the level or order of the fusion where lower

level fusion methods utilize the correspondence between the data, whereas higher-level methods use confidence metrics for decisions or labels from the classifier.

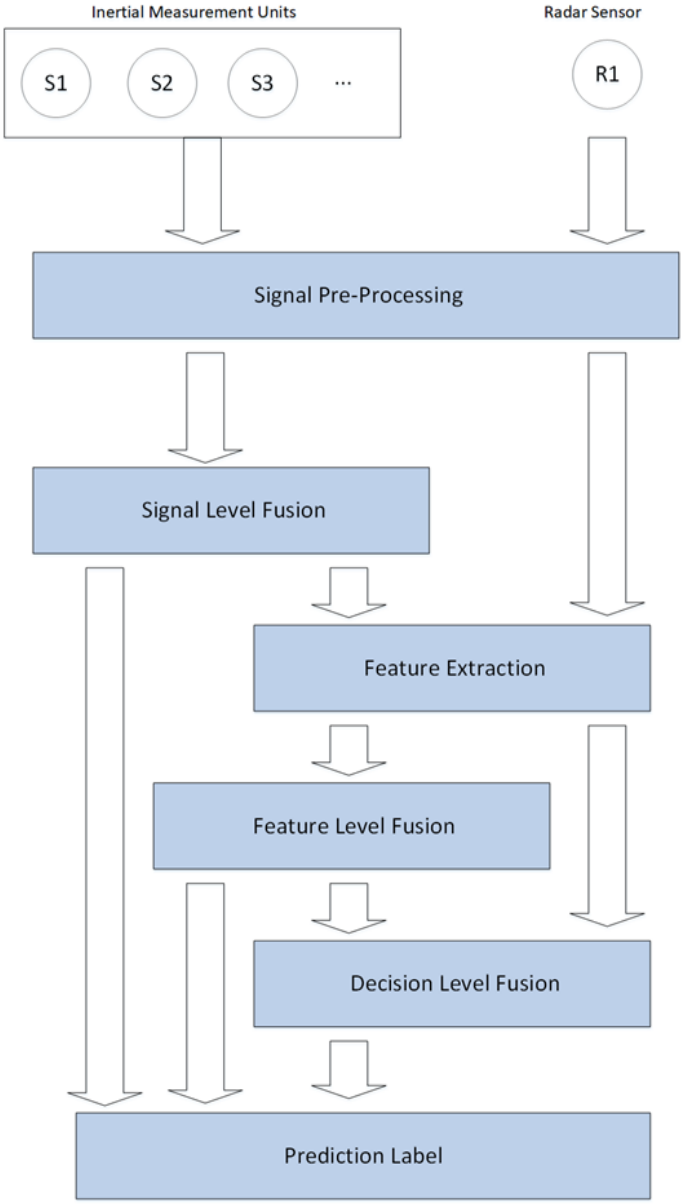

Figure 7.1: Stages and types of sensor fusion algorithms

Signal level fusion is the deepest method of fusion and in the context of human activity recognition [101] it takes place between sensors recording the same quantities from different locations on the body of the monitored subject. It is a complementary fusion method where commensurate data from the sensors is directly combined through mixing or Kalman filtering. The benefit of signal level fusion comes from fusion of data

prior to processing individually. This means instead of processing many sensor output chains, they can be combined and processed in an efficient manner. However, signal level fusion also has limitations as data from different domains are usually sampled at different rates and they can have opposite information causing the fused data to become redundant. To mitigate this, down sampling, where segments of data are removed, or up converting, where data points are resampled, need to be used with this method in most cases. The caveats with these methods are that the removing and resampling can either remove information or make it redundant.

Feature level fusion is a simpler method that combines the features generated after post processing the signal to generate dimensionally reduced, representative information [19]. As features are generated in a uniform observation or time step, the features from the different sensors can be concatenated into a single, larger feature matrix. This has the drawback of increasing the size and dimensionality of the feature space but also it also brings diverse information together. The main benefit of feature fusion is that since the features extracted are originally useful, the new feature subspace will have useful relevant data and minimal noise. In the context of images, feature level fusion can be used for different domains of analysis such as multi-view, multi-temporal and multi-modal analysis.

Multiview analysis, as the name suggests, uses different viewpoints requiring information from complementary, cooperative and competitive sensor networks, which provide a spatially larger observation of the target scene or activity [101]. Multi-temporal analysis acquires the data at different time steps from complementary networks to detect changes in the time signature. Multimodal analysis integrates information from the same event acquired by different sensors to construct a more complex and dynamic representation of the observation and this is the focus of the experiments in this chapter.

Decision level fusion is a higher level information fusion method which considers the confidence or output from the classifiers [27]. In a classification problem, the output of the process chain is the final label provided by the classifier and a measure that expresses the level of confidence in this decision, given as a form of 'loss' as explained in 3. If the fusion is performed on the confidence metric, then it is called a soft fusion.

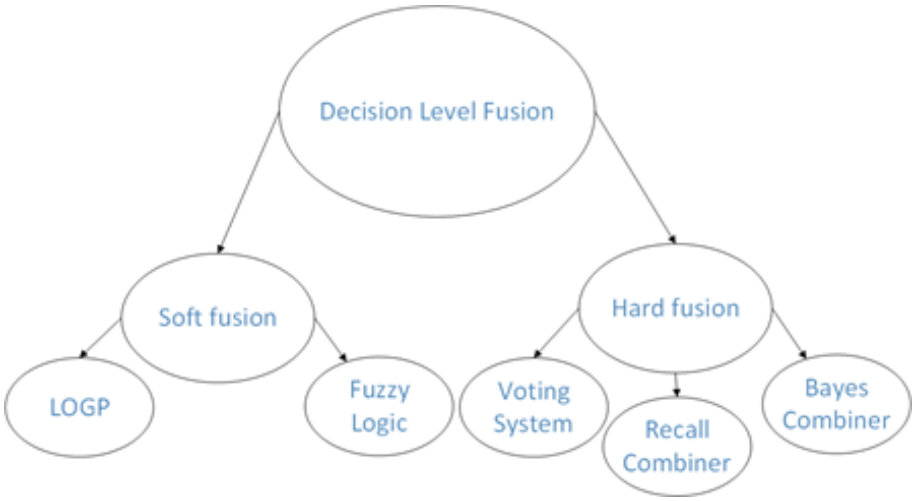

Figure 7.2: Breakdown of decision level fusion methods

In soft fusion, each sensor data, raw or pre-processed, is fed into individual classifiers. Due to the separation of the data the classification complexity is reduced, as there is no need for tuning the parameter again for a new feature set. For all the classes of the set, the classifier detects the likelihood of the detected class by selecting the label with the lowest 'loss' parameter. When there is insufficient or invariant data, the loss parameter for two classes might be close causing an error in the motion or activity detected. In this instance, the 'loss' parameter can be extracted from the two different classifiers, scaled logarithmically, and then added together to correct the classification error. The loss can be mixed in two ways, pooling (soft fusion) and voting (hard fusion).

Logarithmic opinion pool (LOGP) scales the log loss exponentially to a measure of probability in B_n which is a mass Gaussian function [103]. Using a distribution factor d which is $1/N$, the number of classifiers or sensor sources, $B(\alpha|y)$ gives the most likely class and in most cases the correct class.

$$B_n = e^{-S_n(c)} \tag{7.1}$$

$$B(\alpha|y) = \prod Nn = 1B_n(a|y)^d = \prod Nn = 1e^{-S_n(c)^d} \tag{7.2}$$

Fuzzy logic uses the confidence metric as a fuzzy set as described in equation 7.3. S_{Radar}

and $S_{Inertial}$ are two sets of negated binary loss values which include 10 elements related to the 10 activities, as generated from the two SVM classifiers. S_{Fused} is the new binary loss set which represents the sum of errors. The final decision is made by finding the least errors in this set.

$$S_{fused}(c) = min(S_{Radar}(c), S_{Intertial}(c)) \tag{7.3}$$

Rather than merging the confidence level of different sensors in the second stage of hierarchical model, the hard fusion of radar and inertial sensor occurs between the prediction labels through a probability combiner. There are several potential combiners in the literature, such as the majority voting system or weighted voting system, Recall combiner and Naïve Bayes combiner. However, a simple voting-based system as the first three are not suitable due to the decision clashes in most complex scenarios whereas Recall combiner is not ideal for binary classes problems since the performance of Recall combiner is proportional to the number of classifiers.

Naïve Bayes combiner calculates the posterior probability of each class through the prediction label and confusion matrix of individual sensor. The probability of a certain class after fusion is obtained by the equation 7.4 below:

$$logP(C_k|d) \propto logP(C_k) + \sum_{m=1}^{N} log(p_m, c_m, k) \tag{7.4}$$

Where $logP(C_k|d)$ is the probability of interest, denoted for the likelihood of class C_k being the true class. $P(C_k)$represents the number of classifiers suggesting C_k as the prediction label. The classifier belonging to a classifier ensemble whose length is equal to N. p_m, c_m, k refers to the confusion matrix element corresponding to classifier p_m, row c_m and column k. The final prediction label is the class with highest posterior probability after this process. Compared with soft fusion, hard fusion requires less computational load being more efficient with selecting the optimal weight function.

7.2 Review of feature selection

A high degree of accuracy requires features which are discriminative and provide confidence and reliability in the classification results [104, 105]. The main requirement for a well performing classifier in most cases is the input to the model. The size of the input is often a concern, therefore starting with a low sized input set can be helpful especially as the number of sensory information increases the feature space can become orders of magnitude larger, spatially and dimensionally. This makes the feature space redundant due to 'the curse of dimensionality'. To avoid this and to maintain the input size, performing feature selection has been advocated in literature and there is evidence it can reduce the number of features considerably. For modern complex classifiers however, there are intrinsic methods to reduce redundancy.

There are two classes of methods utilised for selecting optimal features: filter methods, based on feature space metrics such as euclidean distance, entropy, correlation coefficients. These metrics are used to find the distance between feature clusters and rank them accordingly. On the contrary, wrapper methods test different combinations in the feature space and for each of them a classifier, in order to find the best performing feature set. Wrapper methods require more iterations and exhaustive search within the feature space, to find the optimal input features for the classifier, therefore they are more resource intensive than the filter methods.

F-score (7.5) [105] finds a feature subset by calculating and ranking the score determined by the distance between data points. The distance between data points belonging to different groups is as large as possible while the distance between data points belonging to the same group or class is as small as possible. It is defined as:

$$F(X^i) = \frac{\sum_{j=1}^{c} n_j(\mu_i^j - \mu^i)^2}{\sum_{j=1}^{c} n_j(\sigma_j^i)^2} \tag{7.5}$$

Where $F(X^i)$ is the fisher score of i^th feature; the parameters n_j and μ_j are the size of the jth class and the mean value respectively; σ^i and μ^i indicates the standard deviation and mean value of the feature subset regarding to the ith feature.

Relief-F is a modification of the original algorithm, the Relief algorithm, to fit multi-

class classification problems [105]. An observation corresponding to a row of feature data set is selected randomly, where the nearest hit belonging to the target class and the nearest miss belonging to the opposite class are generated. The input of the algorithm should be normalized to [0, 1] by dividing the weight with the product of number of features and classes, and the output is a weight between -1 and 1 given for each feature. The weight of features is updated iteratively by the type of nearest miss, weighted by prior probabilities of each class. The Relief-F algorithm is explained in the pseudo-code below.

Algorithm 1: Relief-F Algorithm

Result: Obtain feature set with minimal distance

Set all the weight W(F) to 0;

for *i = 1 : n (n equates to the number of instances)* **do**

Select one observation B_i randomly ;

Find k nearest hit h for each Class C = B_i;

Find k nearest miss m for each Class C $\neq B_i$;

for *F = 1 : number of features* **do**

$$W(F) = W(F) - \sum_{j=1}^{k} \frac{diff(F, B_i, h_j)}{n * k} + \sum C \neq$$
$$class(B_j) \frac{P(C)}{1 - P(class(B_i))} \frac{\sum_{j=1}^{k} diff(F, B_i, m_j)}{n * k}.;$$

Update the weight;

end

end

Where the *diff* function denotes the difference between the values of features for two observations and k is the class. In this case P(C) is equal to 1/10 due to equal numbers of observations being present in each class.

The best combinations of features by using a classifier can be found through sequential feature selection (SFS) [105]. Forward SFS adds features according to classifier performance iteratively increasing the input subspace of the classifier. In backward search it begins with the original feature subset and reduces the dimension one-by-one. Of the feature selection methods, SFS is the most computationally intensive method as it runs (No. of features-1)*No. of features/2 times multi-class SVM. The pseudo-code

below shows the steps of the SFS algorithm.

Algorithm 2: The SFS Algorithm:

Result: find one feature combination which yields the best classification accuracy

Initialized with an empty feature set F(θ) ;

while *While Accuracy is not stagnating or all features have not been used* **do**

Find feature X which yields the best performance with F(θ) ;

The new feature subset selected becomes F(θ)=F(θ)+X ;

Update the new feature set iteratively ;

Until accuracy saturates or stagnates ;

end

7.3 Heterogeneous data fusion

Using a singulars sensors has been attractive in the biomedical field to measure precise information such as muscle movements, heart movements and brain activity perception [106]. Pressure sensor arrays, together with barometers have been recently explored as complementary sources of information. However, the richness of the data coming from the inertial measurement unit and its components: accelerometer, magnetometer and gyroscope have been the focus of research for activity recognition.

7.3.1 Feature extraction from inertial and radar sensors

In cases where the target is a body with multiple sources of micro motion, the overall movement of the limbs is extracted by generating a spectrogram with a radar sensor. The spectrogram and the perceptible data within it need to be in a format that can be interpreted by a machine learning algorithm to classify different activities instead of a human operator [107, 108, 109].

Feature extraction is defined as deriving properties of measured data that are informative and representative of the source information. As discussed for radar in Chapter 6 they can be highly varied [110] and they are also methods of dimensionality reduction as the features contain identifiable information for a certain class but are usually orders

of magnitude smaller than the input [107]. Yet, features express the salient property of the class despite the smaller amount of information compared to the source.

Even though the spectrograms have visually perceptible movements to the human eye, to translate this for an activity recognition application, the properties that define these two distinct movements must be extracted.

Two types of feature extraction methods encompass the features currently used in literature, both discussed in Chapter 6: automatic features and handcrafted features. The prior uses established dimensionality reduction methods such principle components analysis or singular value decomposition to reduce the input into smaller dimensions with highly varied data. The latter on the other hand use a functional pattern within the input to characterize the signal.

Automatic features for radar data are extracted from decomposition of the data where the maximally variant or the most diverse features are identified and extracted. These have been discussed in more detail in Chapter 6. They include PCA , SVD and ICA. These methods do not require precise derivation of information and modifying or fine-tuning the algorithm. They are resilient to different levels of signal to noise ratio and produce salient features.

Handcrafted features for radar data are functional properties of the input, which are commonly identifiable. These features enumerate these noticeable properties which match between different samples of the same class. For micro-Doppler radar, the prominent handcrafted features used in this research are Doppler centroid , Doppler bandwidth , entropy and grey level of histograms, Spread spectrum shape and Cadence peaks as displayed in table 7.1.

Features in Table 7.2 can be used to characterize the temporal and spectral information of the raw inertial data. Temporal inertial features include mean, variance, standard deviation and other statistical components like skewness and kurtosis,together with cross-correlation between different axes. Spectral inertial features are extracted to capture the energy distribution of the signal and include the magnitude of the power spectral density at three different frequency bands, at 0.5-1 Hz, 1-5 Hz and 5-10 Hz; the sum of Fourier Transform coefficients, and the spectral entropy based on the power density function normalized between 0 and 1 of the power spectral density. The mean

Table 7.1: Features for radar sensor

Feature Category	Radar Features	#
Image based	Entropy of spectrogram	1
	Skewness of spectrogram	1
Physical	Centroid of spectrogram (mean & variance)	2
	Bandwidth of spectrogram (mean & variance)	2
	Energy curve of spectrogram	3
Transform based	Singular Value Decomposition	13
	Range Doppler velocity	1
	Range Doppler displacement	1
	Range Doppler dispersion	1
	Energy curve of spectrogram	3
	Step repetition frequency	1
	Step repetition frequency band peak	2
	Number of features	28

Table 7.2: Features for intertial sensor separated by temporal and spectral properties.

Temporal	#	Spectral	#
Norm of XYZ	1	Spectral Power	9
Mean	3	Coefficients Sum	3
Standard Deviation	3	Spectral Entropy	3
Autocorrelation(Mean,STD)	6		
Cross Correlation(Mean,STD)	6		
Variance	3		
RMS* (Root Mean Square)	3		
MAD (Median Absolute Deviation)	3		
Inter-quadrature Range	3		
Range	3		
Minimum	3		
25th percentiles	3		
75th percentiles	3		
Skewness	3		
Kurtosis	3		
Number of features	49	Number of features	15

and variance of specific signal is derived by the equation 7.6 and 7.7, where $x(i)$ is the input signal and N denotes the number of samples.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x(i) \tag{7.6}$$

$$\sigma^2 = (\frac{1}{N}\sum_{i=1}^{N}(x(i) - \mu)^2)^2 \tag{7.7}$$

The correlation function is significant in classifying activities with signal magnitude change along two dimensions, such as the activities where the body modulates back and forth. The correlation between X and Y axis, which equals to the ratio of covariance between two inputs and the product of their standard deviation is one of the correlation features utilised here.

