Multimodal Radar Sensing for Ambient Assisted Living

by

Haobo Li

James Watt School of Engineering

University of Glasgow

A thesis submitted

For the degree of Doctor of Philosophy

2020

Abstract

Data acquired from health and behavioural monitoring of daily life activities can be exploited to provide real-time medical and nursing service with affordable cost and higher efficiency. A variety of sensing technologies for this purpose have been developed and presented in the literature, for instance, wearable IMU (Inertial Measurement Unit) to measure acceleration and angular speed of the person, cameras to record the images or video sequence, PIR (Pyroelectric infrared) sensor to detect the presence of the person based on Pyroelectric Effect, and radar to estimate distance and radial velocity of the person.

Each sensing technology has pros and cons, and may not be optimal for the tasks. It is possible to leverage the strength of all these sensors through information fusion in a multimodal fashion. The fusion can take place at three different levels, namely, i) signal level where commensurate data are combined, ii) feature level where feature vectors of different sensors are concatenated and iii) decision level where confidence level or prediction label of classifiers are used to generate a new output. For each level, there are different fusion algorithms, the key challenge here is mainly on choosing the best existing fusion algorithm and developing novel fusion algorithms that more suitable for the current application.

The fundamental contribution of this thesis is therefore exploring possible information fusion between radar, primarily FMCW (Frequency Modulated Continuous Wave) radar, and wearable IMU, between distributed radar sensors, and between UWB impulse radar and pressure sensor array. The objective is to sense and classify daily activities patterns, gait styles and micro-gestures as well as producing early warnings of high-risk events such as falls. Initially, only "snapshot" activities (single activity within a short X-s measurement) have been collected and analysed for verifying the accuracy improvement due to information fusion. Then continuous activities (activities that are performed one after another with random duration and transitions) have been collected to simulate the real-world case scenario. To overcome the drawbacks of conventional sliding-window approach on continuous data, a Bi-LSTM (Bidirectional Long Short-Term Memory) network is proposed to identify the transitions of daily activities. Meanwhile, a hybrid fusion framework is presented to exploit the power of soft and hard fusion. Moreover, a trilateration-based signal level fusion method has been successfully applied on the range information of three UWB (Ultra-wideband) impulse radar and the results show comparable performance as using micro-Doppler signature, at the price of much less computation loads. For classifying 'snapshot' activities, fusion between radar and wearable shows approximately 12% accuracy improvement compared to using radar only, whereas for classifying continuous activities and gaits, our proposed hybrid fusion and trilateration-based signal level improves roughly 6.8% (before 89%, after 95.8%) and 7.3% (before 85.4%, after 92.7%), respectively.

List of Publications

Journal Papers

[1] X. Liang, H. **Li**, A. Vuckovic, J. Mercer and H. Heidari, "A Neuromorphic Model with Delay-based Reservoir Computing for Ventricular Heartbeat Detection, " *IEEE Transactions on Neural Network and Learning Systems*, under review.

[2] H. **Li**, A. Mehul, J. L. Kernec, "Sequential Human Gait Classification with Distributed Radar Sensor Fusion," *IEEE Sensors Journal*, under review, July 2020.

[3] A. Shrestha, H. **Li**, J. L. Kernec and F. Fioranelli, "Continuous human activity classification from FMCW radar with Bi-LSTM networks," *IEEE Sensors Journal*, Accepted for publication, June 2020.

[4] 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*, Accepted for publication, Oct 2019.

[5] H. **Li**, X. Liang, A. Shrestha, H. Heidari, J. L Kernec, and F. Fioranelli, "Hierarchical Sensor Fusion for Micro-Gestures Recognition with Pressure Sensor Array and Radar," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 4, no. 3, pp. 225-232, Sept. 2020.

[6] X. Liang, H. **Li**, W. Wang, Y. Liu, R. Ghannam, F. Fioranelli, and H. Heidari, "Multi-Sensors Fusion with Hierarchical Support Vector Machine for Gestures Classification," *in Advanced Intelligence System (Wiley)*, 1(7), August 2019.

[7] H. **Li**, A. Shrestha, H. Heidari, J. Le Kernec and F. Fioranelli, "Magnetic and Radar Sensing for Multimodal Remote Health Monitoring," in *IEEE Sensors Journal*, vol. 19, no. 20, pp. 8979-8989, 15 Oct.15, 2019

[8] 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, June 2018.

[9] F. Fioranelli, S. A. Shah, H. **Li**, A. Shrestha, S. Yang and J. L. Kernec, "Radar sensing for healthcare," in *Electronics Letters*, vol. 55, no. 19, pp. 1022-1024, Sept. 2019.

Conference Papers

[1] H. **Li**, J. L. Kernec, A. Mehul, S. Z. Gurbuz and F. Fioranelli, "Distributed Radar Information Fusion for Gait Recognition and Fall Detection," *IEEE Radar Conf. 2020*, Florence, Italy, 21-25 Sept 2020. (Accepted for publication).

[2] 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 Proceedings of IEEE MTT-S 2019 International Microwave Biomedical Conference (IMBioC2019)*, Nanjing, China, 6-8 May 2019.

[3] F. Fioranelli, H. **Li**, J. L. Kernec. V. Busin, N. Jonsson, G. King, M. Tomlinson and L Viora, "Radar-based Evaluation of Lameness Detection in Ruminants: Preliminary Results," *in Proceedings of IEEE MTT-S 2019 International Microwave Biomedical Conference (IMBioC2019)*, Nanjing, China, 6-8 May 2019.

[4] X. Li, S. Li, H. **Li** and F. Fioranelli, "Accuracy Evaluation on the Respiration Rate Estimation using Off-the-shelf Pulse-Doppler Radar," *in Proceedings of IEEE MTT-S 2019 International Microwave Biomedical Conference (IMBioC2019)*, Nanjing, China, 6-8 May 2019.

[5] H. **Li**, A. Shrestha, F. Fioranelli, J. L. Kernec and H. Heidari, "Hierarchical Classification on Multimodal Sensing for Human Activity Recognition and Fall Detection," *in Proceedings of IEEE Sensors Conference*, New Delhi, India, 30 Oct-1 Nov 2018, pp. 1-4.

[6] H. **Li**, A. Shrestha, J. L. Kernec, F. Fioranelli and H. Heidari, "FMCW Radar and Inertial Sensing Synergy for Assisted Living," *in Proceedings of IET International Radar Conference*, Nanjing, China, 17-19 Oct 2018.

[7] A. Shrestha, H. **Li**, F. Fioranelli and J. L. Kernec, "Activity Recognition with Cooperative Radar Systems at C and K band," *in Proceedings of IET International Radar Conference*, Nanjing, China, 17-19 Oct 2018.

[8] H. **Li**, A. Shrestha, F. Fioranelli, J. L. Kernec, H. Heidari, P. Matteo, C. Enea, G. Ennio and S. Susanna, "Multisensory Data Fusion for Human Activities Classification and Fall Detection," *in Proceedings of IEEE Sensors Conference*, Glasgow, UK, 29-31 Oct 2017.

Book Chapter

[1] A. Shrestha, H. **Li**, F. Fioranelli and J. L. Kernec, "Chapter 6: Multimodal sensing for assisted living using radar," *in IET 'Micro-Doppler Radar and its Applications'*, 2019.

Declaration of Authorship

I, Haobo Li, confirm that this thesis and the work presented in it are my own achievement.

Statement of Copyright

Acknowledgements

I would like to say thank you to a lot of people who helped me during my three and half years PhD life, especially those fantastic academic colleagues, without them, this project could not be accomplished.

Firstly I would like to show great appreciate to my PhD supervisor, Dr. Francesco Fioranelli, for every effort he made in my PhD project and for all the kind suggestions and support he gave to my personal life. Then, I would like to thank the rest of my supervision team, Dr. Julien Le Kernec, who gave me invaluable advice and help, especially when I wrote my first publication, Dr. Sevgi Zubeyde Gurbuz, who provided me an exciting opportunity of working in her lab, and Dr. Hadi Heidari, who helped me establish strong connections with his PhD students for combining our research projects. Many thanks also to my friends in the Communication, Sensing and Imagining (CSI) Lab, Aman, Jarez, Paulo, Bruno, Matteo, Adnan, Siming, Xiangpeng, Jin, Bowen, Yingke, Yanmei and Amy, for the discussions and advice of my research work and certainly for their participation of my experiments (being the targets for my radar sensor). I also appreciate Mr. Ajay Mehul for his assistance when I worked in Dr. Gurbuz's lab.

Additionally, I would like to express my gratitude to University of Glasgow Mobility Scholarship and Great Britain China Educational Trust (GBCET), for funding my exchange in University of Alabama.

Last but not least, I would like to thank my parents for their financial support as well as encouragement throughout my PhD.