Spectral features also approximate the energy burst of different activities and the trend visible suggests that activities with fast movement provide a high response in a short frequency band, especially for the falls.

Classifiers

All the training and testing procedures are performed in Matlab utilising the SVM library: libSVM. The dataset was stochastically divided into two parts, with 70% data for training and 30% data for testing on a per class basis, and the per class basis was set for stratification in the test set to prevent class imbalance.

Table 7.3: Comparison of classifiers for activity monitoring.

Classifier	Complexity	Classification time (s)	Accuracy
Linear discriminant	small	0.37	68-70%
K nearest neighbour	small	0.48	75-78%
Support vector machine	large	1.4	84-90%
Artificial neural network	large	2.38	86-92%
Deep neural network1	Very large	>1200	88-95%

By using this deterministic approach, we minimised unwanted bias in the results, which may occur in cases of imbalance between classes in the training and test sets. This process is repeated 10 times [65] for each test, and the average results across all the repetitions are presented [27]. A comparison of all the classifiers are presented in Table 7.3 where approximate accuracy, time taken to generate a classification result and complexity is shown.

7.4 Discrete assisted living activity detection with fusion

In the two key experiments presented in this chapter, a combination of non-contact and wearable sensors were used for monitoring ten daily activities [27, 19]. In the first experiment, a single magnetometer embedded within an inertial sensor was used in parallel with a FMCW radar. In a second experiment, a smartphone with its multiple degree of freedom was used together with FMCW radar to assess the effect of having acceleration and gyroscope information available.

Experimental and hardware setup

The main hardware components in the experiments are the Magnetometer, the IMU and the FMCW radar. While the FMCW radar produced by Ancortek inc mentioned in Chapter 2 operates with a center frequency of 5.8 GHz and has an instantaneous bandwidth of 400 MHz and a pulse repetition frequency (PRF) of 1 kHz. A distance of two meters from the target would replicate real life conditions where the target would be a short distance away from the radar. Variations in movement were present depending on the person and the activity being performed to replicate life-like movement.

In both experiments, the wearable was attached to the participant's wrist with a bracelet on the dominant hand. Considering the inertial sub-sensors, magnetic sensors are either used in conjunction with accelerometer and gyroscopes for activity recognition or ignored, instead favoring data from the other two inertial units. Therefore, the use of magnetic sensors on their own with a non-contact sensing method is intriguing.

This in turn reduces the computational complexity of the processing chain since it has less overall number of sources, which reduces battery consumption.

Radar sensors were collecting data simultaneously with the wearable sensor. Three repetitions for each of the 10 activities were taken for each volunteer with 600 samples in total taken for each sensor [19]. For the second experiment, the data analysed were collected with a group of 20 volunteers aged between 22 and 32 years. Each activity was recorded for three repetitions for each volunteer, generating a dataset of 600 simultaneous readings from each sensor. The radar sensors were placed on a wooden table at approximately 80cm height, pointing to the area where the subjects were performing the different activities at about 1-2 m. The smartphone was held with a Velcro-strap on the wrist of the dominant hand of the participating subjects to keep parity with the other experiment while recording data. The radar system and its antennas were placed on a box, facing the area with the subjects in the direct line of sight. The separation between the antennas was approximately 30 cm and the distance from them to the subjects was approximately 1.5 m; furthermore, vertical polarisation was used for these measurements. These were the activity sets 3 and 4 as previously discussed in Chapter 5.

Experimental design

To make the classification challenge harder, sets of similar activities were decided to be included in the original framework of this set. The listed activities, as mentioned in Chapter 5 were selected due to the variety in the motions required as the whole body with its constituent parts, and similarity to other activities in the set. Most of these movements, other than falls, are commonly performed in daily life and therefore they are a balanced indicators of decreasing mobility.

Fig.7.3 in shows the mD Doppler signature and magnetometer data for four different daily activities. The strongest segment of the spectrogram is used to normalize it to scale the clutter and the noise floor down. However, changing the distance between the target and the radar can cause a corresponding change in the background noise levels of the radar.

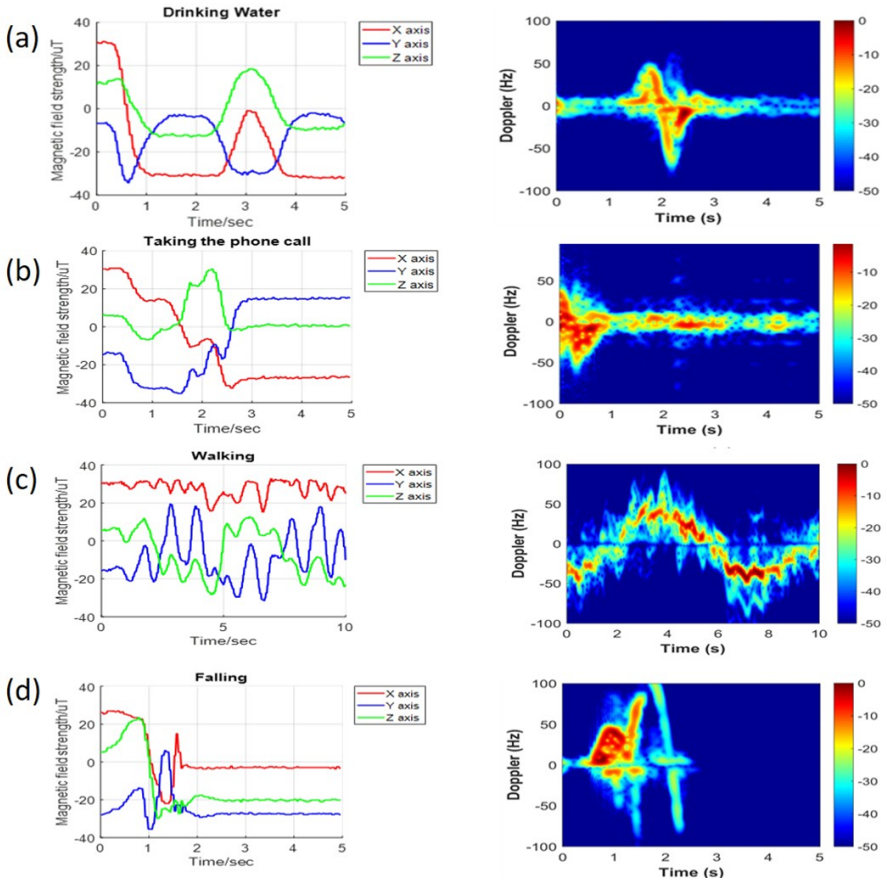

Figure 7.3: Magnetometer and mD spectrogram representations for four activities: a) reaching out to drink water from a cup b) picking up a phone to receive a call c) walking back and forth d) frontal fall.

7.4.1 Classification results

Magnetic sensor and Radar

Fig. 7.4 shows the classification accuracy for different sensors and classifiers and the varying number of optimal features through the SFS algorithm. There is no linear relation between the number of features and the accuracy, therefore the presence of redundant features is assumed. The accuracy over the number of features becomes stable when 30 features are selected for the magnetic sensor and when 10 features for the radar. The average accuracy with these features is 90% approximately. As adding more features will not bring any improvements to the classification and instead it may

reduce the performance, only a subset of all the available features should be used [19].

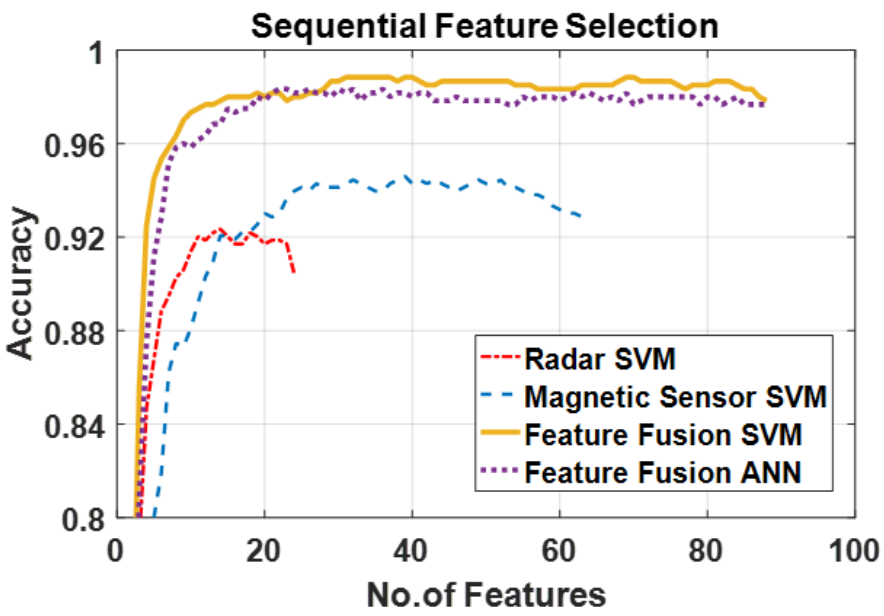

Figure 7.4: Accuracy comparison between sensors with and without fusion

Feature selection after feature fusion derives the salient features from both radar and the magnetic sensor after combining them into a single feature space. The resulting accuracy from feature fusion outperforms the cases of both sensors used individually [19].

Fig. 7.4 shows also the accuracy as a function of the number of features from a fused pool of features from both sensors used as input to a single hidden layer artificial neural network (ANN) classifier with 50 neurons. Results appear to be very similar for SVM and ANN, so classifier selection appears to be unimportant provided fusion and feature selection are used. The validation accuracy produced using ANN with a single hidden layer and between 1 and 50 neurons, utilizing different sensors is shown in Fig. 7.5. Results from using fusion outperforms using each sensor individually with an average of 96% with feature fusion, which is similar for SVM. The near maximum accuracy is reached when over 10 neurons are used in the fusion case, which corresponds to the number of classes [19] hinting at a relation between these parameters. For the individual sensors, the number required is comparable to the optimal features from SFS. This suggests that the ANN is automatically selecting relevant information from

the combined space and selecting salient features.

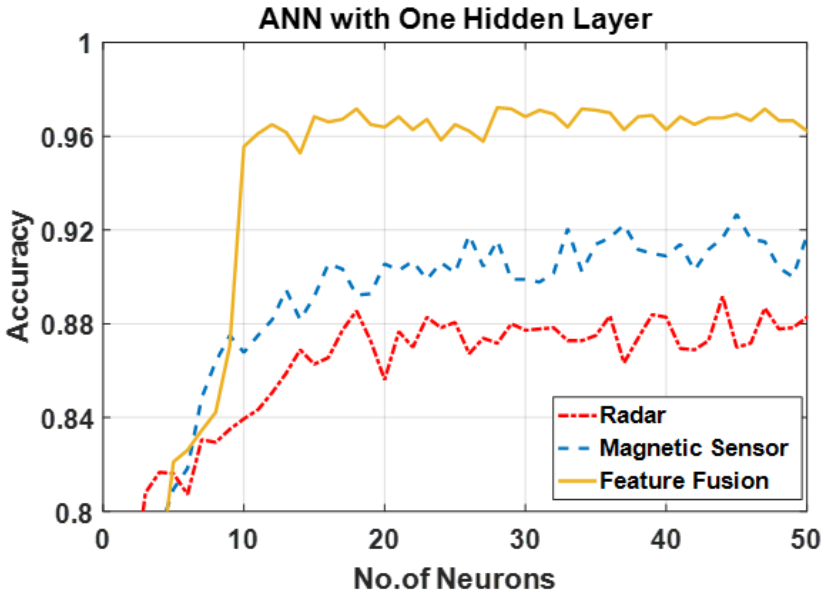

Figure 7.5: Performance of a ANN Accuracy

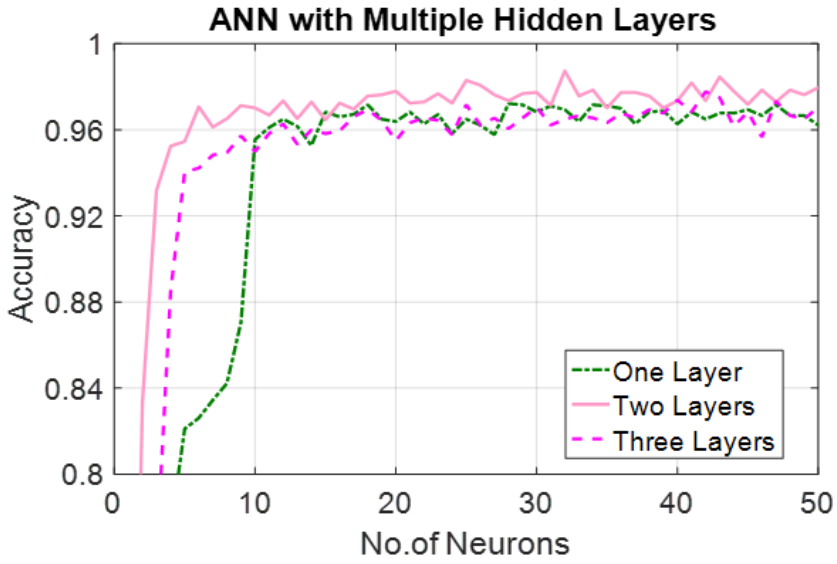

Figure 7.6: Performance dependency of layers of a ANN on accuracy

Fig.7.6 shows the marginality of using multiple hidden layers for the ANN after fusion is performed as there is only a small difference in accuracy when multiple hidden layers are used for the ANN. The number of neurons are varied in the last hidden layer and kept constant at 50 for the other layers however there is a variety of about 0.8% between

using one or two hidden layers with the exception that less neurons are required to reach the optimal point.

In real-life applications, the classifier will often not have prior information from the test subject, yet it should be able to classify the movements from such a person. For the leave one subject out, observations from one specific participant was partitioned from the entire dataset to make the testing set and the remaining participants were used for training. This was performed until all the participants were used as the testing set then the cumulative accuracy was generated [19].

Fig 7.7, which is in [19], shows the lower, average and upper bounds using 'min', 'max', and 'mean' accuracy obtained from these tests. The 'max' and 'min' variables represent the best and worst-case scenario from all the participants while the 'mean' is the average across all participants. The 'difference' variable shows the stratified test and this new approach of "leave one subject out".

The results show that there can be significant variability in accuracy on a subject-by-subject basis, with the extreme case of the magnetic sensor, where both ANN and SVM yield accuracy of approximately 40%. Radar is more robust with both classifiers, as the mean results are 2 to 4% lower than the stratified set. Ultimately, the differences are clearer for the magnetic sensor as the accuracy is 12% lower for both classifiers with this sensor. However, feature fusion has shown to help recover this loss [19].

The minimum value for each of the sensor-classifier combinations displays the challenging issue of activity classification for unknown users as it single sensor classification can be as low as 40%. Despite the use of feature fusion and selection, the difference between this lower bound of performance and fusion is 27% as SVM fusion achieves a much improved accuracy of 67%. For the magnetic sensor, feature fusion helps here as the 'difference' between the two testing methods is reduced from 12% to 4%. Feature fusion therefore is truly necessary for ambient activity monitoring as the additional degree of freedom generated through feature fusion provides a large benefit in the accuracy obtained.

In our experiments, as the participants were given freedom to move in a comfortable way to make the data represent real natural movements. In an extreme case, one participant who moved slowly and was a general outlier has a completely different

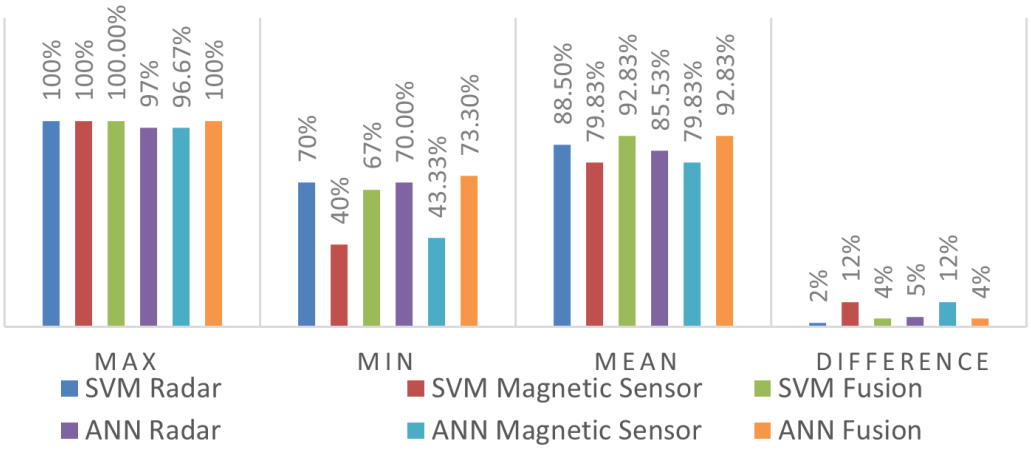

Figure 7.7: Classifier performance when models are tested with an unknown participant. Difference is the delta from the average stratified test when participant information is not explicitly removed from the training set.

signature from the rest of the set. The minimum figure mostly represents cases like this where the target moves uniquely to the training samples, which is expected from new participants who have no prior knowledge of the experiments.

IMU and Radar

Fig. 7.4 presents results for the SFS method for wearable IMU and the radar sensor, with a general summary for all methods provided in table 7.4 (inertial) and table 7.5 (radar). These tables show that the filter methods reduce the number of features used but bring no performance improvements, with less than 2% improvement compared to when all available features are used [27]. Optimal features suggested by filter methods reduce the required number of features by 60% and 35% for inertial sensor and radar , respectively but there are exceptions to this in the Relief-F results for both sensors. For radar Relief-F with KNN increased accuracy by 4% despite being a filter method, while for IMU, Relief-f only reduced required features by 8%. SFS reduced the feature subset while bringing an improvement of 5-7% in accuracy for both sensors with SVM and this was achieved after fusion. KNN on the other hand had no performance improvement for the inertial sensor despite a 9% boost for radar [27]. Finally, for feature fusion between the inertial and radar sensors, the highest classification accuracy was 97.4%

Table 7.6: Decision level fusion results

Method	Average error	Average accuracy (%)
LOGP	9	96.7
Fuzzy logic	14	94.8
Voting system	6	97.8

as shown in Fig. 7.9.

Table 7.4: Comparison of feature selection methods for the inertial sensor

Method	Accuracy (%)	Time(s)*	Features no.
Fscore (SVM)	90.7	1448	73
Fscore (KNN)	88.2	220.2	76
ReliefF (SVM)	91.1	1210.7	164
ReliefF (KNN)	89.3	196.9	58
SFS (SVM)	95.6	14489.5	35
SFS (KNN)	88.25	903.5	69

Table 7.5: Comparison of feature selection methods for the Radar sensor

Method	Accuracy (%)	Time(s)	Features no.
Fscore (SVM)	78.8	220.4	17
Fscore (KNN)	74.1	30.6	17
ReliefF (SVM)	74	213.1	20
ReliefF (KNN)	67	24.2	18
SFS (SVM)	85.6	1316.7	20
SFS (KNN)	79.8	32	19

Table 7.6 shows the results of using the different decision level fusion methods where log opinion pooing fusion was the best performing decision level method with an accuracy of 96.7% with 9 misclassification instances of activities over 10 iterations of testing. Fuzzy logic returned lower accuracy of 94.8% and a higher misclassification of 16 instances.

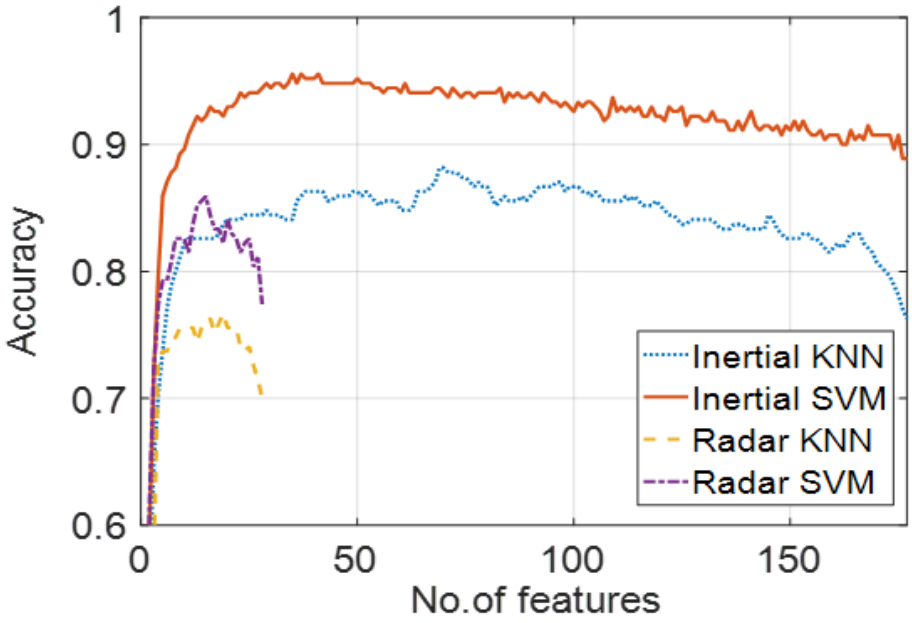

Figure 7.8: Classification accuracy of individual sensors. The improvements through the sequential feature selection and the classifier used are highlighted

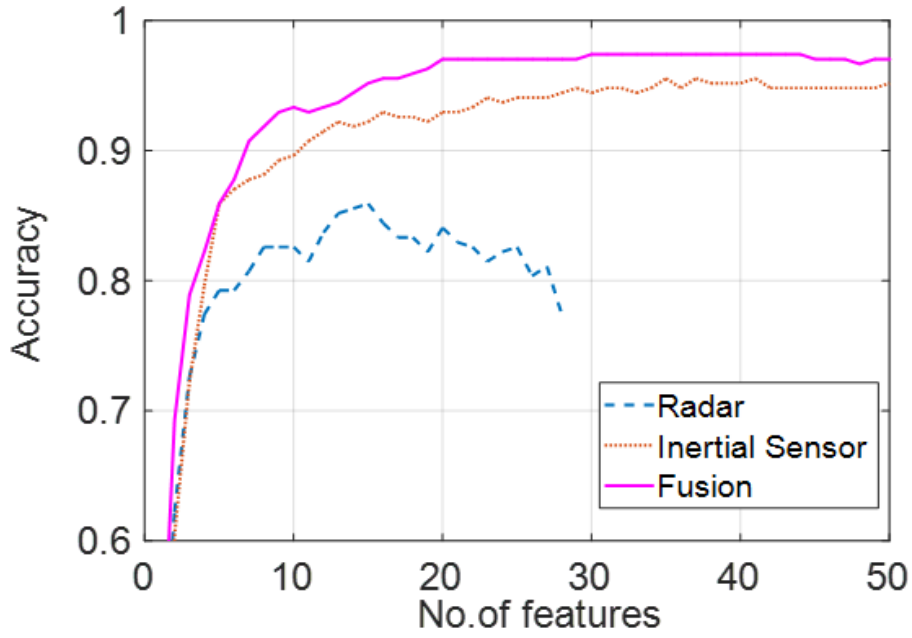

Figure 7.9: Classification accuracy over the number of features utilised.

Table 7.7: Comparison of feature selection methods classification accuracy (%) with radar sensors

Predicted class	CW	FMCW	Fused
A1: Walking	58.82	58.68	61.31
A2: Walking with object	73.89	83.95	75.05
A3: Sitting	70.89	72.89	95
A4: Standing	69.05	80.81	94.22
A5: Picking up an object	83.61	69.94	90.39
A6: Tie up shoes	68.69	74.27	95.14
A7: Drinking from a cup	80.05	78.82	83.14
A8: Taking a call	57.5	72.27	86
A9: Falling forwards	96.23	87.61	95.28
A10: Checking under table	95.67	95.72	96.17

To improve the decision level results, a voting system which considered labels from the classifier for each of the sensor then performed log opinion pooling with the two best performing sensors was constructed. This system gave the best decision fusion result with an average accuracy of 97.8 and 6 error events after 10 iterations.

7.5 Discrete activity recognition: homogeneous data level fusion

In the previous sections, we explored the use of wearable and non-contact sensors, a heterogeneous combination of sensors, to detect activities at a higher rate of accuracy. However, there is a question of what happens when homogeneous sensors are used. To understand this, a smaller experiment with a single CW and FMCW radar were used and feature level fusion was performed to identify the improvements that having two non-contact sensors brings [24]. The processing was from experiment 3 as described in Chapter 5. The same features were extracted for both sensors and the results from the individual sensors and fused case are shown in Table 7.7. The data was split into 70% for training and 30% for testing. The separation was done in a stratified manner so the class ratios were preserved and class imbalance during training was prevented. This process was repeated 20 times after which the average was taken and the mean values are presented in this section. Overall the FMCW radar performs better than the CW

except for some activities which are identified better by the CW but feature fusion is the best performing option. However, in certain cases feature fusion will negate correct decisions from either sensor.

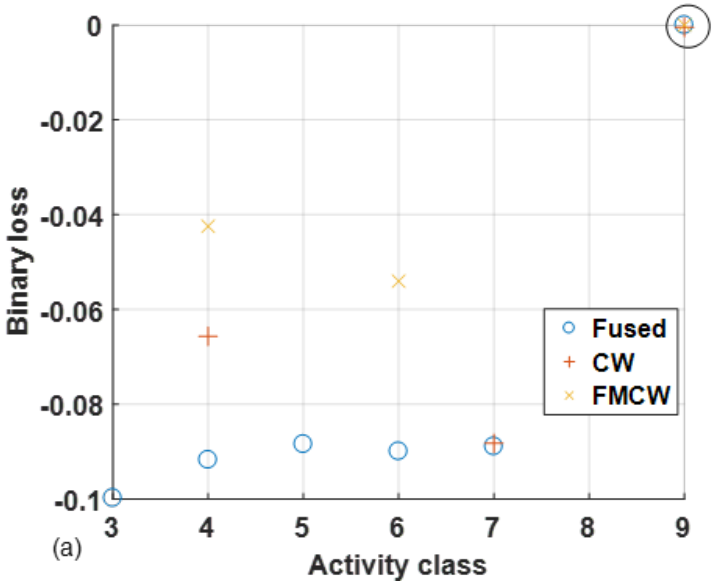

Figure 7.10: Confidence metrics in cases where fusion is required. Case where individual sensors and fusion results in the same class being selected.

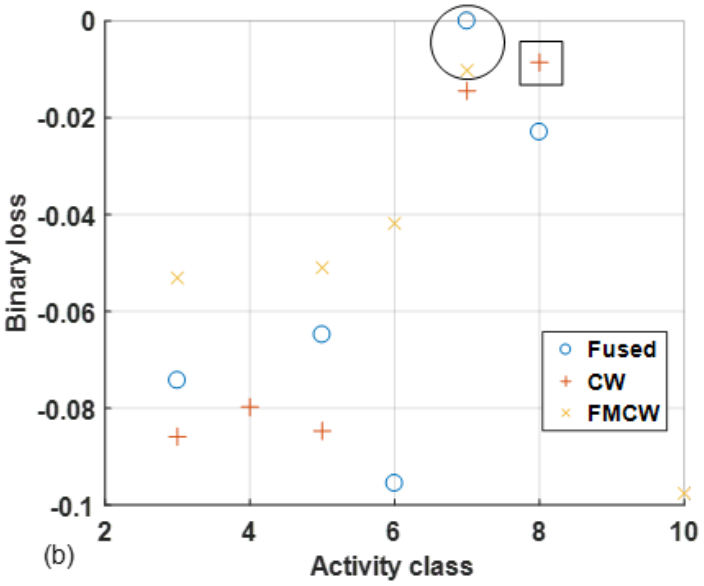

Figure 7.11: Confidence metrics in cases where fusion is required. Case where Fusion is required as only the FMCW sensor selects the correct class while CW is incorrect.

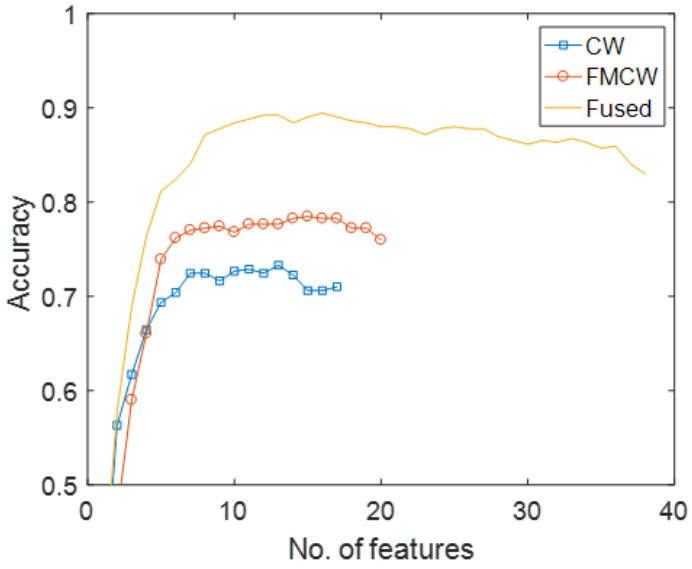

Figure 7.12: Accuracy of the tree radar sensor combinations over increasing features.

One such case is for A2: walking with an object. The accuracy is lower for the fused case at 75.05% compared to FMCW at 83.95%. This is due to the features from CW, which influence the decision as only using CW has 73.89% accuracy. For A1 and A2 the accuracy remains low for both classes even after feature selection and fusion. This is due to both movements having similar motion in walking pattern between the participants.

To understand the effect of fusion fig. 7.10 shows how the confidence score, expressed in loss, changes with fusion. The target class here is A9: fall and we can see the confidence for A9 is the closest to 0 compared to the other activities for all the sensors. This is the ideal case where the sensors correctly identify the class. In fig. 7.5, there is ambiguity in the correct class, as the FMCW identifies A7: Drinking from a cup, while the CW radar picks up the confuser A8: taking a call as the predicted class. Feature fusion helps select the correct class A7 as the incorrect influence from the CW sensor is offset.

Fig. 7.12 shows that SFS does not appear to provide a great increase when the standalone radar systems are used independently. However, it does show that selecting ten features can get the classification accuracy to within 2% of the maximum value attainable in both cases. The real noticeable improvement comes when SFS is used

together with fused features from both radars. With all the features pooled together, there is an improvement in classification to 83% without any feature selection. SFS increases this further to approximately 90%, with only 15 features used out of the 38 total. Out of these 15 features selected in the multi-radar fusion scenario, 4 are extracted from the CW radar data, and 11 from the FMCW radar. Looking at the radar systems individually, the CW selects features from the SVD and physical features (centroid and bandwidth), whereas the FMCW chooses energy curve, centroid, cadence velocity diagram-based features, and SVD-based features. For both systems, the SVD features selected are predominantly from the left singular vectors, in other words, the spectral information. Interestingly, in the fused case, the features selected appear to be different from the individual case. While centroid is present for both cases, the bandwidth from the CW appears to be redundant when fused. Understandably, many of the features generated appear to be covariate with their CW and FMCW counterpart. Additionally, this indicates the presence of less significant features, which could be removed to improve efficiency for implementation into real world systems. There are also class-based differences between the radars for specific activities that is shown by analysing the accuracy of the classifiers on a per class basis. The true positive rate where the classifier has identified the class correctly, in other words the accuracy, is shown in Table 7.7. Here we compare the accuracy for all three situations, where each of the sole radars are used and when they are fused together. Specifically looking at A9: fall; which is the activity of interest in our case the CW system appears to be strong at detecting it compared with FMCW. Fusing here seems to bring the helpful features, which identify falls as the accuracy to a similar level as the CW case. Some redundancy remains, as fall detection accuracy is 1% lower when fused. For activities A3, A4, A5 and A6, fusing improves the accuracy from between 69.05% and 83.61% to between 90.39% and 94.22%. The classification accuracy of this small cluster of activities seems to have increased the most by the cooperative use of radar.

In Table 7.8, aside from the clusters A1/A2 and A7/A8; the remaining diagonal activities are classified to an accuracy above 90%. Notably, for A2: walking with an object, the accuracy is lower for the fused case at 75.05% when compared with FMCW alone with 83.95%. This appears to be due to the features from CW influencing the decision process as it is close to the 73.89% accuracy attained when only CW is used. The

Table 7.8: Confusion matrix for activities for fusion - accuracy and misclassification (%)

%	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	61.32	38.68	0	0	0	0	0	0	0	0
A2	23.94	75.06	0	0	0	0	0	0	0	1
A3	0	0	95	0	2.78	0	0	0	0	2.22
A4	0	0	0.5	94.2	2.42	2.88	0	0	0	0
A5	0	0	0.56	5.78	90.39	3.27	0	0	0	0
A6	0	0	0	1	0.5	95.14	0	1.41	1.95	0
A7	0	0	0	0	0.5	1	83.14	15.36	0	0
A8	0	0	0	2.5	2.5	0.5	8.5	86	0	0
A9	0	0	0	0	0	2.67	0	2.06	95.27	0
A10	0	0	0	0	0	3.83	0	0	0	96.17

Table 7.9: Fall detection false positives - misclassification (%)

Incorrectly predicted as falls (false positive)	CW	FMCW	Fused
A4: Standing	0	3	0
A5: Picking up an object	2.95	2.17	0
A6: Tie up shoes	0	0	1.95

accuracy for A1 and A2 remains low even after feature selection and fusion. This is shown in Table 7.8 where we can see that confusion occurs between the two types of walking movements. However, this is due to both motions having a similar movement and variety in walking pattern from the participants. It is also not a severe outcome as it is not a confuser for A9.