List of Symbols

B - Bandwidth

c - Speed of propagation of EM waves in vacuum

f_B - Beat-note frequency

f_c - Carrier frequency

f_d - Radar Doppler Shift

f_{PR} - Pulse repetition frequency

G_r - Transmitter gain

G_t - Receiver gain

k - Boltzmann constant

L - Global loss factor

N_c - Number of processed chirp

p - Discrete frequency set

p_{cad} - Discrete cadence set

P_t - Transmitted power of the radar

R - Physical distance to the target

R_r - Physical distance from receiver to the target

R_t - Physical distance from transmitter to the target

T_{PR}- Chirp duration

T_s - Noise temperature

t_0 - Time delay

$t_0{'}$ - Time delay when the target has a constant radial velocity

$t_0{''}$ - Time delay when analysing successive chirps

v_0 - Target radial velocity

Z - Number of samples per digitized chirp

ΔA - Angular resolution

ΔR - Range resolution

Δf_D - Doppler resolution

λ - Wavelength

σ - RCS of the target

ω - Doppler frequency set

ω_c - Angular carrier frequency

ω_{cad} - Cadence frequency set

ω_c - Quefrequency frequency set

τ - Pulse duration

List of Abbreviations

AAL – Ambient Assisted Living

ADC – Analogue to Digital Converter

AE – Auto-Encoder

AMR – Anisotropic Magnetoresistance

ANN – Artificial Neural Network

BF – Bayesian Fusion

BKS – Behavior-Knowledge Space

BPF – Band Pass Filter

BN – Bayesian Network

CAE – Convolutional Auto-Encoder

CART – Classification and Regression Tree

ccECG – capacitively-coupled Electrocardiography

CF – Comparator Function

CFR – Channel Frequency Response

CI – Confidence-Indicator

CI4R – Computer Intelligence for Radar

C-LSTM – Convolutional-Long Short-Term Memory

CNN – Convolutional Neural Network

CRC – Collaborative Representation Classifier

CSI – Communication, Sensing and Imaging

CTC – Connectionist Temporal Classification

CV – Cross-Validation

CVD – Cadence Velocity Diagram

CW – Continuous Wave

DAC – Digital to Analogue Converter

DAG – Directed Acyclic Graph

DBN – Deep Belief Network

DC-DCNN – Dual Channel-Deep Convolutional Neural Network

DCNN – Deep Convolutional Neural Network

DCT – Discrete Cosine Transform

DF – Decision Fusion

DMM – Depth Motion Maps

DNN – Deep Neural Network

DRDT – Dynamic Range-Doppler Trajectory

DT – Decision Tree

DS – Dempster-Shafer

DWT – Discrete Wavelet Transform

FA – False Alarm

FF – Feature Fusion

FFT – Fast Fourier Transform

FI – Fuzzy Integral

FPGA – Field-Programmable Gate Array

F1– F1-Score

FMCW – Frequency Modulated Continuous Wave

GA – Genetic Algorithm

GAR – Gait Recognition

GMM – Gaussian Mixture Models

GMR – Giant Magnetoresistance

GR – Gesture Recognition

GRU – Gated Recurrent Units

HAR– Human Activity Recognition

HD – Human Detection

HF– Hard Fusion

HGA – High Gain Amplifier

HHMM – Hierarchical Hidden Markov Model

HMM – Hidden Markov Model

IDFT – Inverse Discrete Fourier Transform

ID3 – Iterative Dichotomiser 3

IMF – Intrinsic Mode Functions

IMU – Inertial Measurement Unit

KF – Kalman Filter

KNN – K-Nearest Neighbour

LDA – Linear Discriminant Analysis

LiDAR – Light Detection and Ranging

LNA – Low Noise Amplifier

LOGP – Logarithmic Opinion Pool

LPC – Linear Predictive Coding

LPF – Low Pass Filter

LSTM – Long Short-Term Memory

L1O – Leaving One Participant Out

MAE – Mean Absolute Error

MDR – Maximum Detectable Range

MDS – Maximum Doppler Shift

MEMS – Micro-Electromechanical Systems

MLP – Multi-Layer Perceptron

MOCAP – Motion Capture

MR – Missing Rate

MSE – Mean Squared Error

MSSIM – Mean Structural Similarity

MTI – Moving Target Indication

Mul-DCNN – Multi-Channel Deep Conveolutional Neural Network

MUR – Maximum Unambiguous Range

MV – Majority Voting

NB – Naïve Bayes

NBC – Naïve Bayes Combiner

NGIMU – Next Generation Inertial Measurement Unit

NLP – Natural Language Processing

OFDM – Orthogonal Frequency-Division Multiplexing

PA – Power Amplifier

PCA – Principle Component Analysis

PIR – Pyroelectric Infrared

PPCA – Probabilistic Principal Component Analysis

PPV – Positive Predictive Value

PR – Personnel Recognition

PSA – Pressure Sensor Array

PSD – Power Spectral Density

RBF – Radial Basis Function

RC – Recall Combiner

RCS – Radar Cross Section

RF – Radio Frequency

RFBT – Random-Forest Bagging Trees

RGB – Red Green Blue

RGB-D – Red Green Blue-Depth

RMS – Root Mean Square

RNN – Recurrent Neural Network

SAE – Stacked Auto-Encoder

SBS – Sequential Backward Selection

SCG – Scaled Conjugate Gradient

SCGRNN – Segmented Convolutional Gated Recurrent Neural Network

SDA – Strong Displacement Activity

SE – Sensitivity

SF – Soft Fusion

SFS – Sequential Forward Selection

SGD – Stochastic Gradient Descent

SGRU – Stacked Gated Recurrent Unit

S-LSTM – Stacked- Long Short-Term Memory

SNN – Shallow Neural Network

SNR – Signal Noise Ratio

SOH – Sum-Of-Harmonics

SP – Specificity

SPHERE – Sensor Platform for HEalthcare in Residential Environment

Sub-KNN – Subspace-K Nearest Neighbour

SVD – Singular Value Decomposition

SVM – Support Vector Machine

TA – Translational Activity

TF – Time-Frequency

TMR – Tunnel Magnetoresistance

USRP – Universal Software Radio Peripheral

UWB – Ultra Wide-Band

VM – Voting Machine

WA – Weighted Average

WB-VNA – Wide-Band Vector Network Analyser

WER – WERnecke's

WMV – Weighted Majority Voting

XOR – Exclusive-OR

ZDA – Zero Displacement Activity

Table of Contents

List of Tables and Figures

1 Introduction

1.1 Background

Ambient Assisted Living (AAL) aims to provide low latency intelligence sensing-based products, services and systems which enable elderly people to live independently in their home, and at the same time reduce the cost of healthcare provision. Due to the increasingly aging population [1]–[3], it is more challenging to provide healthcare for managing multiple chronical conditions (multimorbidity) and provide timely assistance in case of critical events such as a fall accident [3], [4]. Besides physical injuries like hip fracture and head trauma [5], falls can trigger psychological problems of patients including loss of confidence in rehabilitation and fear of living alone [3]. Research also shows that life expectancy after a fall event is highly correlated with the time to receive medical aid [4], [5], and data from the U.S. Census Bureau confirms that patients over 65 who have waited over an hour have a higher chance of death within the next 5 years than otherwise [6]. Beyond the detection of critical events, the continued analysis of daily routines and activity patterns is also beneficial to identify possible changes and anomalies that may be related to worsening health conditions [7], [8]. For instance, changes in daily gait patterns and related metrics, such as gait asymmetry, imbalance, and slower or staggered gait have been associated with increasing fall risk and health anomalies in older people. These might go unnoticed by the subjects themselves until the symptoms are too severe to require hospitalization and acute treatment. Therefore, a reliable fall detection and health monitoring system capable of identifying different human behaviours is very useful, not only for timely emergency response, but also to enable early intervention and treatment monitoring. More broadly, the recent COVID-19 pandemic has highlighted the relevance and benefits of remote monitoring technologies to reduce the need for physical proximity to diagnose and monitor a wide range of conditions that could potentially affect human behaviours (e.g. concussion, stroke, and neuro-muscular disorders). Remote human activity and gait analysis technologies provide the opportunity to monitor the natural mobility of patients, as opposed to constrained settings typically used in hospitals or highly specialised laboratories. Moreover, less invasive technologies deployed in natural settings (e.g. private homes) can provide data more frequently and at less cost than evaluations conducted during hospital visits.

To enable this continued and personalized healthcare monitoring in home environment, different sensing technologies have been suggested in recent years, in the context of AAL. These include wearable sensor-based device ranging from IMU [9]–[11] to pressure sensor [12]–[14], image and video camera-based system [15], [16] such as Microsoft Kinect (an integration of infrared LED and depth camera which can work without illumination of light), ambient sensor-based module [17], [18] including PIR sensor and intelligence floor, and Radio Frequency (RF) sensing like radar [19]–[22], Wi-Fi [23], [24] and RFID [25].

- ***Wearable sensor-based device***: wearable sensor-based device intends to measure fine resolution data related to the movements of individual body parts (e.g. acceleration and angular speed), or the properties of the attached surface (e.g. skin temperature and humidity), or the vibrations caused by vital signals (e.g. pulse rate). All those applications require the involvement and compliance of users (the device needs to be attached to the user via a Velcro-strap or carried in jacket pockets), therefore this sensing approach is not ideal for older users that suffering from cognitive problems (easily forget to wear or charge the device). Moreover, this sensing approach is particularly invasive when measuring the skin temperature and vital signals of human body, which may create discomfort and feeling of being constrained, especially when used in sports. However, wearable sensor-based device is still very popular in AAL thanks to its small size, light weight as well as no privacy issues.
- ***Image and video-camera based system***: image and video-camera based system utilizes cameras located in the monitored area to capture the images or video frames for further recognition. Compared to wearable sensor-based device and ambient sensor-based module, this type of computer vision-based system can identify falls and other human movements in very high accuracy with the help of deep neural network-based algorithms (e.g. CNN). However, the privacy-preserving becomes more challenging as the generation of plain images and videos, especially when used in private homes. In addition, there are lots of other issues to be solved such as demand for line-of-sight as well as huge computational loads for real-time monitoring.
- ***Ambient sensor-based module***: ambient sensor-based module detects human motions in a non-intrusive fashion through the ambient information involving infrared light, vibration of floor and environment noise produced by events of interest. This kind of sensing approach allows user to be sensor-free and its risk of privacy leakage is much less than camera-based system. However, to achieve comparable performance as other sensing technologies, it needs to be deployed in large scale (e.g. 3 modules including 12 PIR sensors are installed for classifying walking direction and speed in [17]), besides that, false alarms are easily induced by other moving objects yielding the same effect. For instance, a glass bottle fall could produce similar sound or vibration as a fall accident of elderly people.
- ***RF sensing***: RF sensing has attracted considerable interest in the sensing research community thanks to its contactless capabilities (whereby the end-user does not need to wear, carry, or interact with any additional device, which can help for acceptance and compliance), and to its lack of plain images or videos to be recorded (which can help for potential issues of privacy). Wi-Fi and radar are the two most representative RF sensing approaches, whereas Wi-Fi is limited to detect 'bulk-motion' such as falls and freezing-of-gaits since the information content of returned signal is not rich. Compared to Wi-Fi, radar information can be represented in a 3D space, containing range (physical distance), time, and velocity (measured through the Doppler

> Effect), sometimes referred to as ‘radar cube’. Among these different domains of radar information, micro-Doppler is typically used, exploiting the small modulations on the received radar signal caused by ‘micro-motion’ of individual body parts (e.g. limbs and head). Hence, numerous studies in the literature have investigated the use of radar sensing for human activities classification, personnel recognition, and presence sensing, even in through-the-wall conditions.

To identify fall events among a set of collected data, a threshold-based method is often used, which considers any events yielding higher signal value than the pre-defined threshold as falling. However, this method can cause high false alarm rate since the threshold is a fixed constant rather than a learnable variable. Asides from that, it is not possible to set multiple thresholds when the system is required to distinguish several daily activities. Therefore, machine learning-based methods [7], [8], [26] are exploited to classify or cluster the collected data, depending on the supervised or unsupervised learning mode. This is significant for fall detection as a classifier can gain experience of identifying the signal patterns of fall events through training process. In addition, machine learning-based methods can be easily extended from binary problems (e.g. falling vs non-falling) to multi-class problems (e.g. walking vs sitting down vs standing up vs falling). In the recent years, neural network-based methods, in particular, CNN (Convolutional Neural Network) [27], [28] and RNN (Recurrent Neural Network) [29], [30] are showed to generally outperform conventional classifiers (e.g. SVM, KNN, DT) in terms of classification accuracy, at the price of additional training complexity.

However, it should be noticed that information from single sensor or one type of sensing technology may not be enough to characterize certain human activities (for instance radar can’t classify drinking water and using mobile phone very well due to the limitation of range resolution) and movements and the sensor signal strength could be attenuated because of external conditions such as bad weather, aspect angle, block of objects, and long distance to monitored target (for instance radar information, particularly micro-Doppler data, can be degraded during the tangentially movement of target to the radar line-of-sight or out of the antenna beam). Therefore, the use of additional radar nodes (multistatic/distributed radar) [19], [20], [31] or additional heterogeneous sensors [32]–[35] avoid any data degradation in a multimodal framework (in our case wearable sensor). This enables to exploit the complementary advantages of different sensing modalities, combine information at the most relevant level (e.g. at the signal, feature, or decision level), and capitalize on a plurality of sensors that are widely available in modern and smart living environments.

1.2 Motivation and Objectives

Motivations:

Each sensing technology has advantages and disadvantages, for instance, radar sensing could monitor human behaviours in a non-invasive, contactless way. However, due to the limitation of the radar range resolution, radar is not able to separate multiple very close objects. Therefore it is not very reliable to depend on radar only when detecting crucial events like falls and classifying micro-motions like human gestures. Oppositely, a wearable sensor can only provide fine-grained information about where it is placed (e.g. wrist, waist and ankle rather than the whole human subject, thus, it is also very hard to use a wearable sensor in recognizing human motion with different styles (e.g. drinking water with different hands). By fusing the strength of radar and wearable sensing, we believe that the classification system will be more robust in distinguishing similar activities and identifying high-risk events. Additionally, existing fusion algorithms may not be optimal for the classification tasks in terms of both accuracy and computational loads. Hence, new sensor fusion algorithms are in high demand to boost the classification performance and maintain a low computational budget at the same time. For the application, it should be noticed that continuous human motions (e.g. a sequence of activities containing walking, sitting down, standing up, drinking water and natural transitions between them) are more realistic compared to single, fixed-length human motion (e.g. 10s walking only), therefore this thesis firstly validates the proposed fusion algorithms on the single human motion classification then moves to more challenging continuous human motion tasks.

Major Objectives:

1. Design experiments and collect data from real subjects.
2. Analyse the radar and wearable data separately.
3. Combine the results of radar and wearable sensor through existing fusion algorithms.
4. Develop more robust sensor fusion algorithms.
5. Validate the fusion algorithms on ‘snapshot’ data, then move to more complicated continuous human motion.

Minor Objectives:

1. Program a data collection platform to synchronize multiple, different sensors (three Xethru impulse radars, one FMCW Ancortek radar, one pressure mat and one Microsoft Kinect) via C++ and design a GUI to save data.
2. Pre-process the data (including Time-Frequency analysis for radar and noise filtering for wearable sensors), extract the statistical features and choose the best feature combinations.

3. Use different machine learning algorithms (from conventional classifiers to deep neural networks) to train a classifier.
4. Compare the fusion result of different sensor combinations, propose a hybrid fusion framework based on soft and hard decision fusion and a signal level fusion algorithm using the trilaterated radar range information.
5. Design a dual Bi-LSTM layers network to classify continuous human motions and optimize its hyper-parameters.

1.3 Main Contributions

Idea Contributions:

- This thesis presents several fusion frameworks to leverage information of radar and wearable sensors for the purpose of increasing the capability for human activity recognition. Different combinations of sensors are validated and compared through various classification algorithms involving conventional classifiers and deep neural networks. As far as we know, we are among the first to explore the possible combination between those two different types of sensing technologies.

Algorithm Contributions:

- Specifically, a novel hybrid fusion approach is proposed to combine the advantages of soft and hard decision level fusion, which subsequently improves the performance of continuous activity recognition and fall detection.
- Furthermore, a trilateration-based signal level fusion algorithm is first implemented to combine the range information from three radar sensors at different spatial positions. Our proposed range trilateration achieves comparable performance as using micro-Doppler information, along with more than 90% saving in terms of computational cost.

Application Contributions:

- Finally, this thesis presents the analysis of continuous sequences of human activities and gaits, as opposed to a more conventional analysis of time-limited "snapshots" of spectrograms. The analysis is performed by a proposed dual Bi-LSTM network, successfully yielding a high gain in the classification accuracy with respect to the conventional sliding window methods over spectrograms.

1.4 Organization of the Thesis

Chapter 2 reviews different sensing technologies for AAL utilized in the literature and investigates the possible fusion approaches between different types of sensors. The classification methods including conventional classifier and deep model are also studied and compared in terms of system robustness and computational cost.

Chapter 3 depicts the FMCW radar signal processing in a mathematical manner and analyses the fundamental parameters of radar system along with their influence on the radar detecting ability. The information contained in different radar domains (micro-Doppler signature, CVD profile and radar cepstrum) are illustrated and discussed. Aside from that, this chapter also compares the wearable IMU signal before and after filtering.

Chapter 4 describes the handcrafted features and machine learning algorithms used in this thesis. Different levels of sensor fusion (signal, feature and decision level) and their pros and cons are also discussed. This chapter also presents a comprehensive analysis of feature selection algorithms proposed in the literature.

Chapter 5 presents the classification results of snapshot activities, the performance of using radar alone and fusion with wearable IMU sensor are compared. The impact of different feature selection methods as well as different classifiers and fusion methods are also discussed.

Chapter 6 extends Chapter 5's work to continuous activities and gaits, a Bi-LSTM network is proposed and compared with the conventional sliding window-based method, and the hybrid fusion techniques are also discussed. In the sequential gait analysis, a novel trilateration-based method is validated on the range information of three impulse radars, aiming to achieve comparable performance as micro-Doppler signatures with much less computational loads.

Chapter 7 focuses on micro-gesture recognition, the strength of UWB impulse radar and Pressure Sensor Array (PSA) are combined via the proposed hierarchical sensor fusion framework to improve the classification rate of both static and dynamic gesture.

Chapter 8 summarizes the thesis and draws a future picture which suggests possible ways to improve related research.

2 Literature Review

This chapter presents a survey of the main relevant literature in the field of human activity recognition using radar and other sensors. The survey considers sensing technologies, classification algorithms (both conventional algorithms based on supervised learning and more recent deep learning approaches), and information fusion algorithms to combine information from multiple sensors.

2.1 Review of Sensing Technologies

2.1.1 CW Radar

Continuous Wave (CW) radar, also referred as Doppler radar, is widely utilized in the field of human activity recognition [36], [37], vital signal monitoring [38] and gesture classification [39]. A CW radar system transmits single-tone radio waves and the signal is not modulated [26]. Meanwhile, the echo signals are received and processed continuously until the radar system is switched off. CW radar is capable of characterizing the human motions by the time-varying radial velocity, however, it can't produce range readings because there is no basis for the measurement of the time delay [26], [40]. CW radar can achieve comparable performance in classifying daily activities with respect to FMCW (Frequency Modulated Continuous Wave) and UWB (Ultra-wideband) impulse radar, in the meantime, the signal processing of CW radar is relatively simple.

At 2009, Y. Kim and H. Ling firstly utilizes micro-Doppler signatures of a 2.4 GHz CW radar incorporated with a linear SVM classifier to classify 7 human activities [36]. They extract six physical features from radar signature and the final accuracy reaches about 90% when all the features are jointly used. Additionally, this work draws a picture for future research in terms of continuous activity classification, radar aspect angle problem and through-wall measurements. Six years later, they upgrade the classifier from SVM to DCNN. Besides testing the new model on the previous dataset [37], the measured Doppler data of a CW radar operating at 7.25 GHz is also used to identify target type. The classification performance of HAR remains the same level as their previous work, whereas the separation of human beings and other objects yields an average accuracy of about 97.6%.

M. G. Amin's group contributes a lot in exploring different feature subspaces for human gait analysis. In [41], micro-Doppler step signatures of left and right legs are calculated on top of radar spectrograms using a 2-D cross-correlation function. The researchers extract 5 different kinds of statistical features, namely, correlation coefficient at different Doppler frequencies, MSE, MAE, MSSIM and difference of maximal Doppler shifts. In the classification stage, a logistic classifier is created to diagnose that the gait is asymmetric or not. In [21], A. Seifert investigates more features including physical features from CVD profile such as cadence frequencies and transform-based features such as SVD and PCA. Moreover, a sum-of-harmonics (SOH)-based model is proposed to estimate the fundamental frequency of the gait and the number of harmonics. The final accuracy of the combined feature set outperforms CVD-based feature set for about 7.2%. In addition, M. S. Seyfioglu and S. Z. Gurbuz extracts 13 numerical features from Doppler spectrogram and CVD profile of a 4 GHz CW signal simulated by NI-USRP module [22]. For the machine learning part, they propose a Convolutional Auto-Encoder (CAE), whose advantage is applying convolutional filtering in the manner of unsupervised auto-encoder, nearly 94% classification accuracy is achieved for separating 12 gait styles and gait-like activities.

RNN-based classifiers have drawn lots of attention thanks to their outstanding performance on the time-series data and the Doppler-time map of CW radar could also be treated as a time-dependent matrix. In [30], [42], M. Wang et al. utilizes stacked-LSTM and GRU networks to classify micro-Doppler signatures of a 25 GHz CW radar separately. The LSTM one indicates an accuracy gain of approximately 1% with respect to GRU in distinguishing 6 daily human motions.

In some special cases, there are more than one moving target in the radar sensing area. Researchers from TU Delft successfully uses measurements of a X-band Doppler radar integrated with CNN to classify 4 activities with the presence of two testing subjects [43].

2.1.2 FMCW Radar

In contrast to CW radar, Frequency Modulated Continuous Wave (FMCW) radar is able to change the frequency of the transmitted signal during the measurement period [44], thus it can also provide range information (mapping the frequency change to time).

The research of FMCW radar in HAR and GR focuses on exploring useful information from different domains such as range-time [45], [46], range-Doppler [19], [47], Doppler-time [29], [32], [48] or a combination of them [49]–[51]. Doppler-time domain, also refereed as Doppler spectrogram, is considered as the most popular area to extract features regarding human motions. In [52], P. Cao et al. utilizes micro-Doppler signatures of a 24 GHz FMCW radar with the combination of a CNN for personnel recognition, nearly 86% average classification accuracy is achieved in the case scenario of distinguishing 10 subjects. In [48], S. A. Shah uses an AlexNet to extract features from Doppler spectrograms of a 5.8 GHz FMCW radar, then a SVM classifier is trained with those transfer learning-based features to distinguish 6 daily activities.

Range-time domain, also known as range profile has gained a lot of interest recently as it takes one less processing step than Doppler spectrogram (without performing Time-Frequency analysis). Z. Zhang et al. [45] proposes a novel network architecture involving a 3-D CNN followed by LSTM and CTC on the range profile to classify both static and dynamic gestures, significant improvement is reported compared to 2-D CNN and HMM-based models.

Range-Doppler is considered as an attractive domain as it contains two kinds of radar measurements integrated with time. In 2017, B. Erol and M. G. Amin [19] extracts envelope-based features from range-Doppler maps of a 24 GHz wideband FMCW radar, KNN and SVM classifiers are jointly used while the fall detection performance of different fusion methods have been compared. In [47], C. Ding uses a novel dynamic range-Doppler trajectory (DRDT)-based method to segment the continuous activities collected by a 5.8 GHz FMCW radar into single frames, subspace-KNN is implemented to classify the segmented data.

Additionally, those information from different domains could be jointly used for gaining subsequent improvement, in [51], M. G. Amin's team extracts the features from both range-time maps and Doppler spectrograms through stacked sparse auto-encoders, in contrast to other literatures, a logistic regression classifier is adopted to predict the labels and the fall detection results are proved to be better than conventional feature extraction and PCA-based approaches. In 2019, engineers from ETH Zurich and Google Soli project [50] designs a novel network architecture using a combination of CNN and LSTM layers for micro, low-effort gesture recognition, 3-D range-Doppler-time stream, as shown in figure 2.1, is acquired from a 60 GHz FMCW radar and the system yields approximately 87% average accuracy with the 'L1O' training and testing scheme.

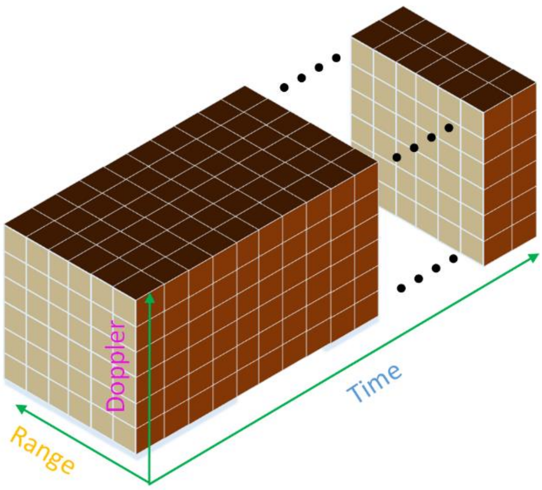


Figure 2.1 3-D radar data (also known as radar cube) reproduced from [49]

2.1.3 Ultra-WideBand (UWB) Radar

The bandwidth for Ultra-wideband (UWB) wireless sensing and communication is set as 7.5 GHz (from 3.1 to 10.6 GHz) by Federal Communications Commission (FCC) [53]. UWB radar usually transmits pulse signals with very short durations (<1ns) and the effective -10dB bandwidth of its transmitted radio wave either exceeds 25% of the operating frequency or is greater than 500 MHz [26], leading to very fine range resolution. Therefore, UWB radar is beneficial for remote respiration and heart rate monitoring [54]–[56], through-wall imaging [57], [58] and short-range activity classification [15], [59],

[60], especially in the presence of passive interference (e.g. rain, mist and metal strips) and frequency jammers. The power consumption of UWB radar is also very low [26], [60], which makes it possible to become a portable device with limited battery life, for the purpose of long-term activity recognition and vital signal monitoring in a private house or hospital. Furthermore, due to the very wide bandwidth (fine range resolution), UWB radar is capable of acquiring more information about the separate elements of one target (e.g. the chest movement in respiration detection), then identify and localize the main scattering centres among others [15], [54], [60]. Meanwhile, the bandwidth of UWB radar allows that the data transfer rate in short-range reaches a very high level. The shortcoming of UWB radar is high system complexity since it generally requires ultra-high sampling frequency at the receiver part, and same as other pulsed radar systems, it can't receive when it transmits, leading to a blind range [15], [60].

Due to the design complexity of UWB radar system, some researchers prefer to simulate UWB radar data using the positions of body joints collected by motion captured cameras (e.g. MOCAP from Graphics lab of CMU) and depth cameras such as Microsoft Kinect. In [61], Y. Lang simulates the micro-Doppler signature of a virtual UWB radar operating at 4 GHz with the help of MOCAP database, after that, a CNN is built for classifying 7 activity patterns, besides that, the effect of grayscale images, radar SNR and STFT window size have been discussed. H. Du et al. [62] also uses the MOCAP database to generate the human Doppler spectrograms of an UWB radar and then compares the classification performance of two different network architectures based on ResNet-18, one of which is fully fine-tuned by the Doppler spectrograms and the other freezes the network weight except the last dense layer (fully connected layer).

In some cases, the measured radar data is taken in a limited size and not enough for training a classifier, in particular, deep neural network. Hence, the simulated radar data could be added with the measure data to increase the size of training dataset. In [63], both simulated and measured UWB radar data are acquired, firstly, the segmented features of micro-Doppler spectrogram are extracted through the convolutional process, then those feature maps are encoded along the time bins via the gated recurrent units to recognize human activities with flexible lengths. Approximately 88% testing accuracy is reported under the 'L1O' cross-validation scheme. Alternatively, the simulated and measured UWB data could be served as training and testing set separately, also known as bottleneck analysis. Y. Lang et al. [64] trains the classification model with the simulated micro-Doppler signatures and evaluate the system performance with the measured data. Additionally, deep features extracted via an AlexNet, physical features such as Doppler centroid and bandwidth along with moment-based features are fused prior to the training of classifier.

At 2010, J. Bryan and Y. Kim [65] uses a P220 UWB radar with around 3.2 GHz effective bandwidth to distinguish 7 human motions in conjunction with SVM and PCA-based features. H. Sadreazami et al.in [59] and J. Maitre et al. in [15] both use range-time maps of one commercial UWB impulse radar sensor (Novelda Xethru X4M03 and X4M200). At 2018, H. Sadreazami proposes a recurrent neural network composed of multiple LSTM layers to detect falls and the results are significantly improved compared to conventional classifiers such as linear and Gaussian SVM. After two years, J. Maitre updates the stacked-LSTM to CNN-LSTM and evaluate its performance on a dataset comprised of 15 distinct activities.

2.1.4 Wearable Sensor

Figure 2.2 Wearable IMU senor from X-IO technologies: without strap (left); with strap (right)

Wearable sensors devices have been considered as a mature technology in the field of health and behaviour monitoring [32], [66]–[68]. Compared to contactless sensing methods based on video-cameras or RF signals, wearable devices require the users to wear or put them in their pocket during their daily lives. Wearable devices are usually utilized to map the human movements by recording the physical characteristics variation (e.g. acceleration, angular speed and magnetic field strength) related to the human body [32]. The most commonly used wearable sensor is the Inertial Measurement Unit (IMU) shown in figure 2.2, which includes an integrated chip with accelerometer, gyroscope and magnetometer [9].

The accelerometer behaviour can be assimilated to a displaced mass on a string, where upon movement it experiences a change in status and the corresponding acceleration is estimated by the displacement of the string [68]–[70]. In the commercial market, piezoresistive, piezoelectric and capacitive components are used to convert the mechanical displacement into an electric voltage. Piezoresistive materials can effectively measure sudden changes of high acceleration, whereas piezoelectric materials are sensitive to the upper frequency range and are more temperature tolerant. On the contrary, capacitor-based accelerometer are sensitive to lower frequency range [69], [70].

Gyroscope is utilized to estimate the angular speed and support maintain direction in navigation applications [71], [72]; it is typically combined with accelerometer to construct 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 tilting and rotations. Modern gyroscope sensors are based on MEMS technology [72], which allows packaging multiple gyroscopes for different axes in one chip.

Magnetic sensor or magnetometer can detect weak bio-field inside human body. It is categorized into magnetic Hall Effect sensor [73] and magnetoresistance sensors, which include anisotropic magnetoresistance (AMR), giant magnetoresistance (GMR) and tunnel magnetoresistance (TMR) [74]. Hall sensor [66], [75] is widely used into human activity recognition due to its sensitivity range, whereas magnetoresistance sensors can capture subtle variations of magnetic field (10^{-6}-10^{-12} Tesla) via an array structure [73], [74]. As the magnetometer is moved when part of the human body is performing a movement in a 3-D space, different voltages are produced from the conductor according to the amplitude of the motion and aspect angle with respect to the Earth magnetic field. This is known as Hall-effect, i.e. the external magnetic field can be related to the floating electric current on the conductor, which in turns produces a variable Hall voltage signal related to the human movement [75].

All three wearable sensors, (accelerometer, gyroscope, and magnetometer) will produce a sampled voltage signal as a function of time that is related to the movement performed by the human subject. Typically, each sensor will produce one separated signal for each axis in a 3D space (tri-axial sensors), with a total of 9 mono-dimensional raw signals to be considered as the starting point information for any activity classification analysis. Figure 2.4 below shows the tri-axial readings of one wrist-worn accelerometer (NGIMU from X-IO technology) for six different daily activities. The signal pattern in (a) is periodic as walking is the repetition of similar body movement such as swinging of limbs, whereas the sudden change of signal amplitude in (e) indicates that the subject is experiencing a fall accident. In addition, the signal returns to be flat when the subject is static.

At the first time, researcher only uses single wearable sensor to recognize human activities, in [76], S. Chembumroong et al. collects experimental data through the eZ430-Chronons sport watch, which is composed of one tri-axial accelerometer and wireless transceiver. The acquired data contains 5 daily human activities, a decision tree-based classifier (C4.5) and an ANN with one hidden layer are trained with the features selected by correlation-based method. Different amount of pruning for decision tree along with different number of hidden neurons for ANN are compared in their work to find the optimal point yielding the highest accuracy. Then, sensing via multiple wearables becomes popular due to the increasing size of dataset and gain of classification performance. C. Zhu and W. Sheng at Oklahoma State University [77] proposes a HMM-based algorithm to classify continuous data stream composed of 4 strong-displacement activities with the help of two wearable IMU sensors (one at waist and the other at ankle). The results of two sensors are fused through specific decision fusion rules and the classification accuracy is in the range of 87% to 92.5%. At 2011, they extends their previous work to 13 activities and 5 dynamic gestures [78]. Moreover, the proposed hierarchical classification system consists of two stages, where the first stage utilizes a normal ANN to divide similar activities into three sub-groups and the second stage plays a role of intra-group classification using previous HMM-based algorithm.

In the recent years, deep neural network has attracted a lot of interest in the context of wearable-based HAR. At 2016, F. J. Ordonez and D. Roggen from University of Sussex applies a Deep C-LSTM network composed of 4 convolutional, 2 LSTM and 1 fully connected layers on the public OPPORTUNITY and Skoda datasets [79]. Compared to others work, the dataset they used is comprised of 19 wearable sensors located at different joints of human body. Besides that, a sliding window-based approach is used to segment the sequence predictions and the classification result at last time step within each segmentation is considered as the final output. The proposed network outperforms the baseline CNN (regular CNN) about 2% and 8% in activity and gesture recognition, respectively. In some cases that the measured data is not enough or less sensors are involved, multiple classifier framework could be considered to leverage the strength of each distinct classifier. In [80], R. M. Gibson et al. collects acceleration data through a Shimmer IMU placed on the chest of testing subjects. Meanwhile, they combine a voting machine-based framework containing 5 different classifiers with a comparator function to detect fall events and identify the level and direction of falls.