In Table 7.8, we see that there are missed detections for A9: falls and misidentifications of other classes as falls. 1.95% of A6 were identified as falls. Some fall events are detected as A6 but also surprisingly A8, which is not a confuser for A9. Although initially it seems the CW performs better individually, Tables 7.9 and 7.10 show that there are improvements to be seen with fusion for incorrect classification and missed detections too. Table 7.9 shows that a lower proportion of A9 is incorrectly detected but with the fused scenario the erroneous detections occur with A6 as opposed to A5 for CW and A5/A4 for FMCW. The CW radar alone performs similarly with only A5 being incorrectly identified as falls but with higher misclassifications 2.95%. The FMCW is the worst performing system here as it incorrectly identifies 3% of A4 and

Table 7.10: Fall detection misses (%)

Missed falls predicted as (false negative)	CW	FMCW	Fused
A3: Sitting	0.56	0	0
A4: Standing	2.11	1.61	0
A6: Tie up shoes	1.11	5.44	2.67
A8: Taking a call	0	5.33	2.06

2.17% of A5 as A9.

A number of falls are misclassified as other activities, specifically, A3, A4, A6 and A8 and the proportion of which is shown in Table 7.10. Here, we see the 1% loss in fall detection accuracy is offset slightly as without fused features, falls are identified as three different classes for both sole radar systems. Fusing reduces this to two classes that are identified incorrectly, and we also see here that the features from the CW system are offsetting the errors from the FMCW. The bias towards A8 that the FMCW has still appears to be there to some extent as 2.06% of A9: falls are still predicted as A8.

7.6 Conclusions

In this chapter, we have overviewed sensor fusion methods and discussed relevant results. Compared to using a single sensor, multimodal sensing significantly improves the activity pattern-recognition performance. Different machine learning algorithms and sensor combinations have been evaluated separately with feature selection techniques utilized, namely Fisher score, Relief-F and Sequential Forward Selection. We have shown the stages where fusion can take place, specifically at signal, feature and decision level. By **using different sensing technologies cooperatively** and **proving that utilizing features from different domains with feature selection increases activity recognition accuracy**, we addressed the question of **how sensing methods can be improved to optimise the classification accuracy in the context of assisted living**.

With heterogeneous fusion, an accuracy 97.4% of sensors is achieved with feature fusion with 20 features, while for homogeneous fusion, twenty-one FMCW-derived radar features are required to achieve an accuracy of 75% where the CW radar performs to a

more limited degree, with 70% accuracy when 19 features are used. With feature fusion between the radars, an improvement of classification to 83% occurs. SFS increases this further to almost 90% with only 15 features. Out of these 15 features selected in the multi-radar fusion scenario, 4 are extracted from the CW radar data, and 11 from the FMCW radar, therefore SFS appears to also highlight the salient sensor.

Furthermore, we demonstrated that fusing the features from the FMCW and CW radars brings about improvements in classification accuracy in general, but also for missed detection and misidentification of falls as other classes. Using data from the radar returns of the 20 participants, classification rates of up to 89.54% was achieved with the help of fusion along with SFS. The central property of feature fusion appears to be bringing in the strengths of the feature set from both sensors and SFS which seem to simply enhance this as it has more choices to make the optimal feature selection.

Chapter 8

Continuous Activity Recognition with Bi-LSTM

Coming to the end of the thesis and after having overviewed the many challenges and solutions for discrete activities, it is clear that the research interest has been progressing towards using edge artificial intelligence (AI) techniques, with deep neural networks coming more and more prominent. This chapter mirrors this trend as the process of how automatic feature generation methods are utilized for activity recognition by generating salient features for classification of complex and continuous activity sequences is reviewed. Utilising a temporal neural network design, specifically LSTM and Bi-LSTM, a spectrogram signature composed of sequences of six activities are classified without handcrafted feature extraction. The main contribution of this chapter is the generation of a network architecture employing the Bi-LSTM network then testing and validating it with continuous data.

This work has been previously covered in the publication [21] and the remainder of this chapter is organized as follows. Initially the discussion on the motivations for using Bi-LSTM networks is conducted, followed by the description of the experimental setup, data collection, and data pre-processing. Section 8.3 presents a description of the algorithms used and the results obtained with LSTM and Bi-LSTM networks used for and offers some insight on optimizing performances. Finally, section 8.5 concludes the chapter and outlines possible future avenues of research.

8.1 Motivation for continuous activities and temporal classification networks

As discussed in the introduction, the research focus on human activity classification with radar has been on discretely separated activities, which are usually performed and recorded individually. For the analysis of continuous activities, discrete data samples can be sequentially concatenated as in [82, 111]. However this does not capture the full realism of unconstrained human movements, where the duration of each action can change, and the inter-activity transitions can happen variably.

To evaluate this realistic scenario, the data set utilised in this work includes sequences of continuous activities performed without restrictions or instructions by the participants. This process also captures the diversity in sequential order and dynamic transitions between the activities. Continuous recordings of radar data can appear in time as sequences of range profiles, stacked together to form range-time matrices, or alternatively they can be represented as micro-Doppler spectrograms [112]. As discussed in Chapter 4, the majority of the works in the literature would interpret these radar data as two dimensional images or three dimensional data cubes, and process them with methods used by the image processing community, such as convolutional neural networks or auto-encoders. In this method, a sliding window of fixed length can be applied across the sequence of radar data to extract images of individuals or sub-sets of activities. In a realistic sequence of human movements however, there is no fixed temporal duration of each action, and furthermore, the transitions between actions can happen randomly. Therefore, rather than akin to images, these continuous radar data are akin to sequences of speech or audio signals where individual words or patterns can appear at any time and for an unconstrained period.

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 Bidirectional LSTM.

The main property of the LSTM is its memory capability to capture long-term dependency between data separated by a significant duration [48]. 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.

Then, bidirectionality is the capability of correlating the data processed at a given timestep with both learned data from past and expected future timesteps [113]. This is again a key property in speech/language processing to capture the connection between different words in a long sentence, but also a relevant capability in radar-based activity classification to understand the kinematic constraints of human movements.

8.2 Training/testing set composition and learning library

A total of 48 different sequences were collected while 3 sequences were repeated recordings performed by each individual subject, and set aside as the validation test. From the remaining 45 sequences, the testing test included one of the 45 sequences, repeating the process 45 times to test all sequences. For the training set, two different approaches were followed to evaluate the effect on the classification performance of prior knowledge/data about a specific participant.

In the approach labeled as "*New*," the two sequences which belonged to the subject under test were removed from the training set, leaving a total of 43 sequences for training. In this case, the test subject is unknown to the classifier, as if a new person joining the experiment. In the approach labeled as "*Known Prior*," two random sequences out of 45 from other participants were removed from the training set, leaving 43 sequences for training therefore maintaining consistency with the previous case. However, the two sequences performed by the test subject were purposely kept in the training set which meant in this case, the classifier did retain some knowledge of the test subject

through the two sequences in the training set, although the order of the activities and the related transitions were different. These two approaches were tested to evaluate if any prior knowledge of an individual human subject would affect the classification algorithm and its performance. For the generation of input data, in this experiment, a 0.2s Hamming window and an overlap factor of 95% is used to generate the micro-Doppler spectrograms. The LSTM and Bi-LSTM neural networks can use either range profiles or spectrograms as inputs the sections below explore the effect of using them.

As for the algorithm library, While others have used various libraries [16] [82], this work utilized the Deep learning toolbox included in MATLAB 2018B. Furthermore, MATLAB was used additionally for radar signal processing and for creating the class labels for each time step by manually labeling the ground truth data from physical observation during the experiments.

8.3 Experimental 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 hyper-parameters of the networks (e.g. learning rate).

In this section, lower case symbols will denote vectors, e.g. x, whereas matrices are denoted by upper case letters H. An arrow pointing right, e.g. $\overrightarrow{H_t}$ indicates the scalar or vector in the next time step whereas an arrow pointing left, e.g. $\overleftarrow{H_t}$ indicates the scalar or vector from the previous time step. $\odot$ denotes the Hadamard product, an element-wise product of two vectors.

This experiment utilised the dataset 5 which has been overviewed in Chapter 5 but to reiterate, the data included six human activities: walking (A1), sitting on a chair (A2), standing up (A3), bending to pick up an object (A4), drinking a glass of water (A5) and simulating a frontal fall (A6). These activities are shown in Fig. 5.6. With the three sequences being :

- A1: A2: A3: A4: A5: A6

- A5: A4: A2: A3: A1: A6
- A4: A5: A1: A2: A3: A6

8.3.1 Doppler LSTM

The first network investigated is a two-stage stacked LSTM network, which is referred to as *Doppler-LSTM* and serves as a baseline for the spectrogram-based results. As the name suggests, the input to this network is the spectrogram, which contains micro-Doppler information and is fed into the network as a sequence of different vectors time bin after time bin.

Table 8.1: Size and property of layers used in *Doppler-LSTM* network

Layer	Size	Properties
Input	240	based on the frequency bins of the input spectrogram
LSTM	2400	number selected to store large sequences in memory
LSTM	2400	number selected to store large sequences in memory
FC	6	Based on the number of possible output classes
Output	1	Single output

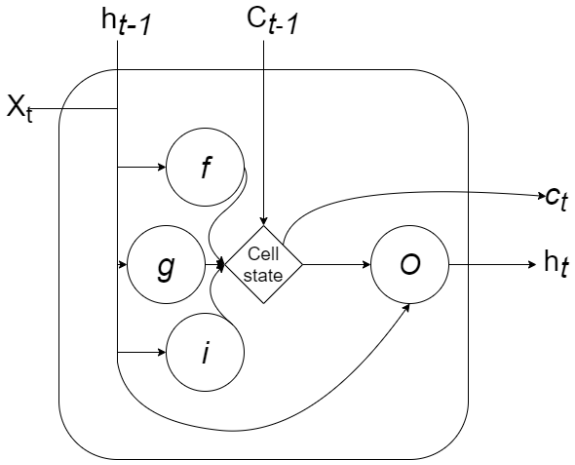

Figure 8.1: Overview of the LSTM cell used in *Doppler-LSTM*Note that the arrows indicate only forward-based temporal information flow from a timestep $t-1$ to the following t

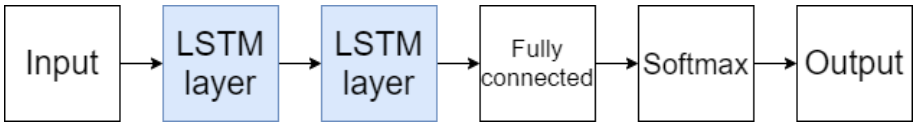

Figure 8.2: The network architecture of *Doppler-LSTM*

With the temporal dependencies accounted for, the level of abstraction in the input data should be assessed, as spectrograms can be considered a mixture of multi-tones where the micro-Doppler movements induce different Doppler frequency components depending on the movements of individual body parts. Using multiple layers has been suggested as the primary method of detecting higher-level abstractions from the input domain [82]. Therefore as shown in fig 8.2, the proposed Doppler-LSTM network has two stacked LSTM layers so that these higher-level abstractions can be identified by the network.It is comprised of:

- An input layer that takes a segment of the spectrogram (250 Doppler bins in each time bin, which is equivalent to one observation) and sends it to the first hidden layer,
- Two stacked LSTM layers that extract and update the salient features in the input data,
- A fully connected layer that connects the activations of the different LSTM layers necessary for classification,
- A softmax that computes the probability distribution of the data belonging to a specified output class,
- An output layer that outputs the class label based on the Softmax distribution.

Note that the arrows indicate the temporal direction of the recurrent LSTMs. In this case, since a standard LSTM layer is used, only forward based recurrence is considered. The neural networks have 2410 hidden cells for both of the LSTM layers and a learning rate of 2^{-4}. Of the editable hyperparameters, the learning rate is of significance as it is the key contributor to vanishing and exploding gradient problems [73]. This hyperparameter affects radar data significantly and can offset good architectural decisions if an incorrect learning rate is used. The state activation function is the hyperbolic tangent,

while the gate activation function is a sigmoid (rectified linear unit is not commonly used with LSTM as it can cause exploding or vanishing gradients.) and gradient descent optimizer is "Adam." Table 8.1 shows a summary of the size and properties of the layers of the *Doppler-LSTM*.

- f forget gate : Control to forget cell state
- i input gate : Control to update cell state
- g cell candidate : Control for information to be added to cell state
- o output gate : Control to add cell state to hidden state

8.3.2 Doppler Bi-LSTM

The second proposed network referred to from hereon as *Doppler Bi-LSTM*, is a modification of the first one and includes: an input layer, an LSTM layer, a BiLSTM layer, a Softmax layer, and a classification layer. The Bidirectional LSTM cell is the main modification of this network and its details are shown in Fig. 8.3, whereas Fig. 8.4 shows the block diagram of all the layers of the proposed network.

Similar to the first network, *Doppler Bi-LSTM* accepts spectrograms as inputs. Differing from the previous Doppler-LSTM, this network processes the forward time-based dependencies first in the initial LSTM layer, and then searches for bidirectional, forward and backward, dependencies in the extracted temporal features. The capability of characterizing and memorizing these forward and backward dependencies in the sequences of data is critical for this network and its performance, as in the sequence of human activities, there are explicit dependencies and kinematic constraints on the order of possible actions. The main equations for a Bi-LSTM cell unit [114] are as follows:

$$\overrightarrow{h_t} = tanh(W_{x\overrightarrow{h}} X_t + W_{\overrightarrow{h}\overrightarrow{h}} \overrightarrow{h}_{t+1} + b_{\overrightarrow{h}}) \tag{8.1}$$

$$\overleftarrow{h_t} = tanh(W_{x\overleftarrow{h}} X_t + W_{\overleftarrow{h}\overleftarrow{h}} \overleftarrow{h}_{t+1} + b_{\overleftarrow{h}}) \tag{8.2}$$

$$y_t = W_{\overrightarrow{H}y}\overrightarrow{h_t} + W_{\overleftarrow{h}y}\overleftarrow{h_t} + b_y \tag{8.3}$$

The main difference in the Bi-LSTM layer versus the previously described LSTM layer comes from each cell having two hidden states, with two parallel pipelines feeding to both previous and next timesteps as illustrated in Fig. 8.3. Note that in this figure, the two different hidden states are denoted with capital H and forward and backward arrows, respectively, as they are in the equations. Differently from the LSTM, in the Bi-LSTM layer, the interconnections between the input, output, and hidden states through the relevant weights do not propagate through the forward and backward cells directly;

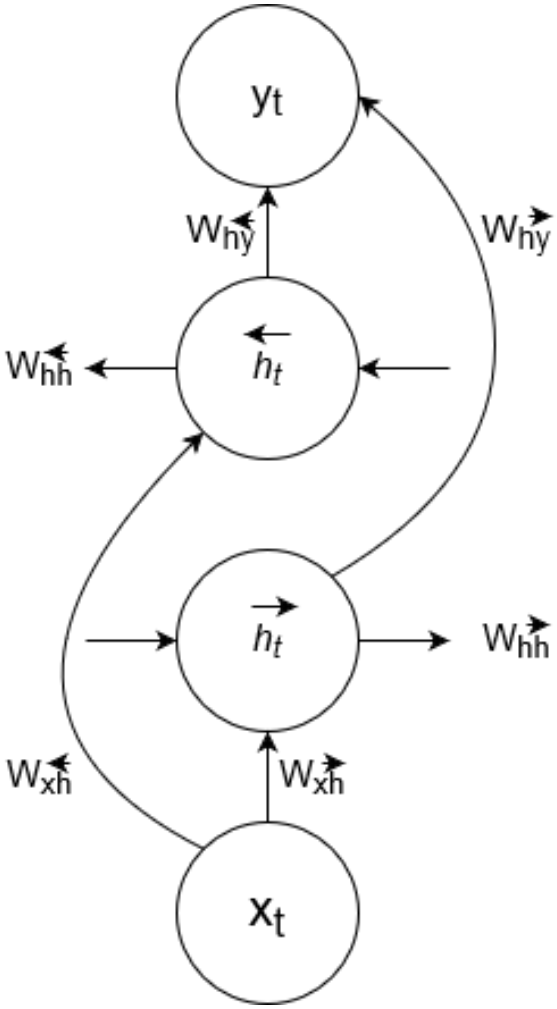

Figure 8.3: Interconnections and weight transfers in a Bi-LSTM cell used in the *Doppler Bi-LSTM*. The arrows show the propagation of the information hidden and cells states between the layers. X_t is the input, h_t is the hidden state with its forward or backward directionality, W_{nn} indicates the weights linking hidden states and outputs/inputs and Y_t is the output.