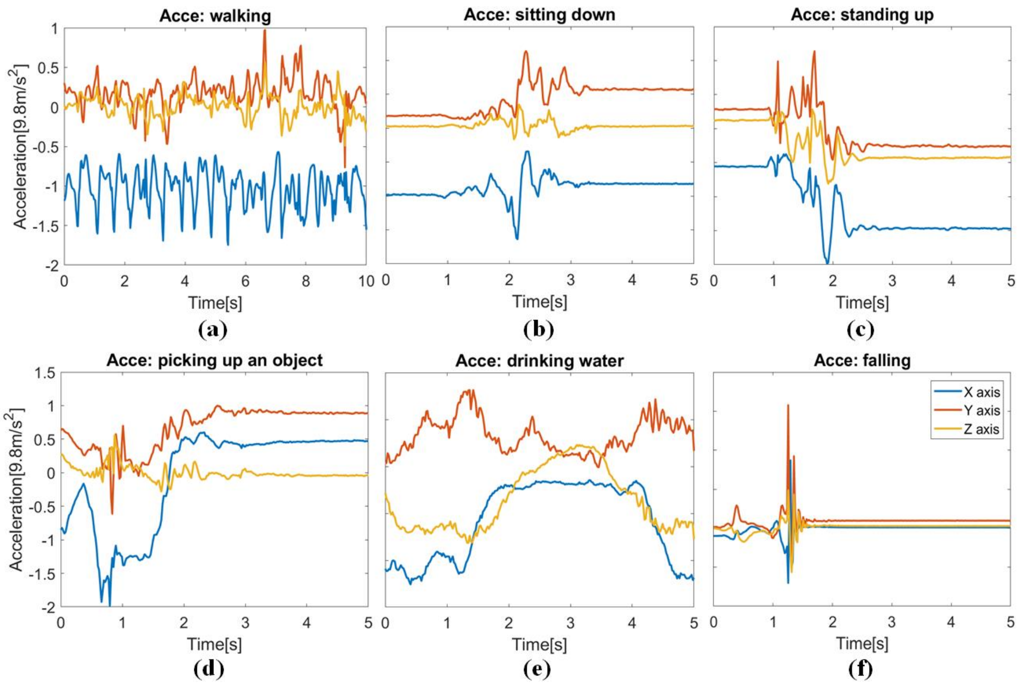


Figure 2.3 The raw acceleration for six different daily activities: (a) walking with normal speed 10s (b) sitting down on a chair and still 5s (c) standing from a chair and still 5s (d) picking up a pen and drop 5s (e) drinking water two times from a cup and put it back when finishing (f) simulating a frontal fall on a mattress

2.1.5 Other Sensing Technologies

- **LiDAR**

  Light Detection and Ranging (LiDAR) [81]–[83], is a remote sensing approach that uses a light pulse emitted by one laser source to measure the distance between sensor and the target. The difference in Time-of-Flight (light return time) could be utilized to construct a 3-D depth map of the monitoring environment [81]. Thus, LiDAR is considered as a key component of the navigation system for autonomous driving [82], [83]. Furthermore, LiDAR system could achieve higher resolution via the combination of laser diodes array and reflection mirror rotating in high speed [81]. However, adverse weather conditions like snow, rain, fog or even a very dusty measuring environment can significantly reduce the performance of LiDAR due to that the light is heavily scattered in aforementioned conditions [81]. LiDAR also can't penetrate solid objects such as a wall and its ability of detection highly relies on the reflectivity coefficient of the surface material of the target.

- **Pyroelectric infrared (PIR) sensor**

  PIR sensors detect the infrared radiation through the Pyroelectric Effect of pyroelectric materials [17], [84], it is often used to trigger the alarm for the presence of invaders in a private house. Compared to increasingly risk of privacy issues when using video camera-based surveillance systems, PIR sensors have higher security standard as no generation of plain images and videos. The outputs of PIR sensor could be digital signal, however, in terms of activity recognition, the analogue voltage output gains more interest as it is correlated with the velocity and direction of a moving target, the distance between sensor and the target and the outline of the target.

  In [17], three PIR modules, each of which contains four individual PIR sensors, are successfully used to distinguish two moving directions (forward and backward) along with three different walking distances and three different speed levels.

- **Video camera**

  Video camera is comprised of three main elements, including a lens that focuses light rays from surrounding environment, a CMOS image sensor that converts the light rays to electrical current and a recorder that digitizes the electrical signal and then encodes the video frames for further storage in camera memory [85]–[87]. Video camera-based systems are usually very reliable in activity/gesture recognition when incorporated with CNN [88], in some cases, researchers consider their results as ground truth. However, the most significant problem for visual system is the privacy invasion, especially when they are used in homes or care institutions. Until now, this has been not fully addressed.

The recorded data has a resolution up to 1920 x 1080 pixels, where each pixel consists of three colour channels (R, G and B), and for each channel, the value varies from the range of 0 to 255 [86], [88].

In [89], a video sequence-based human activity recognition system is proposed for surveillance applications, each recorded activity is represented as a combination of a set of Gaussian Mixture Models (GMM) and a Confident Frame-based method is designed to identify different human motions.

In [90], the video data is compressed to action bank and then fed into the convolutional neural network trained by effective usage of Genetic Algorithm (GA) to select the optimal initial weights. The final output is a decision fusion of multiple CNN results and around 99.98% classification accuracy is reported.

- **Wi-Fi**

Commodity Wi-Fi device is generally used to provide wireless internet access in an indoor area, however, it is reported by [23] and [24] that researchers are using physical layer-based information such as Channel Frequency Response (CFR) to identify falls and freezing-of-gaits. The definition of CFR H is given as follows [23]:

$$Y = H \times X \tag{2.1}$$

Where X and Y denote the transmitted and received radio wave, respectively. The recorded data involves CFR value for different Orthogonal Frequency-Division Multiplexing (OFDM) subcarriers and could be represented as a matrix:

$$H = [h_{sub1}, h_{sub2}, h_{sub3}, h_{sub4} \cdots, h_{subn}] \tag{2.2}$$

The phase data of CFR needs to be removed since amplitude data contains more useful information for characterizing human activities. Compared to traditional received signal strength, OFDM-based features can reveal the signal pattern changes by human activities in a more fine-grained manner.

2.2 Review of the Classification Methods for AAL

2.2.1 Conventional Classifier

The recognition of human motion through the conventional classifier typically includes first extracting the handcrafted features before training the classification model. For the time-series data (e.g. acceleration, angular speed and magnetic field strength from IMU), the handcrafted features involves the statistical parameters (e.g. mean, maximum, minimum and standard deviation) from time domain and energy-based information (e.g. FFT coefficients and spectral entropy) from frequency domain. However, for the image-based data (e.g. Doppler spectrograms from radar and images from depth or video camera), physical features related to the physical characteristic of human activity (e.g. Doppler centroid, bandwidth and envelopes), transform-based features related to the coefficients of a mathematical transform (e.g. DCT, LPC for the radar spectrogram) and DMM (Depth Motion Maps)-based features related to the projection of 3-D depth video onto three orthogonal Cartesian planes (e.g. non-zero region in each DMM) are usually considered. Furthermore, PCA (Principal Component Analysis) could be employed as an alternative way to extract features, it performs a linear transformation on the data to search the direction with the most variance. For the single-channel data, original PCA or 1-D PCA is used to compute the principle components, whereas 2-D PCA and generalized 2-D PCA are the extension of 1-D PCA on multi-channel data (e.g. image).

Conventional classifiers are widely used thanks to easy implementation and relatively low computational cost, typical conventional classifiers involves DT (Decision Tree), NB (Naive Bayes), LDA (Linear Discriminant Analysis), KNN (K Nearest Neighbour) and SVM (Support Vector Machine).

- **Decision Tree**: DT has a structure similar to tree-like flowchart, where each internal node denotes a 'test' on single feature, each branch of internal node indicates the result of the test, and each leaf node (end node) represents the final prediction label generated by considering all features. In [76], [91], DT is used as classifier to distinguish human activities and mobile app traffic, respectively. Additionally, to improve the performance of single tree, multiple DTs could be combined under a majority voting framework and the selection of training samples could be changed from fully random style to 'Bagging', in this condition, DT becomes a more powerful ensemble method 'Random-Forest Bagging Trees' (RFBT). S. Z. Gurbuz et al. test RFBT on a CW radar dataset in [8] and it yields comparable performance as linear and Radials Basis Function (RBF) kernel SVM.
- **Naive Bayes**: NB was proposed in early 1960s and has been widely utilized in document classification task. It assumes that each feature vector is statistically independent and the elements of which are in Gaussian (normal) distribution. The mean and variance of each class are used to compute the probability function in the training phase. Researchers in [20] choose

NB to identify armed and unarmed personnel, whereas in [91] NB is used both individually and in conjunction with other classifiers for the mobile app traffic classification.

- **Linear Discrimination Analysis**: similar as NB, LDA makes the decision on the assumption that the data are Gaussian distributed and the covariance matrix for all classes are equal, whereas the separation of samples with two different classes is determined by the ratio of variance between the classes and variance within the classes. References [92], [93] use LDA as an element in their proposed classifier fusion framework for the purpose of daily activity recognition.
- **K nearest neighbour**: KNN is a non-parametric classification method, which predicts the label based on the majority class of K closest training samples. The Euclidean distance between the testing sample and the training sample is often used as metric of distance. In [19], [21], [41], M. G. Amin's group evaluates the performance of KNN on the radar dataset for fall detection and gait analysis. Besides that, KNN is usually served as a 'comparator' to complex classification algorithms such as a deep neural network, whereas the strength of its variants like Subspace-KNN (Sub-KNN) and weighted KNN are significantly increased. In [47], C. Ding et al. employs Sub-KNN to classify segmented continuous radar data.
- **Support Vector Machine**: compared to aforementioned classifiers, SVM creates a hyperplane to separate the data region based on the chosen support vectors. In the case that data point can't be separated by a linear decision boundary, different kernel functions (e.g. RBF, Quadratic, and Cubic) could be jointly used with linear SVM to map the data into higher dimension where a linear separation is present. In terms of non-neural network-based classifier, SVM is one of the best choices for human motion classification and its performance can surpass neural network when the training data size is relatively small. Reference [12], [24], [36], [48], [65], [84], [94] all utilize linear SVM or SVM with kernel function to detect human targets, classify human activities and recognize micro-gestures.

2.2.2 Deep Learning Model

With the increasingly development of deep learning model, the system performance in many fields such as visual object classification and speech recognition could be improved even on a high base line. A deep model simulates the working of human brain in data processing, it usually contains multiple hidden layers to learn the high-level characterization of input data. To train a deep learning-based classifier, a large dataset is necessary, and as a consequence of that, the training process requests heavy computational budget such as large amounts of parallel computing units and storage states. However, deep learning model can save the step of exploring handcrafted features. Therefore, it has gained a lot interest in AAL, particularly, human activity recognition and gesture classification. Several deep models are studied in this thesis and they could be categorized into supervised and unsupervised learning, where supervised learning-based model involves Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and transfer learning using pre-trained nets, unsupervised learning-based model involves Bayesian Network (BN), Auto-Encoders (AE) and Hidden Markov Model (HMM). Besides that, the hybrid of multiple deep models such as Convolutional Long Short-Term Memory (C-LSTM) is also investigated.

- **Artificial Neural Network (ANN)**: ANN is a combination of connected neurons, it consists of three types of processing layers, namely, input layer, hidden layer and output layer. The dimension of the input layer depends on the input data, whereas the dimension of output layer equates to the number of classes to distinguish. Compared to the non-neural network based classifiers, all the connections between neurons are associated with a weight that updates as learning proceeds. The updating method is known as back propagation which aims to minimize the cost/loss function of the network. When used for activity recognition or similar applications, ANN has the advantage of taking both raw data and handcrafted features as inputs. In [76], ANN is used to classify human activity data collected by accelerometer, in [78] and [80], researchers exploit ANN as decision maker in coarse classification stage and part of classifier fusion framework, respectively.
- **Bayesian/Belief Network (BN)**: BN is a Directed Acyclic Graph (DAG) composed of nodes and directed connections between nodes, where the nodes denote random variables and the connections denote the conditional dependencies between random variables. The network has three different structures, including chain (head to head), fork (tail to tail) and collider (head to tail). BN models the probabilistic function between the cause and the occurrence of events, for instance, when the symptoms are given, it can compute the likelihood that one of several known diseases was the contributor. In [17], J. Yun et al. tests a set of classifiers, notably, BN, DT, NB, ANN as well as linear, quadratic and cubic SVM for distinguishing walking distance, direction and speed, where BN is reported to yield the highest accuracy. In addition, in recent years, Deep Belief Network (DBN) composed of multiple Restricted Boltzmann Machines (RBM) has attracted lots of attention in AAL as it can learn to reconstruct the input data in an unsupervised manner before fine-tuned to supervised learning. In [95], D. Wu uses DBN to process skeleton joint data for gesture recognition.
- **Convolutional Neural Network (CNN) and Transfer Learning**: CNN is first proposed by Y. LeCun et al. [96] in 1990 as a robust classification approach for handwritten digital code, which uses a hierarchical model inspired by the visual cortex structure of human brain. During training process, it intends to learn feature mapping between the input data (mostly images) and its label annotated by researchers. CNN in figure 2.4 (top) is comprised of three key elements, namely, convolution layers that utilizes a filter to process the input image within small receptive fields in a sliding window-based manner, pooling layers that fuses the outputs of several neurons at previous layer into single neuron at next layer to reduce the data size, and fully connected layer

that takes the feature analysis of convolution and pooling layers to predict the image labels. Similar as ANN, the weights of filters and the last fully connected layer are updated with the help of backpropagation method (e.g. SGD). The activation function (e.g. sigmoid function, ReLU) is usually employed on the output of convolution filters to perform a nonlinear transformation of the data, whereas dropout operation that randomly drops part of neurons could be added between those compulsory layers to prevent overfitting of the network. CNN is very powerful in terms of image and video recognition thanks to its ability of being a local feature extractor. In the field of radar image processing, CNN extracts the deep features from different spatial positions, whose content involves information related to cadence/frequency of the movement. In [27], [31], [37], [43], [61], [97], CNN is used to process various types of radar images for AAL, and in [90], [98], researchers adopt CNN to extract localized features of depth images and video data. Moreover, 3D-CNN depicted in figure 2.4 (bot) uses a cube as convolutional filter, rather than a window in 2D-CNN. In [95], [99], 3D-CNN is used to process the time-sequence image data (like a video) for gesture recognition, whereas in [45], Z. Zhang et al. combines the 3D-CNN with LSTM cells to process range maps.