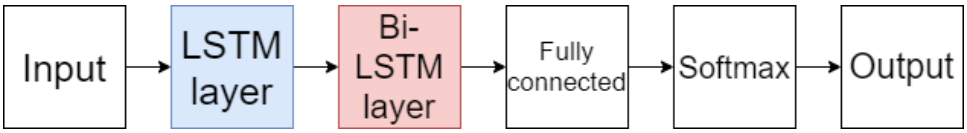

Figure 8.4: Network architecture of *Doppler Bi-LSTM*. The key difference between the previous architecture is the presence of a bidirectional layer.

instead, they interface separately by going through the forward cells ($\rightarrow$) and backward cells ($\leftarrow$) at the same timestep. The hidden states from these forward and backward cells are then combined to generate the output from the Bi-LSTM layer, denoted by y_t. The implication is that the Doppler information corresponding to specific body movements over a long duration in both forward-time and backward-time directions are characterized and captured by the Bi-LSTM layer. Essentially, this means that the network searches and memorizes recurring feature patterns in the past (previous actions) and any linked recurring feature patterns in the future (subsequent actions).

8.3.3 Doppler LSTM and Doppler Bi-LSTM performance analysis

Fig. 8.5 shows the spectrogram of one of the sequences classified by the Doppler-LSTM. Furthermore, it shows the comparison of the classification and ground truth of the activities within this sequence. Initially at t=0, we see that there is a sharp spike that detects A5: Drink while in truth the person was performing A4: Pick, since both of these activities have the central component of moving arms the classifier has a moment of indecision. It then correctly classifies A4: Pick, but it detects A5: Drink with a delay of 3 seconds, after which another "impulse-like" indecision, referring to the sharp spike at about 9 seconds, where A6: Fall is detected. In a fall detection system, the presence of these spikes for erroneous classifications could be undesirable

Table 8.2: Size and property of layers used in *Doppler Bi-LSTM* network. Optimal learning rate of 1e-4 is used

Layer	Size	Properties
Input	240	based on the frequency bins of the input spectrogram
LSTM	2400	number selected to store large sequences in memory
Bi-LSTM	2400	number selected to store large sequences in memory
FC	6	Based on the number of possible output classes
Output	1	Single output

and a potential source of false alarms.

Fig. 8.6 represents an example of results for the Doppler Bi-LSTM network. Note that the sequence of activities is the same as the one presented in the previous figure, but performed by a different subject. For this reason, the ground truth plots are identical, but the input spectrograms appear overall similar. Comparing the classifier output/test outcome in the orange line and the ground truth in the blue line, we can see that test outcome matches the classifier output to a very large extent. However, there are three noticeable segments at time points 4, 20, and 22 seconds where there is a slight mismatch between the test outcome and the observed ground truth. In the first case, at 4 seconds, there is a short delay in detecting the transition from A4: Pick up to A5: Drink. However, one can note that in the spectrogram input in Fig. 8.6, the signature is unclear at that time instance, with a difficult transition detectable by eye. This is typical of transitions where the dynamic range of the macro movement of the body/torso and the micro-movements of the limbs change drastically. The network may respond to this by maintaining the classification from the previous time instances,

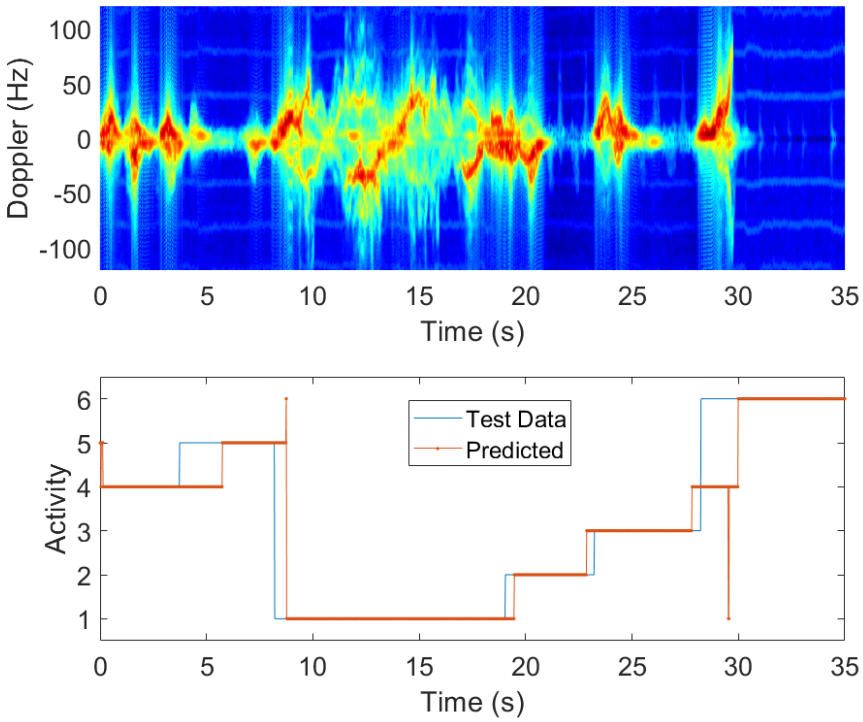

Figure 8.5: Classifier input at the top sub-figure, ground truth in blue, and the test outcome in orange in the bottom sub-figure for a test sequence for the *Doppler-LSTM* network.

so there is a short delay but no erroneous classification occurs. This is similar to the second case, at 20 seconds, where the classifier appears to detect A6: Fall with a short delay. In the third instance, at 22 seconds just before it happens. Reviewing the spectrogram, prior to the A3: Standing occurring, there is a precursory movement which the classifier notes and associates as part of the A6: Fall class, possibly due to the knowledge of the signature at future time instances in the spectrogram provided by the bidirectional capabilities of the network.

Fig. 8.7 shows the classification accuracy for the 45 sequences collected where each one was the test sequence in turn, as discussed in section II.C. Note that the hyperparameters and training/testing approach were kept consistent between the two network architectures. This provides a more effective way to compare the performances of the proposed Doppler LSTM and Bi-LSTM networks across the whole dataset of continuous signatures, rather than observing individual sequences.

The range of classification accuracy is, on average higher for the Bidirectional LSTM

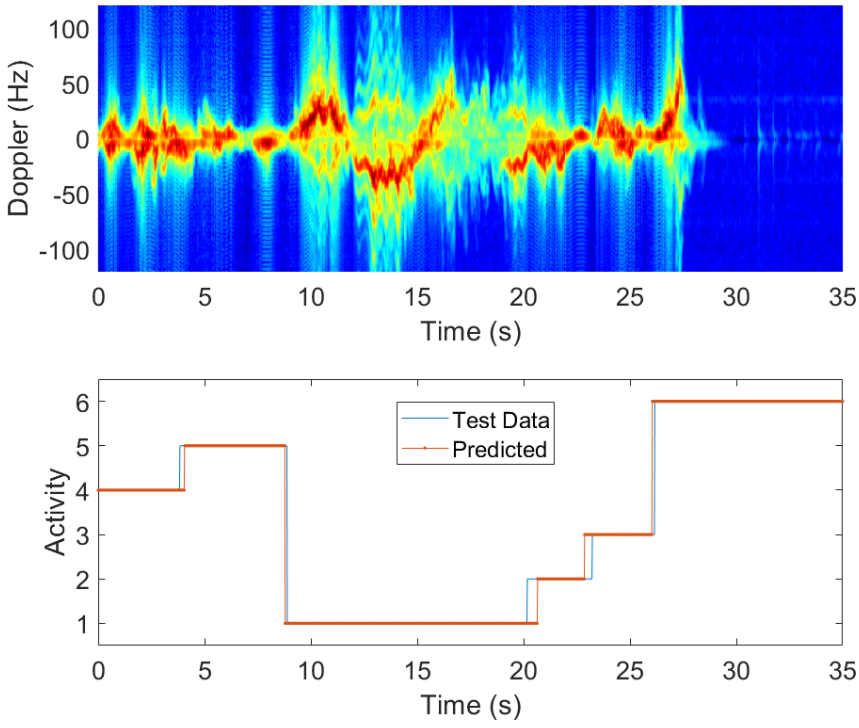

Figure 8.6: Classifier input at the top sub-figure, ground truth in blue, and the test outcome in orange in the bottom sub-figure for a test sequence for the *Doppler-Bi-LSTM* network.

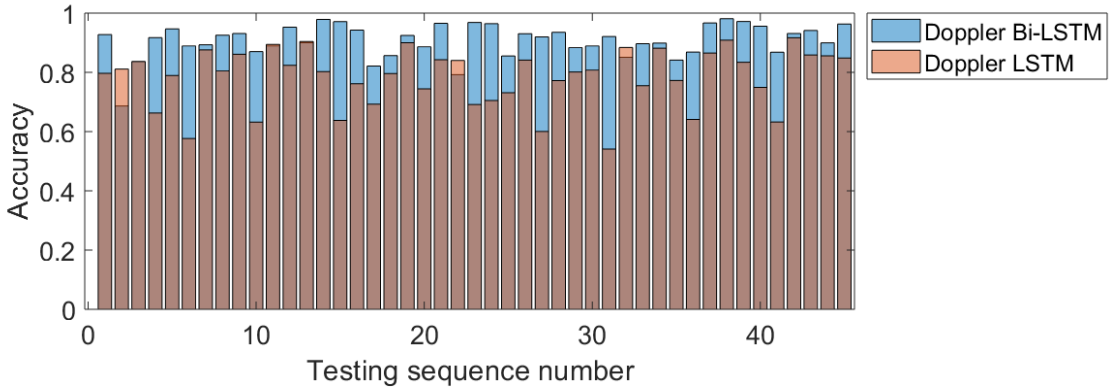

Figure 8.7: Comparison between *Doppler-LSTM* and *Doppler Bi-LSTM* architectures output as classification accuracy over the 45 test sequences. Although the layers between these classifiers are different, the hyperparameters and training and testing methodologies are consistent between both network architectures.

network compared with the unidirectional LSTM, and there is less variability across different subjects and different sequences performed by the same subject. This can be described by the mean and standard deviation across the 45 classification tests, that are recorded in Table 8.3 for the Doppler Bi-LSTM and LSTM networks. The mean increases to approximately 91% from 78%, whereas the standard deviation is reduced; the maximum (best case) and minimum (worst case) are also increased when using a bidirectional architecture of approximately +6% and +15%, respectively.

The metrics in Table 8.3 and the detailed results in Fig. 8.7 for each test sequence show how the bidirectional capability of the proposed Bi-LSTM provides superior capabilities to classify human activities in a continuous sequence with respect to a conventional unidirectional LSTM (for example the increase in accuracy from LSTM to Bi-LSTM is +30% in the best case of sequence #31). The robustness and good generalization of the proposed approach across the diverse set of 45 recorded sequences and 15 subjects are demonstrated. Furthermore, for subsequent sequences where there is a drop in the accuracy, the Bi-LSTM appears to be more robust than the unidirectional LSTM, for example for sequences 5 and 6, where for the Bi-LSTM the accuracy drops from 94% to 88% and for the LSTM it drops from 78% to 57%. As a note, the accuracy in Fig. 8.7 and Table 8.3 is calculated as the number of correct classification of the activity (1 out of the 6 performed) in each time bin of the spectrogram, over the total number of time bins in the 35s total duration of each testing sequence. We think that this

Table 8.3: Accuracy metrics from the tested participants across different LSTM architectures with Doppler and Range input.

Classifier	Mean	Standard deviation	Maximum	Minimum
Bi-LSTM	91	5	98	69
LSTM	78	9	92	54
Range BiLSTM	76	7	87	54

is a conservative approach, which labels as mistakes even the very short spikes with erroneous prediction lasting for only a few time bins, as seen in Fig.8.5. As part of future work, smoothing filters approaches could be applied to the predictions of the LSTM or Bi-LSTM networks to disregard labels for activities that would last only for few time bins, and therefore be unrealistic as the subject could not perform a given activity in such short physical time.

8.3.4 Range Bi-LSTM

In the previous sections, we have analyzed the results of using Doppler spectrograms input to the LSTM and Bi-LSTM networks, but spectrograms need an additional level of processing after the generation of the range profiles to be calculated. This prompts the question of whether sufficient information can be inferred from the data in the range-time domain, leaving to the networks the task of extracting the Doppler information, i.e. the changes between subsequent range profiles implicitly.

Range profiles do not show the different activities in the signature in an easily perceivable manner compared to spectrograms since only the location relative to the radar is given, and in the specific case of our radar, the range resolution is limited to approximately 40cm with the 400 MHz bandwidth. Hence, to the human eye, the different activities in the range-time plots appear much less distinguishable than in the spectrograms, but this may not necessarily be a limitation for neural networks.
Fig. 8.8 shows an example of such a range-time plot in its top part; it is evident how this image is less clear than the corresponding spectrograms in Fig. 8.6 and Fig. 8.5 for the same sequence of actions. A significant difference between spectrograms and range-time plots, is the number of time bins, of temporal units that the LSTM or

Bi-LSTM network will need to process. For 35s of data in each sequence of activities, spectrograms consisted of 1750 time bins with the selected STFT parameters, whereas range-time plots had 35,000 observations or range profiles, as the data were sampled at 1 kHz PRF. This increased size of the data led to a modification of the network with a different number of inputs reflecting the number of range bins in the range profile. This network is referred to as *Range Bi-LSTM*. In terms of its layers and architecture, the *Range Bi-LSTM* is similar to the one shown in Fig. 8.4.

The bottom part of Fig.8.8 shows an example of representative results from this alternative network using range-time data as input. The performance is reduced compared to the Doppler based networks.

At 0, 4, and 5 seconds, we see transient detection of A5: Drink at multiple instances until an apparent misdetection of activity A1: Walking as A2: Sitting as the target comes to a halt which is visible in the range-time plot. This is followed by an early detection of the A2: Sitting. At 28 seconds, multiple instances of A5: Drinking is detected before A4: picking up item is correctly identified, which is reminiscent of the spike transients observed with the Doppler-LSTM. In general, more spikes and instability in providing a steady prediction are shown at other transitions, and there are misdetections of all activities throughout the sequence.

8.3.5 Range-time Bi-LSTM performance analysis

The results in Fig.8.8 for the usage of range-time data as inputs to the Bi-LSTM show a degradation in performance compared to the usage of micro-Doppler information. To view the performance of the network on a sequence by sequence basis across the whole dataset, Fig. 8.9 shows the results for the 45 test sequences, and Table 8.3 shows the overall performance metrics.

There are cases where the Range Bi-LSTM performs well. For example, in the sequences 31-35, an accuracy of approximately 80% is attained. However, it does not maintain this rate for the all of the test sequences as the classification challenge of detecting complex activities designed for this set, and also delayed and transient detection of classes occur as demonstrated in Fig. 8.8. Viewing Table 8.3 while comparing Fig.8.9 and Fig. 8.7, show the performance loss of using the range-time profiles as

inputs to the proposed LSTM networks, despite the potential advantage of avoiding the calculation of spectrograms at the pre-processing stage before the network. To put it into perspective, the best classification accuracy, or the maximum in Table 8.3 (87%), for any range input is 4% less than the mean accuracy for the Doppler Bi-LSTM (91%). In other words, the best case with range input cannot match the average case with Doppler input with a similar or even a range focused network architecture. Directly comparing the mean accuracy shows an improvement of 15% through the use of

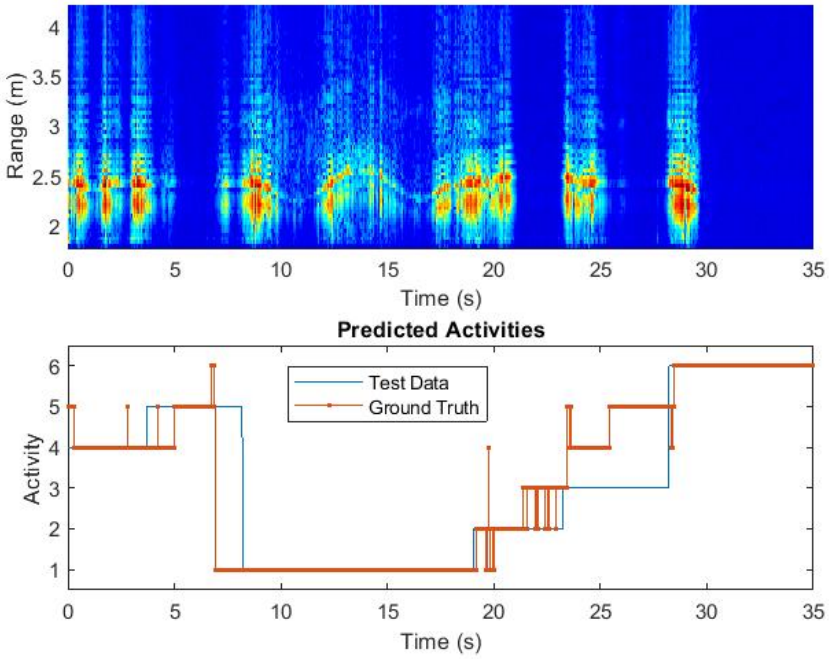

Figure 8.8: Classifier input, ground truth, and the test outcome for a test sequence for the *Range Bi-LSTM* network.

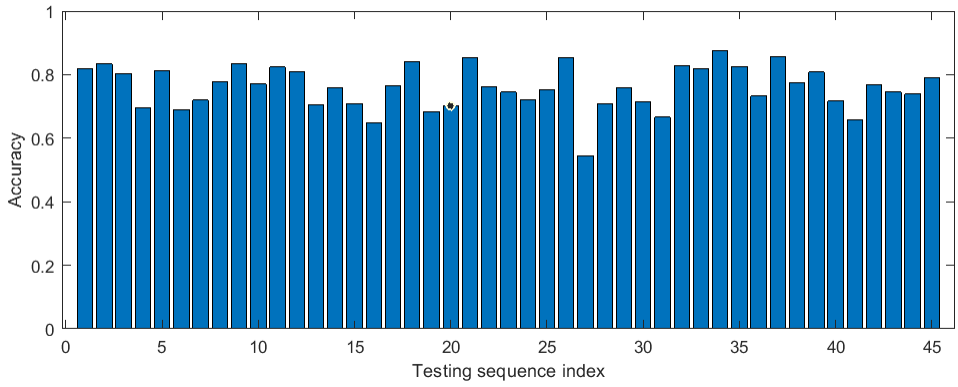

Figure 8.9: *Range Bi-LSTM* output as classification accuracy over the 45 test sequences.

the Doppler Bi-LSTM (91%) instead of Range Bi-LSTM (76%) and improvements in the maximum by 5% and minimum rate by 15% when the former architecture and its corresponding input is used.

8.4 Further experimental validation

In this section we present further tests to validate the proposed methods. We discuss the influence of the classifier having prior information from participants, compared to when no such information is provided. Additionally, the effect of static clutter on the classification is analysed and a comparison with a simpler support vector machine classifier (SVM) is made, with a comparative analysis is provided.

8.4.1 Known Prior vs Unknown

Table 8.5 shows the results from the networks and input domains discussed in this chapter with the "Known Prior" and "New" training and testing methodologies. For

Table 8.4: Size and property of layers used in *Range Bi-LSTM* network. Layers were resized to fit input domain and memory limit.

Layer	Size	Properties
Input	64	based on the range bins of the input spectrogram
LSTM	240	number selected to store large sequences in memory
Bi-LSTM	240	number selected to store large sequences in memory
FC	6	Based on the number of possible output classes
Output	1	Single output

Table 8.5: Accuracy metrics for range vs Doppler domain networks with "Known prior" and "New" training and testing approach.

Classifier	Subject	Mean	Standard deviation
Range Bi-LSTM	Known Prior	74	7
Range Bi-LSTM	New	77	7
Doppler LSTM	Known prior	78	10
Doppler LSTM	New	78	9
Doppler Bi-LSTM	Known prior	91	4
Doppler Bi-LSTM	New	90	6

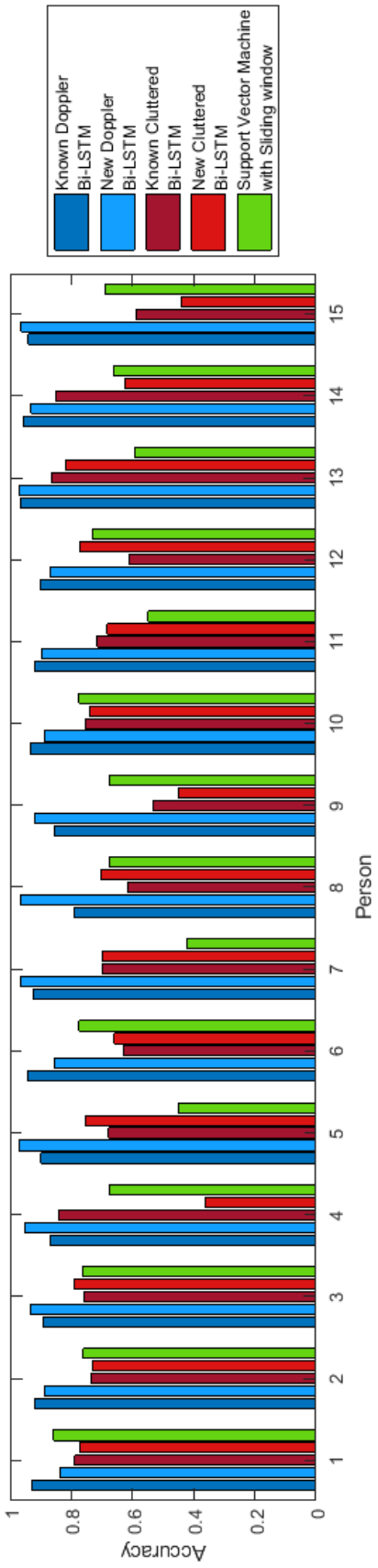

Figure 8.10: Comparison between testing methodologies: where the network had samples of the test subject but with a separate sequence called Known prior; and another where the network had no prior information from the test subject called new. Repeated for scenario with and without clutter and SVM results presented as a classical classifier for comparison. Results generated with the *Doppler Bi-LSTM* network.

each of the participants, one sequence of activities was taken as a test sample with the best performing classifier and input domain combined. In the case of *Known Prior*, the training set had included the other two sequences performed by the same subject (but with a different order of activities), whereas for the *New* set, the classifier did not have information on that specific subject from the other sequences. The Table shows there appears to be a marginal difference in the prediction between a classifier that has prior information from the test subject and one which does not. The only factor inducing a significant change is the selection of the network and corresponding input domain, where it follows the trends discussed at the end of section III-E. Substantially, the Doppler Bi-LSTM outperforms the other architectures/input domains. As the prior knowledge of the test subject induces no significant change in the classification accuracy, we assume that during the data collection the same activities were not reproduced in the exact form in all sequences, despite the test subject remained the same. This was due to the fact that the duration of each activity was unconstrained and that as the order was different, the transitions between the activities happening before or after a given one created diversity in the data. Therefore, each sequence is distinct, and there is no much difference in providing to the network knowledge of other sequences at the training stage for a given test subject. Conversely, it can be seen that, when the best performing combination of network and radar data format is used, the proposed approach based on recurrent LSTM networks is robust enough to generalize across the cohort of 15 subjects and 45 sequences. The analysis of some of the individual test sequences for the "Known Prior" and "New" approaches in Fig. 8.10 shows cases where prior knowledge appears to help with improving accuracy. For example, with person 1 and 6, there is an improvement of more than 10%. However, there are also cases where the opposite occurs, for example, with person 8. Therefore, prior knowledge of the test subject with the classifier does not appear to be a major factor in the overall accuracy.

8.4.2 Influence of static clutter

The role of clutter is another aspect which is questioned about in the research area of using radar for monitoring human activities as indoor environments consistently have objects which generate static clutter and possible multipath. This is usually mitigated

by using MTI filters, as it has been done in this work since it removes the effect of static clutter. To demonstrate the impact of clutter, the signatures without MTI filtering have been used as the input to the classifiers. Fig. 12 shows the results when the spectrogram signature has the filter removed, therefore the effect of background clutter on the Doppler signature is included. Both cases of "Known Prior" and "New" approaches for training the Doppler Bi-LSTM network are reported for the cluttered data.

There are certain cases, e.g. Person 1 and person 13, where the presence of clutter results in a decrease of approximately 12% from the regular case where the MTI filter is present. Similarly, there are cases where an extreme decrease is present, e.g. person 15 where for the "Known Prior" cases the presence of clutter has a 50% decrease in accuracy compared to the filtered/regular counterpart. Incidentally, there appears to be a marginal benefit in this case where prior knowledge is useful in classification as in average, there is a 3% difference between the "Known Prior" and "New" cases with clutter considered as shown in Table 8.6. With Fig.12 and Table 8.6, we see that there is a decrease in performance for all the participants in both "Known Prior" and "New" cases. This suggests that filtering static clutter is essential to ensure accurate recognition of sequences with the proposed method.

8.4.3 Comparison with conventional Support Vector Machine

A simpler classifier, a support vector machine, is used with features derived from segmented windows of the whole sequence to detect the activities, and the results can

Table 8.6: Accuracy metrics for the Doppler Bi-LSTM with vs without static clutter filtered, for "Known prior" and "New" training and testing approach. SVM results also shown for comparison

Classifier	Subject	Mean	Standard deviation
Doppler Bi-LSTM	Known prior	91	4
Doppler Bi-LSTM	New	90	6
Cluttered Doppler Bi-LSTM	Known prior	71	14
Cluttered Doppler Bi-LSTM	New	68	12
SVM	Both	66	11

be then compared to those generated by the proposed LSTM networks. This is done to establish a benchmark for the Bi-LSTM architecture and to validate their use for this classification problem. SVM utilises as input a selection of features extracted from the centroid, bandwidth, and singular value decomposition of the spectrogram signature. This has been previously used in literature to identify discrete activities [115, 18], together with a sliding segmenting window. The window length was 4.5s and overlap was 90% and the kernel of the chosen SVM was linear. These parameters for the sliding segmenting window approach were selected as they resulted in the best performance in previous work [116].

Table 8.6 shows a summary of the results for the SVM to compare them with those of the Bi-LSTM networks; the results for individual sequences were shown in Fig. 8.10. Note that there was only negligible difference in the SVM case between the "Known Priori" and "New" approach for training the classifier, hence they are reported together in the table. In general, the SVM results (green bars in Fig. 8.10 are lower than using Doppler Bi-LSTM networks (blue bars in Fig. 8.10. While the performance can be close in rare instances, e.g. for Person 1, in general the SVM is between 20 to 40 % lower when compared to the regular Doppler Bi-LSTM. This result emphasises the value of the proposed approach with a temporal-aware classifier such as the proposed Doppler Bi-LSTM for recognising continuous human activities.

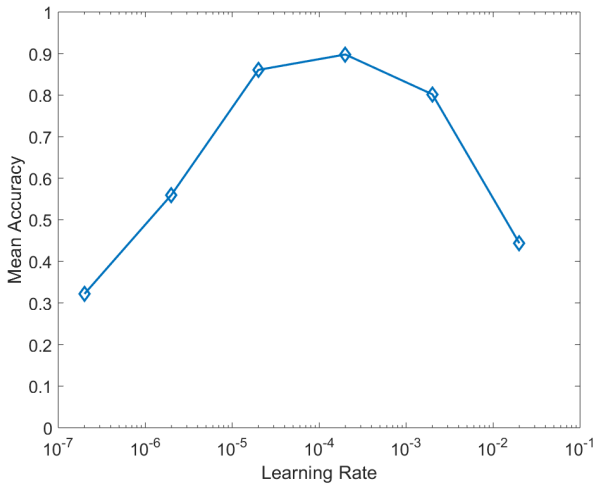

Figure 8.11: Parameter sweep of the learning rate with the best performing architecture and radar data domain: Doppler Bi-LSTM. A suboptimal initial learning rate can be as harmful to the classification accuracy as using a less suitable architecture or input domain.

8.4.4 Optimising Learning rate

A final point of note in the analysis is the selection of the learning rate. Fig 8.11 shows the sweep of the learning rate value with increments of a factor of ten for the Doppler Bi-LSTM architecture. It shows that even with an optimized architecture and input domain if the initial learning rate selected is not optimal; the classification accuracy can degrade significantly to the levels where sub-optimal networks and input domains were used. Note that the mean accuracy presented in this figure refers to the average across the 45 diverse test sequences, as discussed in previous tables.

In suboptimal cases, be it with architecture, input domain or parameter selection, the presence of delayed classification with respect to the ground truth, transient states (spikes in the predicted labels), and complete misclassifications exhibited in fig 8.8 increase significantly, thus reducing the performances.

8.4.5 Line of sight and future direction

One of the recurring questions about using radar for human activities monitoring is its performance when the target is not in a direct line of sight (LOS). In this case, the signal-to-noise ratio of the returned signal will vary, especially if the activity is being conducted at the edge of the antenna beamwidth. This is due to the combination of the effects from the attenuation of the antenna radiation pattern, the RCS fluctuation of the target, and the relation of the micro-Doppler shift with the aspect angle.
To assess this effect in a preliminary test, we recorded a sequence similar to Fig.5.6 with an added A1: walking. The participant started the activities outside of the direct LOS before walking across the LOS to perform the A6: fall event at the edge of the beam. This meant that the aspect angle between the target and the radar varied from 0 degrees to approximately 30 degrees. This sequence was then tested with the Bi-LSTM trained with prior single LOS data, as discussed in the previous sections.

Figure 8.12 shows the input, output, and ground truth of this test. The classifier makes two errors. The first is at 0 seconds where it detects A3: Standing for a small duration, and at 30 s, where it does not immediately detect the transition to A1: Walking. This is due to a non-movement gap visible in the Doppler-time map between

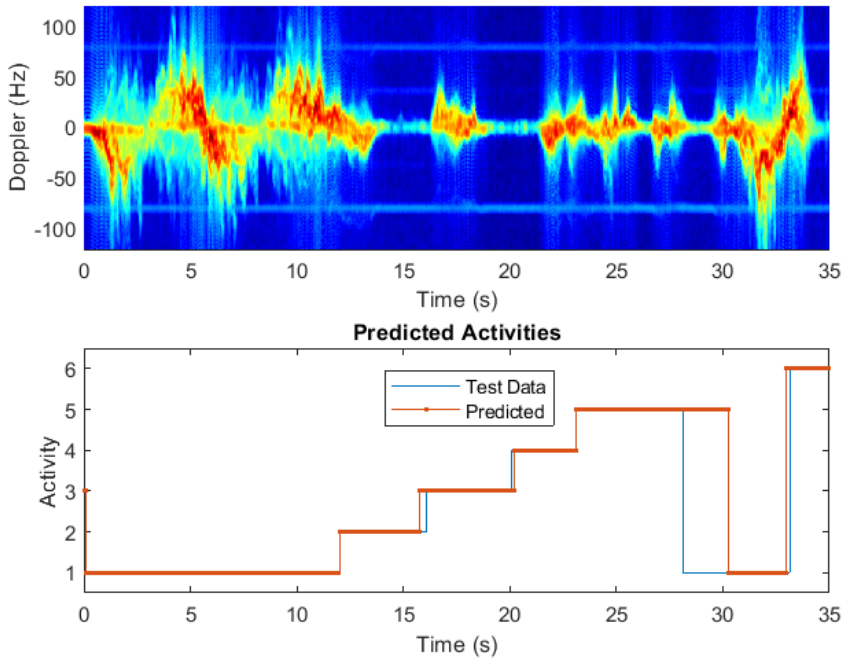

Figure 8.12: Classifier input at the top sub-figure, ground truth in blue and the test outcome in orange in the bottom sub-figure for a test sequence with varying aspect angles and signal to noise ratio for the *Doppler-LSTM* network.

28-30 seconds. Other than this, it appears to track the activities and the sequence well, with an average of 91.56% overlap between the predicted and ground truth data. This is consistent with other findings in literature where an angle up to 30 degrees gives an acceptable performance [16]. When this angle is larger (i.e, more than 30 degrees) or the target is out of the beam, performances may degrade further and a different radar deployment would necessary, such as for instance multistatic, interferometric, or multi-platform. Improving the robustness of classification for irrespective of the aspect angles is the subject of further research, but beyond the scope of this article.

8.5 Summary

This chapter analysed continuous sequences of experimental radar data to classify human activities and movements. Unlike the majority of current work in the literature, the data were not collected as separate recordings for each specific activity, but as a continuous stream where transitions between each activity can happen at any time and have unconstrained duration. Here we **demonstrated improved recognition**

of continuous activities and activity transitions without feature extraction through the use of temporal classification algorithms to show **how to leverage advances/evolution in machine learning to automatically segment and classify human activity in a continuous data stream with a single radar sensor.** These sequences were processed using novel recurrent Bidirectional LSTM networks that interpret the data as a temporal series, rather than as 2D images (i.e. matrices of pixels), as more conventional classification approaches based on convolutional networks do.

The proposed approach was validated with experimental data collected using a C-band FMCW radar with 15 participants performing 6 activities. Sequences with 3 different combinations of these activities were recorded to capture and classify diverse transitions between them. Different architectures for the recurrent networks were investigated, namely conventional LSTM and Bi-LSTM layers, as well as the effect of key hyperparameters such as the learning rate and of different formats of the input radar data, namely spectrograms and range-time sequences.