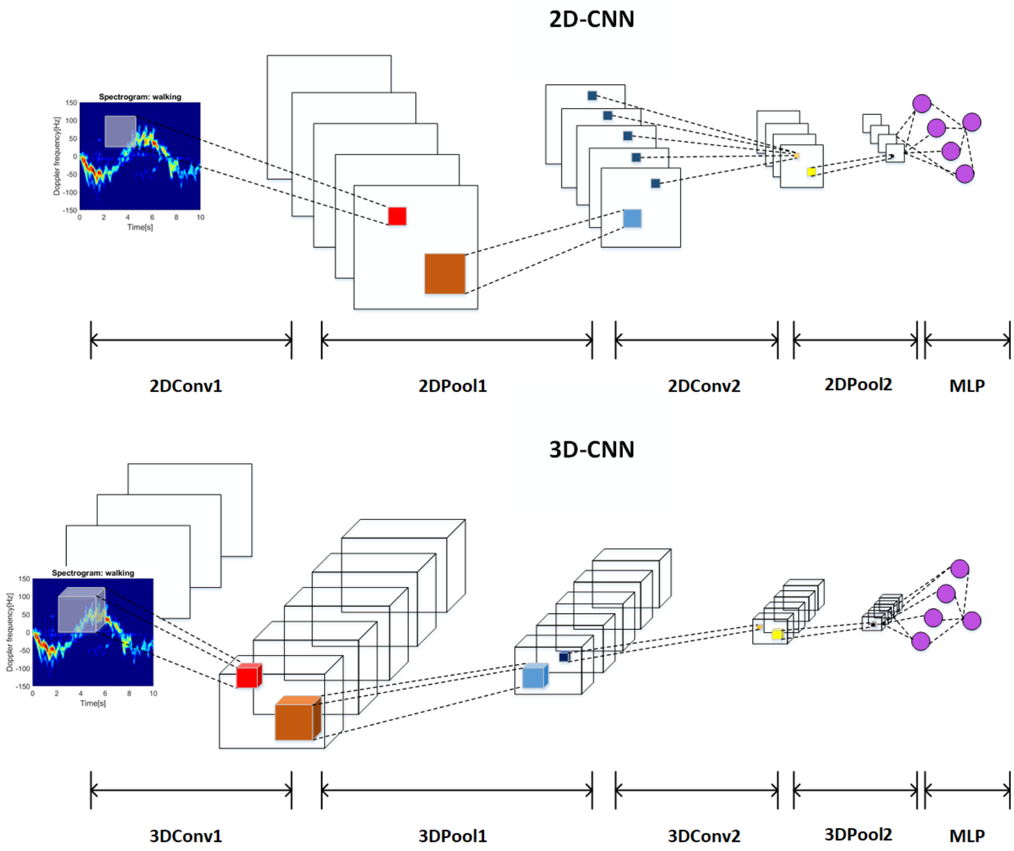


Figure 2.4 Top: regular 2D-CNN is utilized as a local feature extractor and classifier for spectrogram-based HAR. Adopted from [37]; Bot: 3D-CNN is exploited to learn high-level representations of radar range-time maps. Adopted from [45]

Pre-trained CNN could also be used to classify new, unknown images in a transfer learning fashion. It learns the features specific to the new dataset with the help of knowledge gained by previous tasks. Compared to regular CNN, this method simply fine-tunes the network original weights, leading to a significant reduction of required computational power and training time. In [62], a ResNet-18 is retrained with radar micro-Doppler signatures to classify human activities, and in [48], [64], researchers use AlexNet as a deep feature extractor and a conventional classifier to predict activity pattern.

- **Recurrent Neural Network**: with the growing interest in continuous activity recognition, RNN has gained lots of attention because of its ability to model sequence data. Differ from other types of neural networks, RNN has internal memory to remember the information from previous outputs and utilizes it to affect the output of current input. This recurrent structure allows RNN to handle input data of any length and share the weights across time. However, it can't access historical information from a long time ago (also referred as long-time dependencies). In order to solve this drawback, Long Short-Term Memory (LSTM) was proposed by researchers and yields significant improvement in NLP. LSTM cell is comprised of four key components: input gate, forget gate, cell candidate and output gate, where forget gate 'forgets' the redundant information learned by previous outputs and cell candidate learns new features from current input. By using the loop of 'forgetting' and 'relearning', LSTM can refer to the data points with long history. Another variants of RNN is Gated Recurrent Units (GRU), which is considered as a simplified version of LSTM. GRU has less parameters to compute in the training phase as it omits the output gate.

  In terms of AAL, RNN-based model has the advantage to explore the temporal relationships within continuous activity stream, which is critical to localize and identify the transition of two different activities. In [30], [59], LSTM is trained with radar range profile and micro-Doppler spectrogram, for the purpose of characterize human activities. In [42], M. Wang et al. uses GRU on the dataset in [30] and comparable performance as LSTM is reported. In [15], [45], [50], [79], [98], LSTM is combined with CNN to form a two-stage network structure, where spatial-temporal features from images or 3-D data cube (e.g. radar range-Doppler-time) are extracted by CNN and LSTM is utilized to search the global long-term dependencies among those features.

- **Auto-Encoder**: An AE is a class of feed-forward nets that attempts to reconstruct the input data at the output side under certain rules. For instance, for a given input matrix X and output matrix Y, AE aims to approximate $Y \approx X$, as depicted in figure 2.5. It is able to learn the high-level representations of unlabelled data by first encoding and then decoding the inputs. When the similarity of input data is high, AE could explore the subtle difference within the data by implementing layer-wise unsupervised pre-training. Moreover, pre-training in an unsupervised manner appears to be a regularizer, and for this perspective, it protects the neural network from overfitting. For AAL, AE has two common variants, in particular, Stack Auto-Encoder (SAE) that combines multiple sparse auto encoders to represent the data in a more compact manner, and Convolutional Auto-Encoder (CAE) that performs encoding and decoding by convolution and deconvolution. In [51], a SAE is employed to acquire the deep features of range-Doppler maps and softmax layers are added for identify the fall accidents. In [100], a network structure illustrated in figure 2.6 has been used to distinguish similar aid and unaided activities, where CNNs are utilized for the encoding phase and the decoding phase is replaced by a fully connected layer along with softmax classifier. The results show that the performance of CAE is 4% and 10% better than CNN and regular AE.

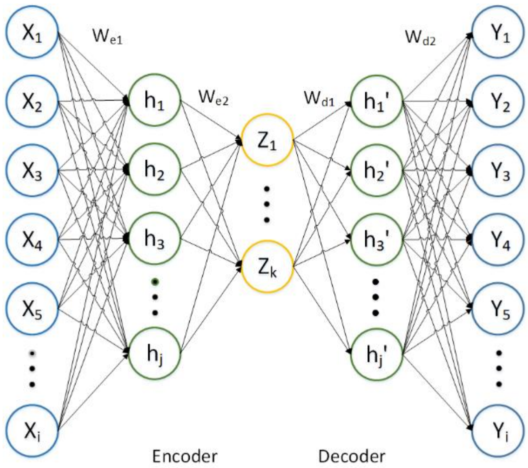


Figure 2.5 A simple schematic of AE, reproduced from [100]

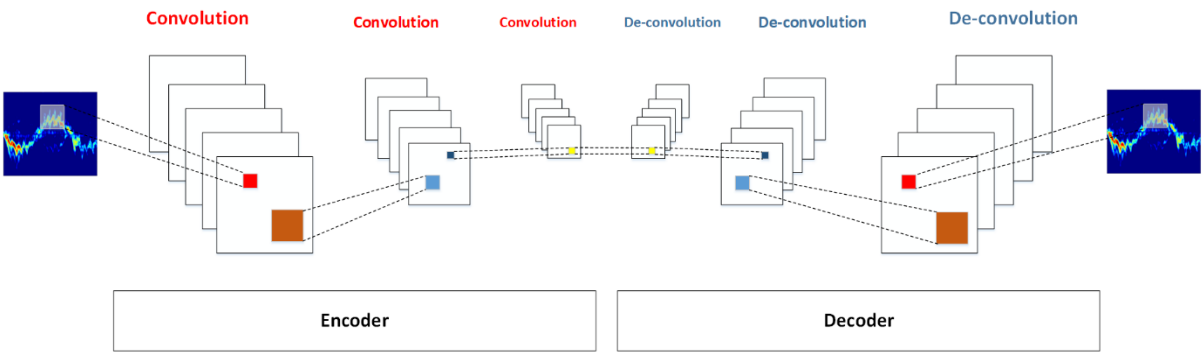


Figure 2.6 CAE is chosen as encoder and decoder to reconstruct radar spectrograms, adopted from [100]

- **Hidden Markov Model**: HMM is an augmented Markov chain, it can estimate the probability of process Y based on the current state X, however, the previous states have no influence on the process Y. Compared to original Markov chain, the events of interest in HMM cannot be observed, thus, those events are referred as hidden events (e.g. tomorrow's weather). Regarding the classifier training, forward-backward algorithm and its variant Expectation-Maximization (EM) are the most common choices to learn the two probabilities in HMM, namely, transition probabilities (between the states) and emission/output probabilities (between the state and the observation). With the successful application in the speech recognition, especially part-of-speech tagging, HMM has received lots of attention in AAL. It is able to compute the likelihood of the sequential activities for each class and the class yielding the maximal likelihood is chosen as the output label. In [95], D. Wu et al. use HMM to classify the outputs of DBN and 3D-CNN. In [77], [78], C. Zhu et al. proposes a Hierarchical HMM to distinguish continuous activity patterns.

- **Hybrid Model**: Each deep model has pros and cons, and may not be optimal for the tasks. Hybrid deep model incorporates components of different types of neural networks to gain the strength of all those networks. For instance, integrating CNN and LSTM could reinforce the ability of HAR since convolutional operations extract the localized deep features and LSTM cells find the global time correlations based on the feature analysis of CNN. In [15], [45], [50], [63], [79], [98] all utilizes such structure depicted in figure 2.7 that several convolutional and pooling layers followed by LSTM or RNN layers. However, those hybrid deep models are very complex in terms of network architecture, therefore, the training of them have to be operated on high performance computing clusters. To overcome this, the combination of deep model and conventional classifier (such as AlexNet plus KNN in [64]) could be considered, with the price of losing few classification accuracy.

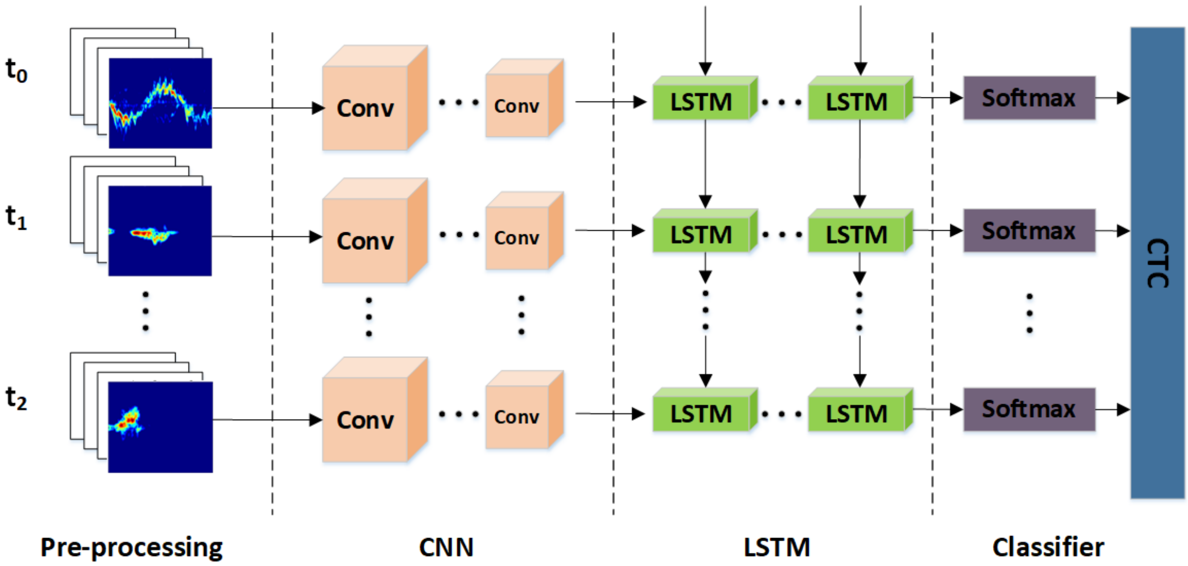


Figure 2.7 Hybrid deep model built by engineers from Google Soli, composed of CNN, LSTM and CTC layers, inspired from [50]

2.2.3 Summary of the Classification Methods

A summary of classification methods (both conventional classifiers and deep models used in AAL) is shown in table 2.1, with a comparison of classifier robustness to human motion classification and computational cost to classifier training.

Conventional classifiers	**Description**	**Robustness**	**Computational cost**
Decision Trees	a tree-like flowchart which recursively separates the input features into different classes, multiple DT could be combined to form a forest via an ensemble method	60%-70% classification accuracy	<2 mins classifier training time
Linear Discriminant Analysis	separating the data points based on the ratio of variance between the classes and variance within the classes (this method assumes that all the data are Gaussian distributed and each class has same co-variance matrix)	65%-75% classification accuracy	<2 mins classifier training time
Naïve Bayes	using mean and variance of each class to compute the probability function during the training (this method assumes each feature is statistically independent and the data are in Gaussian distribution)	60%-70% classification accuracy	<2 mins classifier training time
K Nearest Neighbor	predicting the label of testing sample based on the majority class of K closest training samples	65%-75% classification accuracy	<2 mins classifier training time
Support Vector Machine	constructing a hyper-plane to separate the data points, different kernel function could be jointly used with linear SVM to map the data into higher dimension, such as RBF-SVM and Quadratic-SVM	70%-85% classification accuracy	<10 mins classifier training time
Deep learning models	**Description**	**Robustness**	**Computational cost**
Artificial Neural Network	a feed-forward neural network that accepts both raw data and features as inputs, gradient-based method is used to update the weights between each layer	70%-85% classification accuracy	<30 mins classifier training time
Bayesian Network	a DAG composed of neurons and directed connections between neurons, often utilized to model the probabilistic function between the cause and the occurrence of events	70%-85% classification accuracy	<30 mins classifier training time
Convolutional Neural Network	capturing high-level features from different spatial positions with the help of convolutional operations in a sliding window manner, an ideal local feature extractor when used to classify image-based data	80%-95% classification accuracy	1h to few weeks classifier training time, depending on the size of dataset and depth of network
Recurrent Neural Network	has the ability of modeling sequence data, the previous outputs are stored in internal memory units and can affect the output of current input, variants are often utilized, such as GRU and LSTM	80%-95% classification accuracy	1h to few days classifier training time, depending on the size of dataset
Transfer Learning	retraining off-the-shelf CNNs such as VGG-16, AlexNet, GoogleNet with small amount of new image data (e.g. radar spectrograms), less computationally intensive than regular CNN	80%-90% classification accuracy	1h to few hours, depending on the transfer net and size of dataset

Auto-Encoders	an unsupervised neural network that intends to reconstruct the input data at output side in a encoding-decoding fashion, sensitive to highly similarity dataset	75%-90% classification accuracy	30 mins to few hours, depending on the type of AE
Hidden Markov Model	estimating the probability of process *Y* based on current state *X*, could be applied to compute the likelihood of the sequential activities for each class, very dominant in the field of wearable data analysis	75%-90% classification accuracy	30 mins to few hours, depending on the size of dataset
Hybrid Deep Models	the combination of two or more deep classification models, leveraging the strength of each model to achieve subsequent improvement, such as C-LSTM	90%-95% classification accuracy	few days to few months classifier training time, depending on the size of dataset and the complexity of network combination

Table 2.1 The summary of machine learning methods for AAL

2.3 Review of Fusion Techniques

Sensor fusion can significantly improve HAR performance compared to using single sensor. For instance, one sensor may not be able to recognize some specific events correctely due to the limiltation of sensing approach (e.g subject using the non-dominant hand without wearable sensor) or the attenuation of signal strength (e.g. subject moving vertical to radar line-of-sight). In those cases, other sensors could be used as potinetial 'helpers' to the sensor with undesirable performance. However, sensor fusion also inherits the drawbacks of different sensors when it combines the strength of them.