The results show that the proposed Bi-LSTM architecture outperforms unidirectional LSTM, as the former can capture connections within the data in both the backward (past) and forward (future) temporal directions. This is particularly important for the classification of continuous sequences of human movement data, as the activity/movement performed at a specific time has a strong dependence on what was performed previously and influences what the subject can perform afterward. Classification accuracy over 90% was achieved for the optimized Bi-LSTM architecture across 45 different sequences of activities tested with a leave-one-subject-out cross-validation approach, demonstrating promising robustness and generalization capabilities for the proposed approach. Micro-Doppler data yielded higher accuracy than using range-time profiles as inputs to the networks. It is anticipated that the range information can be more relevant for classification when the subjects perform activities at an unfavorable aspect angle for Doppler-based measurements, or where a radar with wider bandwidth and finer range resolution is available (for example those operating in the mm-wave spectral region). It was also shown the benefit in carefully designing the network architecture and select its hyperparameters to fit the selected radar input data domain, as optimal architectures for the micro-Doppler domain did only provide sub-optimal performances when fed with range profiles.

An open problem faced by the radar research community for human monitoring is when multiple people are in the radar field of view and the recognition of activities while subjects are occluded by other subjects or objects. Techniques to separate the signatures of multiple subjects have been proposed using the fine range information of ultra-wideband (UWB) radar [117] or the separation of the scatterers points of multiple subjects in the 3D radar cube [118]. These techniques could help separate and decompose the total signature into individual signatures that can then be subsequently processed by the proposed classification approach. The thorough investigation of this challenging scenario is left to future work, with different subjects, activities, environments, and trajectories or aspect angles.

Chapter 9

Conclusion and Future Work

Throughout the years of this research, the research topics of activity recognition with radar and radar for living targets has increased considerably in breadth and number of papers. Considering this topic only has been in focus since 2008, this represents a significant increase in the attention around the use of radar. This is also in line with the market treatment of radar as a whole, with many more applications in the civilian domain beyond the traditional defence applications most people would associate radar with.

In this thesis, we overviewed the inception of radar activity with humans and described the iterative steps through which this process occurred. Although the use of sensors to monitor and classify activities is nothing new, the mature technologies which already exist in this space come with a number of detractors which radar does not have. In this digital age, radar has a key strength of preserving privacy more than any other sensor, as it cannot be used to identify an individual or their location outside of the area under test with plain, optical images. This is especially true in the case of pervasive sensors such as camera based systems or inertial sensors which have GPS sensors, these systems have been popular recently because of their use in smartphones and other smart devices.

In the beginning of this research the literature had a number of open research questions available, which grew over the past few years as the state of the art methods evolved. The research activities described in this thesis contributed to address some parts of these questions, as summarised below.

9.1 Summary of contributions

In 2016, when this thesis' activity started, the majority of the works used the classic feature extraction methods often using a single source and a specific category of features, either textural or spectral etc. They also utilised older, more legacy classification algorithms and the experimental design was at times ambiguous and it did not consider the variety in the participants, especially older people and different genders which led to two initial queries.

The first was an inquiry about an issue which is common in many machine learning (ML) applications for activity recognition, regardless of the sensor. By demonstrating how the variety of subject physiology affects the classification rates in chapter 5 and 6 we addressed the first question of **how the target population and handcrafted feature extraction influence the accuracy of the classification algorithm.** This chapter draws from [29] and [31]. This work mirrors the trend at the time of the machine learning community regarding the importance of the input, putting it ahead of other factors such as the algorithms or hyperparameters. By evaluating the effect of variety in age, gender and physiology of the subjects on the accuracy of activity detection, it addressed an open question of presenting signatures of the type of people the hypothetical scenarios in our field have proposed.

The second inquiry was regarding the limit of the use-cases radar micro Doppler (mD) based classification had been applied to so far. By applying methods previously established we showed a **novel use of radar based gait recognition to detect lameness in animals which are physiologically different from humans.** Although this research was a sub-segment of our field, presenting these results in [30] and [24] showed a pivot in potential application of this technology, generating new interest, in this area but also with the use of mD radar. With this innovative use case we showed that mobility testing could be automated with results approaching to the level of human experts in detecting lameness or gait problems in a variety of farmed animals. We showed that activities in this case could be detected at rates between 68% to 92.4%, but also showed how individual discrete activities were more challenging to classify compared to segmented of repeated discrete activities which could be identified above 91.9%. We showed that lameness in dairy cows can be identified to 80%, up to

100% in sheep, and up to 92% in horses for a three-class problem of classifying severe, mild, and absence of lameness.

As the introduction of radar for activity recognition was an improvement on the sensing technologies, it was another avenue where new innovations were being made. By **using different sensing technologies cooperatively** and **proving that utilizing features from different domains with feature selection increases activity recognition accuracy**, we addressed **how sensing methods can be improved to optimise the classification accuracy in the context of assisted living** in chapter 7, which composes of research generated in [27, 26] and [20]. As the features generated by different properties of the input signal required selection of the highly salient ones, by using different sensors and feature selection we were able to demonstrate higher accuracy in activity recognition through exploiting the benefits of increased variance through multi-domain inputs. We demonstrated, using heterogeneous sensor fusion, where sensors were measuring different physical movements, and homogeneous sensor fusion, where sensors were measuring the same physical motions at different frequencies, we demonstrated an increase in the accuracy of activity recognition. Classification rates of up to 89.54% was achieved with the help of fusion along with SFS. With heterogeneous fusion, an accuracy 97.4% of sensors was achieved with feature fusion with 20 feaures, while for homogeneous fusion, twenty-one FMCW-derived radar features are required to achieve an accuracy of 75% where the CW radar performs to a more limited degree, with 70% accuracy when 19 features are used. With feature fusion between the radars, an improvement of classification to 83% occurs with SFS increasing this to 90% with only 15 features.

Over time the complexity of the activities and recognition of life-like sequences became a major topic of interest as people seldom do a single activity in isolation. We **demonstrated improved recognition of continuous activities and activity transitions without feature extraction** through the use of temporal classification algorithms typically used in the audio AI community. In fact, both radar and audio signals are physical movements expressed as spectral information to show **how to leverage advances/evolution in machine learning to automatically segment and classify human activity in a continuous data stream with a single**

radar sensor. We used spectrograms containing sequences of activities as inputs to an LSTM and bi-LSTM based classifier to classify the challenging signature containing not only activities of various duration but also the variable transitions between them. This piece concluded the research in this topic by presenting the best case single radar sensor performance for continuous activity recognition without specified feature generation for a realistic activity sequence with an accuracy over 90% was achieved for the optimized Bi-LSTM architecture across 45 different sequences of activities tested with a leave-one-subject-out cross-validation approach.

9.2 Impact

In the thesis we briefly discussed the challenge of obtaining data for activity recognition with radar. One of the main impact of this research is the generation of a large data-set which has already been used in multiple research papers and it is also the subject of two classification challenges. It has also been used in [119] by Addabbo et. al. as the input for a temporal convolutional network (TCN) to classify activities. This work reassuringly presents a classification scheme to classify the activities in our datasets to a high accuracy.

This work has also led to collaborations between different institutions and fields and through outreach events conducted due to this project, it has promoted the feasibility and option of using radar for activity recognition. One notable example of this being our work with the University of Glasgow veterinary school.

Finally, the research conducted for this thesis and discussed in the prior chapters have been part of broadening the scope of activity recognition with radar and has introduced various challenges present in the literature and methods to resolve them. In other words, it has expanded the knowledge base of how radar can be used for human targets for non-defence roles, establishing an idea of civilian radar which is expected to progress due to the uptake of radar for autonomous driving technologies and use in commercial smart devices.

In a pleasant turn of event, the work here [21] has been cited, amongst other researchers, by Prof. Youngwook Kim in his work [120] who was an inspiration behind this research

and one the author of the notable papers [16] which started the period of interest in this topic.

9.3 Limitations

One of the main caveats of this research is that radar has not reached maturity technologically when we consider activity recognition application. Although there are more readily available sensors that are readily available due to the autonomous driving market such as the TIDA-01570 module from Texas Instruments Incorporated and VMD3 module from RFbeam Microwave GMBH, the support of devices and usability is far inferior compared to dedicated intertial sensors such as the shimmer3 inertial measurement unit (IMU) from Shimmer.

Another limitation is the problem that radar faces in the sense of aspect angle as if the radar is oblique to the target then the signal to noise ratio increases considerably and makes activity recognition more challenging.

For participants, as the experiments are physical and have injury risks associated with them, the internal review board process to conduct these experiments safely means that many types of activities which should be identified by such a system cannot be simulated. Furthermore, it also means that obtaining data from senior citizens is challenging, even if these systems are primarily designed for their benefit.

Another general limitation and challenge is the size of the dataset and the generalisation capability to new and diverse scenarios. In fact, there can always be more activities, longer sequences, more confusers and more participants of diverse nature; however these are all increasing as more attention is attracted to activity recognition with radar.

9.4 Future Work

There were many research avenues that we could have undertaken instead of the ones discussed above and different solutions are also being offered for the research questions we have answered. Regarding the input data, recent contributions in the literature have explored the usage of generative adversarial network (GAN) in [121, 122, 123, 124]

to address the need of a large amount of data for training deep neural networks for classification. GANs have been shown to be an effective tool to generate synthetic radar data starting from a small set of experimental radar data, therefore it could be an avenue to generate data which is difficult to obtain in large volumes, such as with senior citizens. However, there are outstanding research challenges to evaluate the fidelity and reliability of synthetic data produced through them, and their closeness to the real-life data.

In [121] the authors used GANs to generate radar signatures for walking gaits at different speeds, and [122] applied GANs to data of six human actions, including a variety of full body motions. Notably, GAN counters have also been proposed in literature, with [123] proposing a novel approach to use the adversarial learning of GANs combined with a PCA-based kinematic sifting approach to reject synthetic radar samples presenting unrealistic data which contain artefacts that would normally be absent in experimental data. There is also the topic of the increasingly congested electromagnetic spectrum and approaches to use joint radar-communication waveforms in [125] to show how mD signatures of pedestrians and other automotive targets.

Another research path is the use of point clouds generated from multi-receiver radar targets as discussed in [120]. By leveraging the point information from several parts of the body there is a possibility to record richer movement information. This represents an added domain which will be used to enhance activity recognition with radar.

In short, radar is following the developments in the AI space to generate innovative solutions to the limitations presented.

Bibliography

[1] World Health Organization, "WHO Global Report on Falls Prevention in Older Age.," *Community Health*, p. 53, 2007.

[2] Nice, "Falls in older people : assessing risk and prevention," *NICE Clinical guideline*, 2013.

[3] L. Z. Rubenstein, "Falls in older people: Epidemiology, risk factors and strategies for prevention," *Age and Ageing*, vol. 35, pp. 37–41, 2006.

[4] D. Park, H. Bang, C. S. Pyo, and S. Kang, "Semantic open iot service platform technology," *2014 IEEE World Forum on Internet of Things (WF-IoT)*, pp. 85–88, March 2014.

[5] I. H. Lopez-Nava and M.-M. Angelica, "Wearable Inertial Sensors for Human Motion Analysis: A review," *IEEE Sensors Journal*, vol. 16, no. 22, pp. 1–1, 2016.

[6] M. von Keyserlingk, J. Rushen, A. de Passillé, and D. Weary, "Invited review: The welfare of dairy cattle—key concepts and the role of science," *Journal of Dairy Science*, vol. 92, no. 9, pp. 4101–4111, 2009.

[7] M. Bruijnis, H. Hogeveen, and E. Stassen, "Assessing economic consequences of foot disorders in dairy cattle using a dynamic stochastic simulation model.," *Journal of dairy science*, vol. 93 6, pp. 2419–32, 2010.

[8] S. Archer, N. Bell, and J. Huxley, "Lameness in uk dairy cows: a review of the current status," *In Practice*, vol. 32, pp. 492 – 504, 2010.

[9] J. R. Winter and L. Green, "Cost-benefit analysis of management practices for ewes lame with footrot.," *Veterinary journal*, vol. 220, pp. 1–6, 2017.

[10] G. Nieuwhof and S. Bishop, "Costs of the major endemic diseases of sheep in great britain and the potential benefits of reduction in disease impact," *Animal Science*, vol. 81, pp. 23–29, 2005.

[11] G. M. J. Taylor, "The national equine health survey," *The Veterinary Nurse*, vol. 5, pp. 176–176, 2014.

[12] F. Flower and D. Weary, "Gait assessment in dairy cattle," *Animal : an international journal of animal bioscience*, vol. 3, pp. 87–95, 2009.

[13] K. Keegan, "Objective measures of lameness evaluation," in *American College of Veterinary Surgeons Symposium*, 2012.

[14] H. Clayton and H. Schamhardt, "Techniques for gait analysis," *Equine Locomotion*, pp. 55–76, 2001.

[15] C. J. Rutten, A. Velthuis, W. Steeneveld, and H. Hogeveen, "Invited review: sensors to support health management on dairy farms.," *Journal of dairy science*, vol. 96 4, pp. 1928–1952, 2013.

[16] Y. Kim and H. Ling, "Human activity classification based on micro-doppler signatures using a support vector machine," *IEEE T. Geoscience and Remote Sensing*, vol. 47, pp. 1328–1337, 2009.

[17] A. Shrestha, C. Loukas, J. L. Kernec, F. Fioranelli, V. Busin, N. Jonsson, G. King, M. Tomlinson, L. Viora, and L. Voûte, "Animal lameness detection with radar sensing," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, pp. 1189–1193, 2018.

[18] H. Li, A. Shrestha, F. Fioranelli, J. L. Kernec, H. Heidari, M. Pepa, E. Cippitelli, E. Gambi, and S. Spinsante, "Multisensor data fusion for human activities classification and fall detection," *2017 IEEE SENSORS*, pp. 1–3, 2017.

[19] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "Magnetic and radar sensing for multimodal remote health monitoring," *IEEE Sensors Journal*, vol. 19, pp. 8979–8989, 2019.

[20] A. Shrestha, H. Li, F. Fioranelli, and J. L. Kernec, *Multimodal sensing for assisted living using radar*. Micro-Doppler Radar and Its Applications, Institution of Engineering and Technology, 2020.

[21] A. Shrestha, H. Li, J. L. Kernec, and F. Fioranelli, "Continuous human activity classification from fmcw radar with bi-lstm networks," *IEEE Sensors Journal*, vol. 20, pp. 13607–13619, 2020.

[22] H. Li, X. Liang, A. Shrestha, Y. Liu, H. Heidari, J. L. Kernec, and F. Fioranelli, "Hierarchical sensor fusion for micro-gesture recognition with pressure sensor array and radar," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 4, pp. 225–232, 2020.

[23] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "Bi-lstm network for multimodal continuous human activity recognition and fall detection," *IEEE Sensors Journal*, vol. 20, pp. 1191–1201, 2020.

[24] A. Shrestha, H. Li, F. Fioranelli, and J. L. Kernec, "Activity recognition with co-operative radar systems at c and k band," *The Journal of Engineering*, vol. 2019, pp. 7100–7104, 2019.

[25] A. Shrestha, J. L. Kernec, F. Fioranelli, Y. Lin, Q. He, J. Lorandel, and O. Romain, "Elderly care: activities of daily living classification with an s band radar," *The Journal of Engineering*, vol. 2019, pp. 7601–7606, 2019.

[26] H. Li, A. Shrestha, F. Fioranelli, J. L. Kernec, and H. Heidari, "Hierarchical classification on multimodal sensing for human activity recogintion and fall detection," *2018 IEEE SENSORS*, pp. 1–4, 2018.

[27] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "A multisensory approach for remote health monitoring of older people," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 2, no. 2, pp. 102–108, 2018.

[28] A. Shrestha, C. Murphy, I. Johnson, A. Anbulselvam, F. Fioranelli, J. L. Kernec, and S. Gurbuz, "Cross-frequency classification of indoor activities with dnn transfer learning," *2019 IEEE Radar Conference (RadarConf)*, pp. 1–6, 2019.

[29] A. Shrestha, J. L. Kernec, F. Fioranelli, E. Cippitelli, E. Gambi, and S. Spinsante, "Feature diversity for fall detection and human indoor activities classification using radar systems," in *International Conference on Radar Systems (Radar 2017)*, 2017.

[30] A. Shrestha, J. L. Kernec, F. Fioranelli, J. Marshall, and L. Voûte, "Gait analysis of horses for lameness detection with radar sensors," in *International Conference on Radar Systems (Radar 2017)*, 2017.

[31] F. Fioranelli, S. A. Shah, H. Li, A. Shrestha, S. Yang, and J. L. Kernec, "Radar sensing for healthcare," *Electronics Letters*, vol. 55, pp. 1022–1024, 2019.

[32] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote. Sens.*, vol. 11, p. 1068, 2019.

[33] H. Griffiths, P. Knott, and W. Koch, "Christian hülsmeyer: Invention and demonstration of radar, 1904," *IEEE Aerospace and Electronic Systems Magazine*, vol. 34, no. 9, pp. 56–60, 2019.

[34] H. Griffiths, "Early history of bistatic radar," *2016 European Radar Conference (EuRAD)*, pp. 253–257, 2016.

[35] P. Molchanov, *Radar Target Classification by Micro-Doppler Contributions*. Tampere University of Technology, 2014.

[36] V. Chen, *The Micro-Doppler Effect in Radar*. Artech House, 2011.

[37] T. Mitchell, *Machine Learning*. McGraw Hill, 2017.

[38] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[39] 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. April, pp. 1933–1937, 2015.

[40] X. Shi, F. Zhou, L. Liu, B. Zhao, and Z. Zhang, "Textural feature extraction based on time-frequency spectrograms of humans and vehicles," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1251–1259, 2015.

[41] J. A. Nanzer, "A review of microwave wireless techniques for human presence detection and classification," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 5, pp. 1780–1794, 2017.

[42] S. Björklund, T. Johansson, and H. Petersson, "Evaluation of a micro-doppler classification method on mm-wave data," in *2012 IEEE Radar Conference*, pp. 0934–0939, 2012.

[43] A.-K. Seifert, A. M. Zoubir, and M. G. Amin, "Radar-based human gait recognition in cane-assisted walks," in *2017 IEEE Radar Conference (RadarConf)*, pp. 1428–1433, 2017.

[44] B. W. Silverman and M. C. Jones, "E. fix and j.l. hodges (1951): An important contribution to nonparametric discriminant analysis and density estimation: Commentary on fix and hodges (1951)," *International Statistical Review / Revue Internationale de Statistique*, vol. 57, no. 3, pp. 233–238, 1989.

[45] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[46] T. Hastie, J. Friedman, and R. Tibshirani, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction Second edition.* Springer Series in Statistics, 2009.

[47] S. Zubeyde Gurbuz, *Deep Neural Network Design for Radar Applications.* Radar, Sonar & Navigation, Institution of Engineering and Technology, 2020.

[48] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735–80, 1997.

[49] K. Chaccour, R. Darazi, A. H. El Hassani, and E. Andrès, "From fall detection to fall prevention: A generic classification of fall-related systems," *IEEE Sensors Journal*, vol. 17, no. 3, pp. 812–822, 2017.

[50] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "Rt-fall: A real-time and contactless fall detection system with commodity wifi devices," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 511–526, 2017.

[51] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. D. Ho, "Radar signal processing for elderly fall detection: The future for in-home monitoring," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 71–80, 2016.

[52] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, A. Bourdoux, W. De Neve, and T. Dhaene, "Indoor person identification using a low-power fmcw radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3941–3952, 2018.

[53] 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.

[54] F. Luo, S. Poslad, and E. Bodanese, "Human activity detection and coarse localization outdoors using micro-doppler signatures," *IEEE Sensors Journal*, vol. 19, no. 18, pp. 8079–8094, 2019.

[55] X. Bai, Y. Hui, L. Wang, and F. Zhou, "Radar-based human gait recognition using dual-channel deep convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9767–9778, 2019.

[56] F. V. Schultz, R. C. Burgener, and S. King, "Measurement of the radar cross section of a man," *Proceedings of the IRE*, vol. 46, no. 2, pp. 476–481, 1958.

[57] P. Jukkala, J. Ylinen, and A. Raisanen, "Use of a millimeter wave radiometer for detecting pedestrian and bicycle traffic," in *1987 17th European Microwave Conference*, pp. 585–589, 1987.

[58] J. L. Geisheimer, W. S. Marshall, and E. Greneker, "A continuous-wave (cw) radar for gait analysis," in *Conference Record of Thirty-Fifth Asilomar Conference on Signals, Systems and Computers (Cat.No.01CH37256)*, vol. 1, pp. 834–838 vol.1, 2001.

[59] A. Lin and H. Ling, "Doppler and direction-of-arrival (ddoa) radar for multiple-mover sensing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 4, pp. 1496–1509, 2007.

[60] A. Lin and H. Ling, "Through-wall measurements of a doppler and direction-of-arrival (ddoa) radar for tracking indoor movers," in *2005 IEEE Antennas and Propagation Society International Symposium*, vol. 3B, pp. 322–325 vol. 3B, 2005.

[61] V. Chen and H. Ling, *Time-frequency transforms for radar imaging and signal analysis.* Artech House, 2002.

[62] M. Otero, "Application of a continuous wave radar for human gait recognition," in *Signal Processing, Sensor Fusion, and Target Recognition XIV* (I. Kadar, ed.), vol. 5809, pp. 538 – 548, International Society for Optics and Photonics, SPIE, 2005.

[63] 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, 2016.

[64] V. Chen, "Evaluation of bayes, ica, pca and svm methods for classification," in *RTO-MP-SET-080: Target Identification and Recognition Using RF Systems*, 2004.

[65] C. Clemente, A. W. Miller, and J. J. Soraghan, "Robust principal component analysis for micro-doppler based automatic target recognition," in *3rd IMA conference on Mathematics in Defence*, (GBR), 2013.

[66] J. J. M. de Wit, R. I. A. Harmanny, and P. Molchanov, "Radar micro-doppler feature extraction using the singular value decomposition," in *2014 International Radar Conference*, pp. 1–6, 2014.

[67] F. Wang, M. Skubic, M. Rantz, and P. E. Cuddihy, "Quantitative gait measurement with pulse-doppler radar for passive in-home gait assessment," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 9, pp. 2434–2443, 2014.

[68] S. Gurbuz, C. Clemente, A. Balleri, and J. Soraghan, "Micro-doppler-based in-home aided and unaided walking recognition with multiple radar and sonar systems," *IET Radar Sonar and Navigation*, vol. 11, no. 1, pp. 107–115, 2017.

[69] C. Clemente, L. Pallotta, A. De Maio, J. J. Soraghan, and A. Farina, "A novel algorithm for radar classification based on doppler characteristics exploiting orthogonal pseudo-zernike polynomials," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 1, pp. 417–430, 2015.

[70] H. N. Le, S. Phung, and A. Bouzerdoum, "Human gait recognition with micro-doppler radar and deep autoencoder," *2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 3347–3352, 2018.

[71] 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)*, pp. 167–170, 2017.

[72] 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.

[73] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[74] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 16–28, 2019.

[75] P. Cao, W. Xia, M. Ye, J. Zhang, and J. Zhou, "Radar-id: Human identification based on radar micro-doppler signatures using deep convolutional neural networks," *IET Radar, Sonar & Navigation*, vol. 12, 2018.

[76] 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, 2016.

[77] 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, 2020.

[78] M. S. Seyfioğlu, A. M. Özbayoğlu, and S. Z. Gürbüz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1709–1723, 2018.

[79] M. S. Seyfioglu, B. Erol, S. Z. Gurbuz, and M. G. Amin, "Dnn transfer learning from diversified micro-doppler for motion classification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 5, pp. 2164–2180, 2019.

[80] N. Twomey, T. Diethe, M. Kull, H. Song, M. Camplani, S. Hannuna, X. Fafoutis, N. Zhu, P. Woznowski, P. Flach, *et al.*, "The sphere challenge: Activity recognition with multimodal sensor data," *arXiv preprint arXiv:1603.00797*, 2016.

[81] G. Cui, M. Wang, X. Yang, and L. Kong, "Human body and limb motion recognition via stacked gated recurrent units network," *IET Radar, Sonar & Navigation*, vol. 12, 2018.

[82] M. Wang, Y. D. Zhang, and G. Cui, "Human motion recognition exploiting radar with stacked recurrent neural network," *Digital Signal Processing*, vol. 87, pp. 125–131, 2019.

[83] S. Yang, J. L. Kernec, F. Fioranelli, and O. Romain, "Human activities classification in a complex space using raw radar data," in *2019 International Radar Conference (RADAR)*, pp. 1–4, 2019.

[84] X. Li, Y. He, Y. Yang, Y. Hong, and X. Jing, "Lstm based human activity classification on radar range profile," in *2019 IEEE International Conference on Computational Electromagnetics (ICCEM)*, pp. 1–2, 2019.

[85] T. Kose, Y. Terzioglu, K. Azgin, and T. Akin, "A single-mass self-resonating closed-loop capacitive mems accelerometer," in *2016 IEEE SENSORS*, pp. 1–3, 2016.

[86] K. Bliley, D. Holmes, P. Kane, R. Foster, J. Levine, E. Daniel, and B. Gilbert, "A miniaturized low power personal motion analysis logger utilizing mems accelerometers and low power microcontroller," in *2005 3rd IEEE/EMBS Special Topic Conference on Microtechnology in Medicine and Biology*, pp. 92–93, 2005.

[87] B. Hazarika, N. Afzulpurkar, C. Punyasai, and D. Kumar Das, "Design, simulation modelling of mems based comb-drive tunneling effect gyroscope," in *2012 9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pp. 1–4, 2012.

[88] H. Dong and Y. Gao, "Comparison of compensation mechanism between an nmr gyroscope and an serf gyroscope," *IEEE Sensors Journal*, vol. 17, no. 13, pp. 4052–4055, 2017.

[89] V. Nabaei, R. Chandrawati, and H. Heidari, "Magnetic biosensors: Modelling and simulation," *Biosensors and Bioelectronics*, vol. 103, pp. 69–86, 2018.

[90] G. L. Z. Chen, "Human activity classification with neural network using radar micro-doppler and range signatures," in *IET International Radar Conference (IET IRC 2020)*, pp. 222–227(5), Institution of Engineering and Technology, 2021.

[91] Y. Chen, W. Wang, Q. Liu, Y. Sun, Z. Tang, and Z. Zhu, "Human activity classification with radar based on multi-cnn information fusion," in *IET International Radar Conference (IET IRC 2020)*, pp. 538–543(5), 2021.

[92] J. Li, X. Chen, G. Yu, X. Wu, and J. Guan, "High-precision human activity classification via radar micro-doppler signatures based on deep neural network," in *IET International Radar Conference (IET IRC 2020)*, pp. 1124–1129(5), 2021.

[93] P. Cadart, M. Merlin, G. Manfredi, J. Fix, C. Ren, I. Hinostroza, and T. Letertre, "Classification in C-band of Doppler signatures of human activities in indoor environment," in *IET International Radar Conference (IET IRC 2020)*, 2021.

[94] M. Amin, *Radar for Indoor Monitoring: Detection, Classification, and Assessment*. CRC Press, 2017.

[95] D. Tahmoush and J. Silvious, "Remote detection of humans and animals," *2009 IEEE Applied Imagery Pattern Recognition Workshop (AIPR 2009)*, pp. 1–8, 2009.

[96] Y. Kim, S. Ha, and J. Kwon, "Human detection using doppler radar based on physical characteristics of targets," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, pp. 289–293, 2015.

[97] H. Whay, A. Waterman, and A. Webster, "Associations between locomotion, claw lesions and nociceptive threshold in dairy heifers during the peri-partum period.," *Veterinary journal*, vol. 154 2, pp. 155–61, 1997.

[98] A. V. Nuffel, I. Zwertvaegher, L. Pluym, S. van Weyenberg, V. Thorup, M. Pastell, B. Sonck, and W. Saeys, "Lameness detection in dairy cows: Part 1. how to distinguish between non-lame and lame cows based on differences in locomotion or behavior," *Animals : an Open Access Journal from MDPI*, vol. 5, pp. 838 – 860, 2015.

[99] H. J. Thomas, G. Miguel-Pacheco, N. Bollard, S. Archer, N. Bell, C. Mason, O. Maxwell, J. Remnant, P. Sleeman, H. Whay, and J. Huxley, "Evaluation of treatments for claw horn lesions in dairy cows in a randomized controlled trial.," *Journal of dairy science*, vol. 98 7, pp. 4477–86, 2015.

[100] V. Busin, L. Viora, G. King, M. Tomlinson, J. LeKernec, N. Jonsson, and F. Fioranelli, "Evaluation of lameness detection using radar sensing in ruminants," *Veterinary Record*, vol. 185, no. 18, pp. 572–572, 2019.

[101] J. Raol, *Multi-sensor data fusion with MATLAB®*. CRC Press, 12 2009.

[102] E. Cippitelli, F. Fioranelli, E. Gambi, and S. Spinsante, "Radar and rgb-depth sensors for fall detection: A review," *IEEE Sensors Journal*, vol. 17, no. 12, pp. 3585–3604, 2017.

[103] C. Chen, R. Jafari, and N. Kehtarnavaz, "A real-time human action recognition system using depth and inertial sensor fusion," *IEEE Sensors Journal*, vol. 16, no. 3, pp. 773–781, 2016.

[104] R. C. King, E. Villeneuve, R. J. White, R. S. Sherratt, W. Holderbaum, and W. S. Harwin, "Application of data fusion techniques and technologies for wearable health monitoring," *Medical Engineering & Physics*, vol. 42, pp. 1–12, 2017.

[105] S. Z. Gürbüz, B. Tekeli, M. Yüksel, C. Karabacak, A. C. Gürbüz, and M. B. Guldogan, "Importance ranking of features for human micro-doppler classification with a radar network," in *Proceedings of the 16th International Conference on Information Fusion*, pp. 610–616, 2013.

[106] H. Heidari, N. Wacker, S. Roy, and R. Dahiya, "Towards bendable cmos magnetic sensors," in *2015 11th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, pp. 314–317, 2015.

[107] V. C. Chen, D. Tahmoush, and J. Miceli, eds., *Radar Micro-Doppler Signatures: Processing and Applications*. Radar, Sonar and Navigation, Institution of Engineering and Technology, 2014.

[108] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. D. Ho, "Radar signal processing for elderly fall detection: The future for in-home monitoring," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 71–80, 2016.

[109] F. Fioranelli, M. Ritchie, and H. Griffiths, "Analysis of polarimetric multistatic human micro-doppler classification of armed/unarmed personnel," in *2015 IEEE Radar Conference (RadarCon)*, pp. 0432–0437, 2015.

[110] J. Zhu, R. San-Segundo, and J. Pardo, "Feature extraction for robust physical activity recognition," *Human-centric Computing and Information Sciences*, vol. 7, 2017.

[111] C. Ding, H. Hong, Y. Zou, H. Chu, X. Zhu, F. Fioranelli, J. Le Kernec, and C. Li, "Continuous human motion recognition with a dynamic range-doppler trajectory method based on fmcw radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6821–6831, 2019.

[112] V. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Micro-doppler effect in radar: phenomenon, model, and simulation study," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 2–21, 2006.

[113] M. Schuster and K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[114] Z. Yu, V. Ramanarayanan, D. Suendermann-Oeft, X. Wang, K. Zechner, L. Chen, J. Tao, A. Ivanou, and Y. Qian, "Using bidirectional lstm recurrent neural networks to learn high-level abstractions of sequential features for automated scoring of non-native spontaneous speech," in *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 338–345, 2015.

[115] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "A multisensory approach for remote health monitoring of older people," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 2, no. 2, pp. 102–108, 2018.

[116] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "Activities recognition and fall detection in continuous data streams using radar sensor," in *2019 IEEE MTT-S International Microwave Biomedical Conference (IMBioC)*, vol. 1, pp. 1–4, 2019.

[117] T. Sakamoto, T. Sato, P. J. Aubry, and A. G. Yarovoy, "Texture-based automatic separation of echoes from distributed moving targets in uwb radar signals," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 352–361, 2015.

[118] H. Du, T. Jin, M. Li, Y. Song, and Y. Dai, "Detection of multi-people micromotions based on range–velocity–time points," *Electronics Letters*, vol. 55, no. 23, pp. 1247–1249, 2019.

[119] P. Addabbo, M. L. Bernardi, F. Biondi, M. Cimitile, C. Clemente, and D. Orlando, "Temporal convolutional neural networks for radar micro-doppler based gait recognition," *Sensors*, vol. 21, no. 2, 2021.

[120] Y. Kim, I. Alnujaim, and D. Oh, "Human activity classification based on point clouds measured by millimeter wave mimo radar with deep recurrent neural networks," *IEEE Sensors Journal*, vol. 21, no. 12, pp. 13522–13529, 2021.

[121] X. Shi, Y. Li, F. Zhou, and L. Liu, "Human activity recognition based on deep learning method," *2018 International Conference on Radar (RADAR)*, pp. 1–5, 2018.

[122] I. Alnujaim, D. Oh, and Y. Kim, "Generative adversarial networks for classification of micro-doppler signatures of human activity," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, pp. 396–400, 2020.

[123] B. Erol, S. Gurbuz, and M. Amin, "Motion classification using kinematically sifted acgan-synthesized radar micro-doppler signatures," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, pp. 3197–3213, 2020.

[124] H. G. Doherty, L. Cifola, R. Harmanny, and F. Fioranelli, "Unsupervised learning using generative adversarial networks on micro-doppler spectrograms," *2019 16th European Radar Conference (EuRAD)*, pp. 197–200, 2019.

[125] G. Duggal, S. Vishwakarma, K. Mishra, and S. S. Ram, "Doppler-resilient 802.11ad-based ultrashort range automotive joint radar-communications system," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, pp. 4035–4048, 2020.