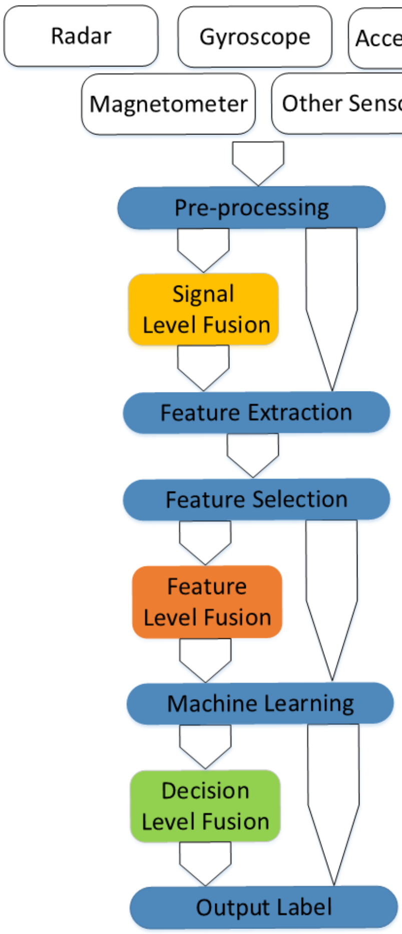


Figure 2.8 Hierarchical model of sensor fusion proposed by [101]

Fusion can take place at three different levels, notably, signal level, feature level and decision level simultaneously, the fusion methods are summarized in figure 2.8 in the format of a hierarchical model. The sources of fusion are not necessary to be different sensors, different information domains of same sensor or even different classifiers could be fused. Signal level fusion is operated on the pre-processed raw data and feature level fusion cascades the feature matrix of different sensors prior to the classification, whereas decision level fusion utilizes the outputs of distinct classifiers to generate a new output label. Additionally, decision level fusion can be implemented on two kinds of classifer outputs, namely, confidence level and classification results, referred as soft and hard decision fusion respectively.

2.3.1 Fusion of Homogeneous Sensors

At 2011, C. Kownacki exploits Kalman filter to combine the accelerazation and angular speed signals in the fashion of real-time processing [102]. Same year, fusion between multiple wearable sensors has been implemented by C. Zhu and W. Sheng in [78] through a decision fusion framework, which combines the prediction labels of two ANNs in the first stage of proposed hierarchical classification system. Besides IMU sensor, X. Liang et al. in [12] use a wrist-worn Pressure Sensor Array (PSA) wristband composed of 5 distinct capactive-based pressure sensors to classify 4 static gestures, where the feature vectors of every pressure sensor are concatenated before fed into classifier.

Information from different receivers of multistatic radar system could also be fused, in [20], F. Fioranelli. et al. exploit SVD-based features from micro-Doppler signatures of NetRad multistatic pulsed radar to identify unarmed/armed personnel and a majority voting framework is utilized to combine the prediction results of three receivers. Z. Chen et al. in [31] proposes a Mul-DCNN and verify its performance on the dataset collected by F. Fioranelli et al., the proposed network cascades the output feature maps of three receivers after 4 convolutional layers and the classification results of fusion are roughly 3% better than the best solo radar receiver. Fusion can also take place between multiple independent radar systems. B. Erol and M. G. Amin explore fusion of two 24 GHz FMCW radars at signal, feature and decision levels in [19], where averaging of pixels in two radars range-Doppler maps and DWT-based fusion are taken at signal level and soft outputs of two radars are simply added at decision level. The fall detection results show that feature fusion with DWT outperforms signal and decision fusion for about 4% and 16%, respectively.

However, fusion between different classifiers could be possible solution to increase robustness when the measurement is taken by single sensor. In [80], a Voting Machine (VM) consisting of three conventional classifiers (KNN, PPCA and LDA) and two neural netowrks (ANN and RBF) is proposed to detect the occurrence of fall along with its level and direction, whereas a Comparator Function (CF) defines the final output through XOR calculation of different VM results and some additional rules. In [49], B. Erol and M. G. Amin extend their previous work on time-intergrated range-Doppler maps to 3-D radar cube comprised of slow time, fast time and Doppler frequency. In the classificaiton stage, the soft outputs of two supevised methods, notably, LDA and SNN are fused using minimum, maximum, average and product as the aggregation rules. In [91], G. Aceto and D. Ciuonzo et al. evaluate different classifier fusion techniques including soft and hard combiners on the mobile app traffic dataset, where KL weights-based algorithm proposed by J. A. Benediktsson et al. in [103] yields the highest performance in soft fusion and Naïve Bayes Combiner (NBC) surpasses others when using hard fusion.

2.3.2 Fusion of Heterogeneous sensors

Compared to fusion between sensors yielding similar information, it is more interesting to combine heterogeneous sensors since they can observe an event in different perspectives. At 2016, N. Twomey et al. from University of Bristol propose the sensor fusion challenge for Sensor Platform for HEalthcare in Residential Environment (SPHERE) project [104], whose main task is to predict the labels of daily activities given the sensor data from wrist-worn accelerometer, RGB-D cameras and passive environmental sensors such as PIR modules.

Wearable sensor could be jointly used with depth camera, motion capture system or ventilation sensor to improve human activity classification performance. In [105] C. Chen and N. Ketarnavaz et al. implements LOGP soft fusion on the Collaborative Representation Classifier (CRC) confidence levels of a wearable IMU sensor and a depth camera, where fusion classification results of 27 activities are approximately 15.1% better than using IMU only. At 2018, N. Dawar and N. Ketarnavaz et al. from same research group uses regular CNN to classify depth images of depth camera, whereas a framework comprised of convolutional and LSTM layers is created to identify inertial signals. The combination of those two sensors is operated by multiplying the confidence levels and the classification accuracy of same dataset in [105] is improved by 1.3%. In terms of continuous activity recognition, the data sequence is segmented via a difference-based method and classification results of activity transitions achieves about 98.8%.

In [106], C. Zhu and W. Sheng et al. collect data of 5 daily activities using a wearable motion sensor (VN100) and a marker-based motion capture system (OptiTrack) containing 12 cameras, where a low-level classifier (could be a KNN, SVM, NB and ANN) is applied to predict the label of both sensors based on the extracted features. In the fusion stage, a posterior probability-based decision level fusion is proposed and the fusion results are significantly improved with the help of entropy adjustment. Ventilation sensor measures the expansion and contraction correlated with respiration rate and volume during the activity period. It is first time that S. Lin and X. Gao et al. in [107] use two accelerometers (MMA7260QT) placed at the hip and wrist in conjunction with a ventilation sensor attached at abdomen to distinguish 13 activities. Regarding the fusion, feature matrices of each sensor are combined and the results are almost 12% higher than the best solo sensor.

Similar as wearable sensor, it is possible to fuse radar data with information from other sensing sources. In [33], P. Molchanov et al. from NVIDIA Research utilize the depth data from a depth sensor to calibrate the FMCW radar range-Doppler maps and depth image encoded with velocity map of the hand is obtained for detecting gestures. In [99], they expand the previous work by using a 3D-CNN to process the gesture images from same FMCW radar, RGB camera and depth sensor, where the fusion results are increased by about 5% with respect to using radar individually. ECG measures the electrical signal caused by the heart beats and it is usually used to evaluate the heart conditions and detect some cardiovascular abnormalities. I. D. Castro et al. combines the capacitively coupled ECG (ccECG) signals with a CW radar to better estimate the heart rate. Features are extracted from 4 ccECG channels and radar backscatter signals, whereas a Confidence-Indicator (CI)-based classifier and Bayesian-based decision level fusion are proposed to increase the coverage between the predict results and ground truth. Moreover, skeleton joint information is considered as an 'Enhancer' to RGB-D images as it contains the 3-D locations of different human body parts. D. Wu et al. from IDIAP in Switzerland proposes a deep dynamic neural network to be the local feature extractor [95], which contains a Gaussian-Bernouilli Deep Belief Network (DBN) for handling skeleton information and a 3D-CNN for processing the images of RGB-D camera. Those two classifiers are combined in two different manners, notably, intermediate fusion (feature fusion before the softmax layer) and late fusion (soft fusion), whereas a HMM-based model is utilized to recognize 20 gestures based on the output posterior probabilities of first-stage classifiers. The best fusion results are approximately 3% better than using RGB-D images alone.

2.4 Summary of the Chapter

The literature reviewed in this section is summarized in table 2.2 and compared in terms of sensing approach, central frequency (radar and RF sensing only), inputs of the classifier (raw data or features), number of sensors, fusion methods, activity mode (snapshot or continuous), number of classes to distinguish, classifier and classification performance along with cross validation method used in the training and testing phases.

Ref	**Sensing approach**	**Central freq. [GHz]**	**Inputs of the classifier**	**No. of sensors (fusion methods)**	**Activity mode**	**No. of classes**	**Classifier**	**Performance /CV**
[50]	FMCW radar	60	Range-Doppler-time maps (3D)	1	Snapshot	11	CNN+LSTM	87%/L1O
[45]	FMCW radar	24	Range-time maps	1	Snapshot/Continuous	8	3-D CNN+LSTM	96%/Holdout 89.3%/L1O
[51]	FMCW radar	25	Doppler-time and range-time maps	1	Snapshot	2	SAE	97.1%/Holdout 87%/L1O
[52]	FMCW radar	24	Doppler-time maps	1	Snapshot	4, 6, 8, 10, 12, 16, 20	CNN	>85% when no. of classes<10/ Holdout
[47]	FMCW radar	5.8	Features from range-Doppler maps	1(MV)	Snapshot/Continuous*	6 for snapshot, 7 for continuous*	Sub-KNN	94.2% for snapshot; 91.9% for continuous*/ K-fold
[93]	FMCW radar and simulated radar data (Kinect)	25	Features from radar data cube	1(DF)	Snapshot	5	LDA+SNN	93.6% /Bottleneck analysis 97.2%/fine-tuning
[48]	FMCW radar	5.8	Doppler-time maps	1	Snapshot	6	AlexNet+SVM	78.25%/training and testing on data from

								different environments
[19]	FMCW radar	24	Features from range-Doppler maps	2 (WA,DWT, FF, DF)	Snapshot	2	KNN, SVM	66.84% for solo radar 86.99% for best fusion
[42]	CW radar	25	Doppler-time maps	1	Snapshot	6	SGRU	91.8%/K-fold
[30]	CW radar	25	Doppler-time maps	1	Snapshot	6	SLSTM	92.7%/K-fold
[21], [41]	CW radar	24	Features from Doppler-time maps and CVD	1	Snapshot	5	KNN	93.8%/K-fold 80.4%/L1O
[43]	CW radar and simulated CW data	X-band	Doppler-time maps	1	Snapshot	4	DCNN	96.1% for model data; 86.9% for real data / Holdout
[31]	CW radar	24	Doppler-time maps	1	Snapshot	8	DCNN	96.9%/Holdout
[100]	CW radar	4	Doppler-time maps	1	Snapshot	12	CAE	94.2%/K-fold
[37]	Doppler radar	2.4	Doppler-time maps	1	Snapshot	4 for HD;7 for HAR	DCNN	97.6% for HD; 90.3% for HAR/K-fold
[27]	Doppler radar	5.8	Doppler-time maps	1	Snapshot	7	DCNN	87.1%/N/A
[36]	Doppler radar	2.4	Features from Doppler-time maps	1	Snapshot	7	SVM	92.8%/K-fold
[61]	Simulated UWB radar data (MOCAP)	4	Doppler-time maps	1	Snapshot	7	CNN	98.34%/Holdout
[108]	UWB Doppler radar	7.25	Doppler-time maps	1	Snapshot	5	VGG-16	80.3%/K-fold
[62]	Simulated UWB radar data (MOCAP)	4	Doppler-time maps	1	Snapshot	6	ResNet-18	97.9%/Holdout
[64]	UWB radar and simulated UWB radar data (MOCAP)	4.3	Doppler-time maps	1(FF)	Snapshot	7	AlexNet+KNN	49.7%/ Bottleneck analysis
[65]	UWB radar	4.3	Features from Doppler-time maps	1	Snapshot	7	SVM	89.88%/Holdout
[63]	UWB radar and simulated UWB radar data (MOCAP)	4	Doppler-time maps	1	Snapshot/Continuous*	6	SCGRNN	88.19%/L1O
[59]	UWB impulse radar	7.3	Range-time maps	1	Snapshot	2	S-LSTM	89.8%/K-fold
[15]	UWB impulse radar	7.3	Range-time maps	1(MV)	Snapshot	15	CNN+LSTM	95%/ Holdout 73%~97%/ L1O
[97]	Simulated radar data (MOCAP)	10	Features from Doppler-time maps	1(FF)	Snapshot	3	DC-DCNN	96.24%/K-fold
[109]	Simulated radar data (Kinect)	15	Features form Doppler-time maps	1	Snapshot	4	KNN	>90%/Holdout
[31]	Multistatic pulsed radar	2.4	Doppler-time maps	3(FF)	Snapshot	2 for PR; 2 for GAR	DCNN	97.42% for PR; 99.63% for GAR /Holdout
[20]	Multistatic pulsed radar	2.4	Features from Doppler-time maps	3(MV)	Snapshot	2	NB	95.75% for best solo 97.22% for fusion/Holdout

[94]	WB-VNA	3.5	Features from spectrum of range-time maps	1	Snapshot	2	RBF-SVM	90.7%/K-fold
[79]	Wearable IMU sensor	N/A	Raw acceleration, angular speed and magnetic field strength	19(FF)	Continuous	4 for HAR; 17 for GR	C-LSTM	93% for HAR(**F1**); 86.6% for GR(**F1**)/Hold out
[110]	Wearable PSA	N/A	Raw wrist pressure level	5(FF)	Snapshot	3	SVM	>90%/Holdou t
[78]	Wearable IMU sensor	N/A	Features from raw acceleration and angular speed	1 for GR, 2 for HAR (DF)	Continuous	9 for ZDA and TA, 4 for SDA; 5 for GR	ANN+ HHMM	98.6% for ZDA and TA; 89.1% for SDA; 92% for GR/Holdout
[77]	Wearable IMU sensor	N/A	Features from raw acceleration and angular speed	2 (DF)	Continuous	4	HMM	87%~92.5%/ N/A
[76]	Acceleromete r	N/A	Features from raw acceleration	1	Snapshot	5	DT, ANN	94.1% for DT; 90.5% for ANN/K-fold
[92]	Acceleromete r	N/A	Features from raw acceleration	1(VM+CF)	Snapshot	5 for HAR; 6 for falls	ANN,KNN, RBF, PPCA and LDA	97.5% for HAR 98.3% for fall direction 93.85% for fall lvl/
[23]	Wi-Fi	5	Features from CSI	1	Continuous	2	RBF-SVM	93.3%(**SE**); 89.3%(**SP**)/L 1O
[24]	RF signal	2.4	Features from CFR	1	Snapshot	2	SVM	99%/Holdout
[17]	PIR sensor	N/A	Features from analogue output	12(FF)	Snapshot	2 for direction; 3 for distance; 3 for speed lvl	BN	>92% for direction and speed lvl; >94% for distance/K-fold
[84]	PIR sensor	N/A	Features from analogue output	12(FF)	Snapshot	6	KNN,SVM	74.38% for KNN; 97.71% for SVM/N/A
[89]	Video camera	N/A	GMM from video frames	1	Snapshot	5	Confident Frame-based method	0.49~32.4%(**MR**); 0.24~2.26%(**FA**)/Holdout
[90]	Video camera	N/A	Action bank from video frames	1	Snapshot	50	CNN+GA	99.98%/K-fold
[91]	Mobile phone	N/A	Service bursts	1(SF, HF)	Snapshot	49	NB. Multinomia l NB, RF, SVM, DT	79.2% for SF;75% for HF/Holdout
[99]	FMCW, color camera, depth camera	24	Range-Doppler map, RGB images and depth images	3(FF)	Snapshot	10	3-D CNN	89.1% for radar only 94.1% for fusion/L1O
[105]	Wearable IMU sensor, depth camera	N/A	Features from raw acceleration, angular speed and depth images	3(SF)	Snapshot	27	CRC	76.4% for IMU, 91.5% for fusion /L1O
[98]	Wearable IMU sensor, depth camera	N/A	Features from raw acceleration, angular speed and	2(DF)	Continuous	7 for HAR; 5 for GR	CNN, CNN+LST M	98.8% for HAR;//L1O; 97.6% for GR/subject-specific

			depth images					
[111]	ccECG, CW radar	5.8	Features from ccECG signals and baseband IQ data	2(BF)	Continuous	N/A	Confidence Indicators	52.2% for ccECG (**Coverage**) 63.2% for fusion(**Coverage**)/N/A
[33]	FMCW radar, depth camera	25	Range-Doppler maps, 3-D postions	4(calibrate radar with depth camera)	Snapshot	10	N/A	N/A
[107]	Accelerometer, ventilation sensor	N/A	Features from raw acceleration and respiration signal	3(FF)	Snapshot	13	SVM	77.6% for best solo; 88.1% for fusion/L1O
[95]	RGB-D camera	N/A	Skeleton information, RGB and depth images	2(FF, SF)	Snapshot	20	DBN+3DCNN+HMM	83.6% for best solo; 86.4% for fusion/Holdout
[106]	Wearable motion sensor and RGB camera	N/A	Features from raw acceleration and 2-D positions	13(BF)	Snapshot	5	Lower level classifier and clustering	67.2%~89.1% /N/A

Table 2.2 Summary of previous work

Through the literature survey, the following main points and observations can be drawn.

- Fusion between multiple sensing sources (e.g. multistatic radar, radar and video camera, wearable sensor and depth camera, multiple wearables) has been reported in others' previous work, whereas the combination of radar and wearable has not been explored so far.
- Most of the papers evaluate their classification methods on snapshot data (only involves single human motion), few papers extend the snapshot data to simplified/short continuous data (only involves one pair of human motions such as 'from walking to sitting on a chair'). However, very limited literatures have completed a comprehensive analysis on long, unsegmented data stream that containing various daily motions in different orders and natural transition between two different motions.
- In most of the literatures, the continuous data sequence needs to be first segmented into smaller frames via a sliding window before sent to the classifier. Only a small amount of researchers prefer to train a classification model with unsegmented data sequence.
- In terms of classification performance, deep model that combines multiple, different types of processing layers outperforms regular DNN, however, the network structure of those hybrid deep models are very complex and the computational cost are extremely high. Therefore, it is necessary to design a 'light' deep model with comparable performance. There are two possible solutions for further exploring: increase the robustness of regular DNN by adding more learning rules such as Bi-directional LSTM, and incorporating deep models with conventional classifiers such as CNN+SVM.
- Not too many researchers try to combine radar systems operating in different modes (e.g. FMCW and UWB impulse radar, FMCW and CW radar) or simply at different frequency bands.

- Only few papers discuss about signal level fusion of radar data and they mainly focus on the simple averaging and Discrete Wavelet Transform (DWT).
- Decision level fusion, especially hard probability combiner (e.g. Recall and Naïve Bayes Combiners) has been marginally discovered in the field of human activity recognition and gesture classification.
- In the table 2.2, all the sensor fusion methods are utilized independently, there is no proposal regarding a hierarchical framework that incorporates different types of fusion methods.
- 'Leaving one participant out' (L1O) cross validation method has been very popular across recent published works as it can simulate the real-world case scenario that the classifier can't access the data of testing subject in the training process.

3 Sensor Signal Processing

This chapter presents the pre-processing of raw FMCW radar and wearable IMU sensor data. The radar part involves the mathematical modelling of transmitted and received chirp signals, extraction of range and Doppler information, and generation of micro-Doppler signatures, CVD and cepstrum. The wearable part discusses the removal of signal noise using BPF.

3.1 FMCW Radar

In this section, chirp signals of FMCW radar system are analysed in a mathematical fashion and further discussion of three main radar information domains (range-time, range-Doppler and Doppler-time) is presented.

The transmitted signal of FMCW radar could be modulated in different patterns [40], [112], for instance, figure 3.1 illustrates the classic sawtooth modulation pattern, and it is observed that the signal frequency is varied by time in a linear fashion (also known as chirp signal). Besides that, triangular, square-wave and sinusoidal modulation patterns have been used for different measurement purposes. In the signal processing period, the backscattered signal is mixed with the transmitted signal to generate beat-notes [26], [40]. Since the frequency difference is proportional to the distance between radar and the target, the range information could be easily extracted by demodulating the beat-note and computing the amount of frequency change [26]. However, the resolution of range measurement, referring to the minimum separable distance of two very close targets, is linked with the radar bandwidth, as depicted in equation 3.1 [26].

$$\Delta R = c / 2B \tag{3.1}$$

Where ΔR refers to the range resolution, c is the speed of light and B denotes the modulation bandwidth of FMCW radar, Hence, FMCW radar is unable to distinguish very close targets without enough bandwidth. For the Doppler information part, FMCW radar operates in the same way of CW radar. FMCW radar is one of the most popular choices in the context of short-range sensing, as it can provide both range and Doppler information.

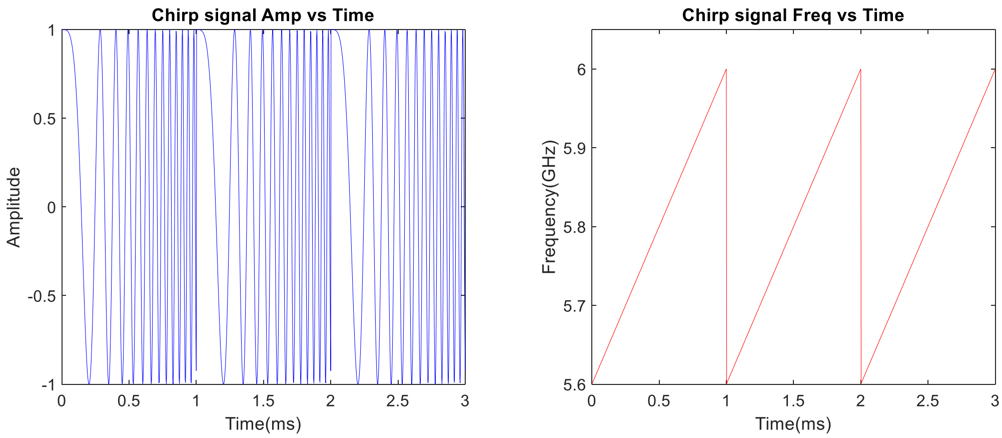


Figure 3.1 FMCW radar chirp signal in time-amplitude (left) and time-frequency (right) domains

The velocity estimation of FMCW radar is same as CW. When the target is moving, there is a frequency shift between the transmitted and received signal due to the Doppler Effect. The frequency of received echo signal is given as follows [26]:

$$f_r = \frac{1 + v_0 / c}{1 - v_0 / c} f_c \tag{3.2}$$

where f_c is the frequency of the transmitted wave, c denotes the speed of light and v_0 is the radial velocity of the moving target. The Doppler shift f_d is then obtained [26]:

$$f_d = f_r - f_c = 2v_0 \frac{f_c}{c - v_0} \tag{3.3}$$

Since v is much smaller than speed of light, the equation 3.3 could be rewritten as below [26]:

$$f_d = 2v_0 \frac{f_c}{c} \tag{3.4}$$

Thus, the radial velocity of the target could be estimated.

Figure 3.2 shows an easy implementation of FMCW radar, the transmitter part is comprised of a DAC for converting a digital control order into analogue voltage, an oscillator for generating specific radio frequency based on the input control voltage, a Band Pass Filter (BPF) for removing the harmonics and spikes of the generated electronic signal, a -3dB power divider (e.g. Wilkinson power divider) for leaving the receiver a reference frequency, a Power Amplifier (PA) for amplifying the transmitted signal to the required transmit power, and a transmitting antenna/antenna array to emit the signal into the free space, whereas the receiver part contains a Low Noise Amplifier (LNA) for magnifying the very weak received signal, a mixer for multiplication of the received echo signal with the transmitted signal, a High Gain Amplifier (HGA) for magnifying the mixed signal again, and an ADC for converting the analogue signal back to a digital signal. Compared to pulsed radar system using high peak pulse signal, FMCW radar transmits its wave continuously with equal power. Hence it avoids the compromise between signal power and pulse length (ideally, pulse radar should transmit extremely high power within a short pulse, so that the range resolution and detectable range could be both maximized, however, it is not physically possible to generate this pulse yet).

Both FMCW and pulsed radar system have three different configurations, namely, monostatic where the transmitter and receiver share a common antenna, bistatic where the transmitter and receiver are placed at two different locations over a considerable distance, and multistatic that involves more than one transmitter or receiver pairs, with all the antennas located separately.

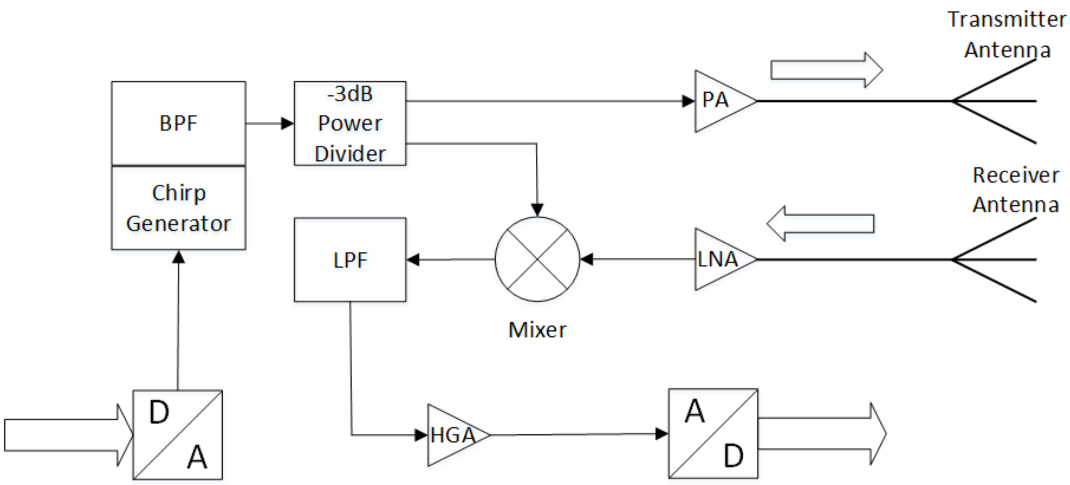


Figure 3.2 FMCW radar architecture

3.1.1 Transmitted and Received Chirp Signals

Transmitted signals could be modulated in a linear or nonlinear modulation, however, we only discuss the linear fashion since it has been applied on the radar products mentioned in this thesis (Ancortek 580B and 2500B). Chirp signal [44], as shown in figure 3.1, is composed of a linear frequency sweep within a short duration. Ideally, a single chirp, whose duration is T_{PR}, could be expressed as a function of carrier frequency f_c and effective bandwidth swept by the chirp B, as shown in equation 3.5 [44]:

$$v_T(t) = A_T \cos[\phi(t)] = A_T \cos\left(\omega_c t + \pi B f_{PR} t^2\right), \omega_c = 2\pi f_c;\ f_{PR} = \frac{1}{T_{PR}}; -\frac{T_{PR}}{2} \le t \le \frac{T_{PR}}{2} \qquad (3.5)$$

where A_T denotes the signal amplitude (constant, related to signal power), ω_c denotes the angular frequency, f_{PR} denotes the reciprocal of the chirp duration T_{PR}. Thus, Bf_{PR} denotes the sweep rate. Assuming the presence of one stationary target, the received echo signal should be same with the transmitted one except a time delay t_0:

$$v_R(t) = A_R \cos[\phi(t-t_0)] = A_R \cos\left(\omega_c(t-t_0) + \pi B f_{PR}(t-t_0)^2\right) \tag{3.6}$$

where A_R indicates the amplitude of the echo signal, proportional to the radar RCS of the target. The transmitted and received signals are multiplied at the receiver part with the help of a mixer. Hence, the mixer output could be written as product of two cosine functions, as depicted in equation 3.7:

$$v_{MF}(t) = A_T A_R \cos\left(\omega_c t + \pi B f_{PR} t^2\right)\cos\left(\omega_c(t-t_0) + \pi B f_{PR}(t-t_0)^2\right) \tag{3.7}$$

By using trigonometric identities, equation 3.7 could be turned into equation 3.8, whose second term yields frequency of twice ω_c, thus, this part is removed through the LPF placed after the mixing stage.

$$\begin{aligned} v_{MF}(t) &= \frac{A_T A_R}{2}\cos\left(\pi B f_{PR} t^2 + \omega_c \tau - \pi B f_{PR}(t-t_0)^2\right) \\ &+ \frac{A_T A_R}{2}\cos\left(2\omega_c t + \pi B f_{PR} t^2 - \omega_c \tau + \pi B f_{PR}(t-t_0)^2\right) \end{aligned} \tag{3.8}$$

The signal still existing, also referred as beat-note, is shown in equation 3.9.

$$v_{MF}(t) = \frac{A_T A_R}{2}\cos\left(\omega_c t_0 + 2\pi t B f_{PR} t_0 - \pi B f_{PR} t_0{}^2\right) \tag{3.9}$$

The beat-note frequency could be obtained by taking the derivative of the phase of the cosine function after low pass filtering as following:

$$\omega_B = \frac{d(\omega_c t_0 + 2\pi t B f_{PR} t_0 - \pi B f_{PR} t_0{}^2)}{dt} = 2\pi B f_{PR} t_0 \tag{3.10}$$

FMCW radar utilizes the frequency of beat-note to compute the time delay t_0 from which the signal travels and back. Therefore, the distance between radar and the target S_0 (only considering transmitter-target path, so half of the total path), which is a function of time delay and speed of RF wave propagation (equivalent to speed of light).

$$f_B = \omega_B / 2\pi = B f_{PR} t_0 = \frac{2 B f_{PR} S_0}{c}, t_0 = \frac{2S_0}{c} \tag{3.11}$$

However, in the circumstance that the target has a constant radial velocity v_0 (could be positive or negative, depending on the moving direction of the target), the 'new' time delay t_0' related to time index t from equation 3.12 is applied in equation 3.10 to recalculate the beat-note frequency, and due to that the speed of light c is much larger than the radial velocity of the target, the second quadratic term could be ignored, as shown in equation 3.13 and 3.14.

$$t_0' = \frac{2S(t)}{c} = \frac{2(S_0 + v_0 t)}{c} \tag{3.12}$$

$$\omega_B' = \frac{d}{dt}\left(\omega_c \frac{2S_0 + 2v_0 t}{c} + 2\pi t B f_{PR} \frac{2S_0 + 2v_0 t}{c} - \pi B f_{PR}\left(\frac{2S_0 + 2v_0 t}{c}\right)^2\right) = \frac{2\omega_c v_0}{c} + 2\pi B f_{PR0} t_0 \quad (3.13)$$

$$f_B' = \omega_B' / 2\pi = \frac{2v_0 f_c}{c} + B f_{PR} t_0 \quad (3.14)$$

According to equation 3.14, the first element of the beat-note frequency indicates the Doppler shift, whereas the second element correlated to the time delay could be used to estimate the range of radar to the target. In the case scenario of AAL, the target radial velocity is much smaller compared to speed of light, therefore the beat-note frequency within a single chirp mainly depends on the target range, and this part of information can be easily extracted by performing a FFT on the beat-note signal. It should be noticed that the range/Doppler measurements will have ambiguity if the target velocity is large enough to shift the beat-note frequency to another Doppler bin, this will be further discussed in section 3.13.

Now we expand the analysis of single chirp to successive chirps, the time delay to the target t_0'' could be rewritten as equation 3.15:

$$t_0'' = \frac{2S_0 + 2v_0(t + n_c T_{PR})}{c} \quad (3.15)$$

where t now denotes the time within the n_c^{th} chirp and n_c denotes the number of chirps analysed. The beat-note frequency in the n_c^{th} chirp could be derived by substituting t_0'' in equation 3.10:

$$\omega_B'' = \frac{d\phi_B(t)}{dt} = \frac{d}{dt}\left(\begin{array}{l}\omega_c \frac{2S_0 + 2v_0(t + n_c T_{PR})}{c} + 2\pi t B f_{PR} \frac{2S_0 + 2v_0(t + n_c T_{PR})}{c} \\ -\pi B f_{PR}\left(\frac{2S_0 + 2v_0(t + n_c T_{PR})}{c}\right)^2\end{array}\right) \quad (3.16)$$

$$= \frac{2\omega_c v_0}{c} + 2\pi B f_{PR} t_0 + \frac{4\pi B v_0 n_c}{c}$$

$$f_B'' = \frac{2 f_c v_0}{c} + B f_{PR} t_0 + \frac{2 B v_0 n_c}{c} \quad (3.17)$$

Equations 3.12 and 3.13 show the beat-note frequency in the n_c^{th} chirp after eliminating the terms which are small compared to others or to one radian. Comparing equation 3.13 with 3.10, the beat-note frequency in the n_c^{th} chirp is identical to the one in a single chirp aside from an additional term, which indicates the target movement within the duration of N_c chirps. In the case that large numbers of chirps are integrated, meaning a very long observation time, this term may not be negligible.

3.1.2 Double Fast Fourier Transform Processing

Similar as equation 3.14, Doppler and range information of the beat-note signal could be extracted through transferring from time to frequency domain. However, there are two different techniques to perform this spectral analysis, namely, double and single FFT processing. Double FFT, as its name, requires two FFT processes for displaying the range to the target and target radial velocity separately, where the former FFT is performed on a single chirp to show time delay (range measurements) and the latter FFT is performed on N_c successive chirps to show Doppler shift (velocity measurements). However, single FFT-based approach exploits a single but longer FFT process on N_c successive chirps to extract both information. These two approaches are equivalent in terms of parameters involved as well as processing time. In the following thesis, double FFT is discussed in detail as it is used.

The equation 3.15 shows the FFT process of the beat-note signal in equation 3.18:

$$\phi_B(t) = \phi_0 + 4\pi f_c \frac{v}{c} n_c T_{PR} + 2\pi\left(\frac{2v_0}{c} f_c + Bf_{PR}t_0 + \frac{2v_0}{c} Bn_c\right)t, \phi_0 = \omega_c t_0 \tag{3.18}$$

$$V(f) = \int_{-\frac{T_{PR}}{2}}^{\frac{T_{PR}}{2}} A\cos(\phi_B(t))e^{-j2\pi ft}dt \tag{3.19}$$

After resolving the integral in equation 3.19, it could be rewritten as following:

$$\begin{aligned} V(f) = {} & \frac{AT_{PR}}{2}\left(\frac{\sin(2\pi(f - f_B))T_{PR}/2}{2\pi(f - f_B)T_{PR}/2}\right)e^{-j\phi_0 + j2\pi t\frac{2v_0}{c}n_c T_{PR}} \\ & + \frac{AT_{PR}}{2}\left(\frac{\sin(2\pi(f + f_B))T_{PR}/2}{2\pi(f + f_B)T_{PR}/2}\right)e^{j\phi_0 - j2\pi t\frac{2v_0}{c}n_c T_{PR}} \end{aligned} \tag{3.20}$$

where f_B denotes the beat-note frequency. It can be seen that the spectrum of single chirp is in the format of sin *x*/*x* with the centre at the beat-note frequency. This FFT process is applied on digitized beat-note signals, therefore the sampling frequency needs to be at least twice of the maximum expected beat-note frequency for fulfilling Nyquist law. In our case, the chirp signal is sampled *K* times within duration T_{PR}, due to that, *K*/2 frequency samples are produced in the spectral analysis. Additionally, those frequency samples could be utilized to compute target time delay, which is a basis for the range estimation.

$$
\begin{bmatrix}
H_{11} & H_{12} & \cdots & H_{1k} & \cdots & H_{1\frac{K}{2}} \\
H_{21} & H_{22} & \cdots & H_{2k} & \cdots & H_{2\frac{K}{2}} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
H_{n_c 1} & H_{n_c 2} & \cdots & H_{n_c k} & \cdots & H_{n_c \frac{K}{2}} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
H_{N_c 1} & H_{N_c 2} & \cdots & H_{N_c k} & \cdots & H_{N_c \frac{K}{2}}
\end{bmatrix} \tag{3.21}
$$

A matrix shown in equation 3.21, whose dimension is N_c x $K/2$, are constructed by repeating the first FFT on each single chirp. The rows denote the slow time (chirp number times PRI), each element of one row is a frequency domain sample correlated with the target time delay. For the perspective of time, the duration of each row is T_{PR}, thus it takes $N_c T_{PR}$ to fill in this matrix. The columns represent the fast time, also referred as range bins. Each range bin contains all digital samples at a specific frequency bin of the beat-note signal, hence, the amplitude change within the range bin is fairly small and could be considered as a constant. However, the phase change within the range bin, as depicted in equation 3.16, could be utilized to obtain Doppler shift due to the movement of the target by performing a second Fourier transform on N successive chips.

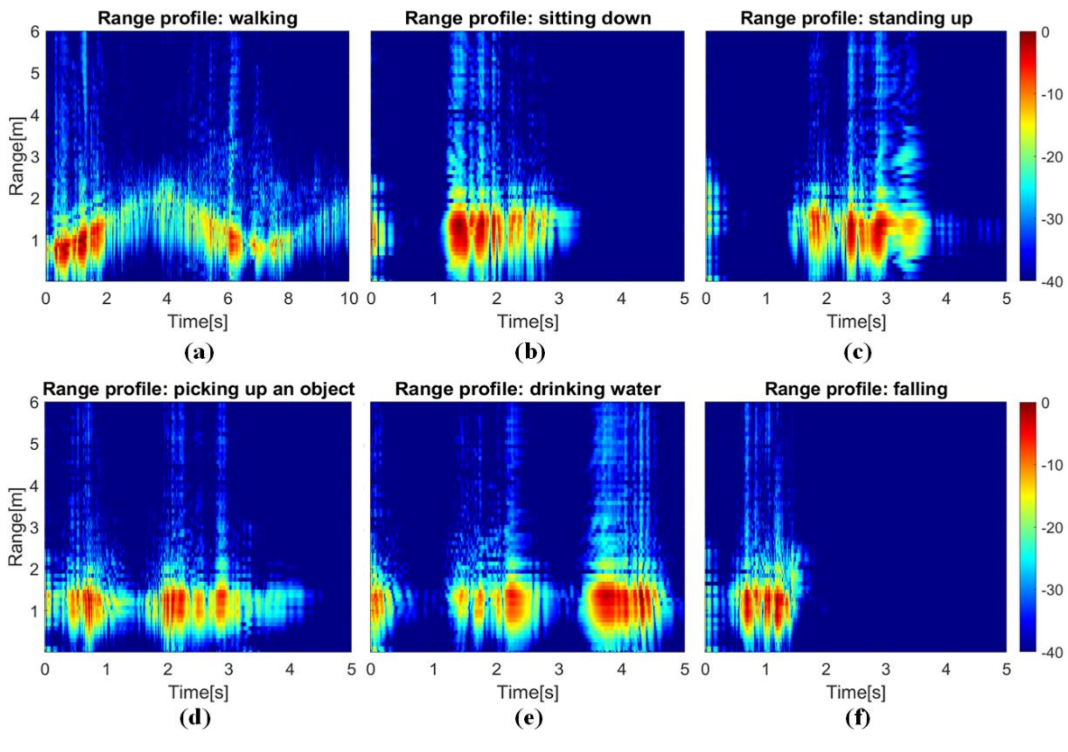


Figure 3.3 Range-time maps for six different activities: (a) walking with normal speed 10s (b) sitting down on a chair and still 5s (c) standing from a chair and still 5s (d) picking up a pen and drop 5s (e) drinking water two times from a cup and put it back when finishing (f) simulating a frontal fall on a mattress

Figure 3.3 above illustrates range-time maps obtained by the first FFT process for six different human activities, where the colour level of these heatmaps indicates the received signal strength from 0 to -40 dB. (a) represents walking with the swinging of arms as its periodic range pattern, (b) and (c) are symmetric so they denote sitting down and stand up separately, whereas (f) is the simulation of a fall. It should be noticed that a 4^{th} order Butterworth Notch filter (cut-off frequency -0.075 and 0.075 Hz) is utilized as a Moving Target Indication (MTI) [113] approach on the range measurement, in order to filter the static clutter, therefore no range readings are available when the subject is not moving (e.g. sitting on the chair, standing still and lying on the mattress after falling). Additionally, the smearing across the range is caused by the radar device itself, which could be ignored due to low signal amplitude (<-25dB).

The phase factor of the first FFT result could be rewritten as following [44]:

$$H_{n_c k} = E(f) e^{j 2\pi t \frac{2v_0}{c} t_{n_c}} \tag{3.22}$$

where $E(f)$ is the amplitude of frequency domain sample and $n_c T_{PR}$ is substituted by t_n to express the discrete flow of time from the first chirp to the n_c^{th}chirp. The second FFT of $H_{n_c,k}$ over N_c successive chirps, starting at 0 and ending at $N_c T_{PR}$, is shown in equation 3.23 [44]:

$$H_k = N_c T_{PR} E(f) \left(\frac{\sin(2\pi(f - 2v_0 f_c / c)) N_c T_{PR} / 2}{2\pi(f - 2v_0 f_c / c) N_c T_{PR} / 2} \right) \tag{3.23}$$

Like the estimation of range, equation 3.23 indicates the shift of the centre of sin *x*/*x* function caused by the Doppler Effect. Hence the radial velocity of the target v_0 could be derived through the amount of Doppler shift after the second Fourier transform.

The output of the second Fourier transform contains values calculated at N_c discrete frequency points. Since the chirp duration is correlated with the MDS, for avoiding the ambiguity on Doppler measurements, T_{PR} should be equal or shorter than 1/2MDS. In addition, the sin *x*/*x* function depicted in equations 3.20 and 3.23 yield high-level side lobes, which may result in the overlapping of main lobes corresponded to other targets. To overcome this, the data needs to be multiplied by a 2-D window function prior to double FFT processing, for instance 2-D hamming window in our case. However, it should be noticed that the reduction of side lobe level cannot be done without compromising the main lobe.

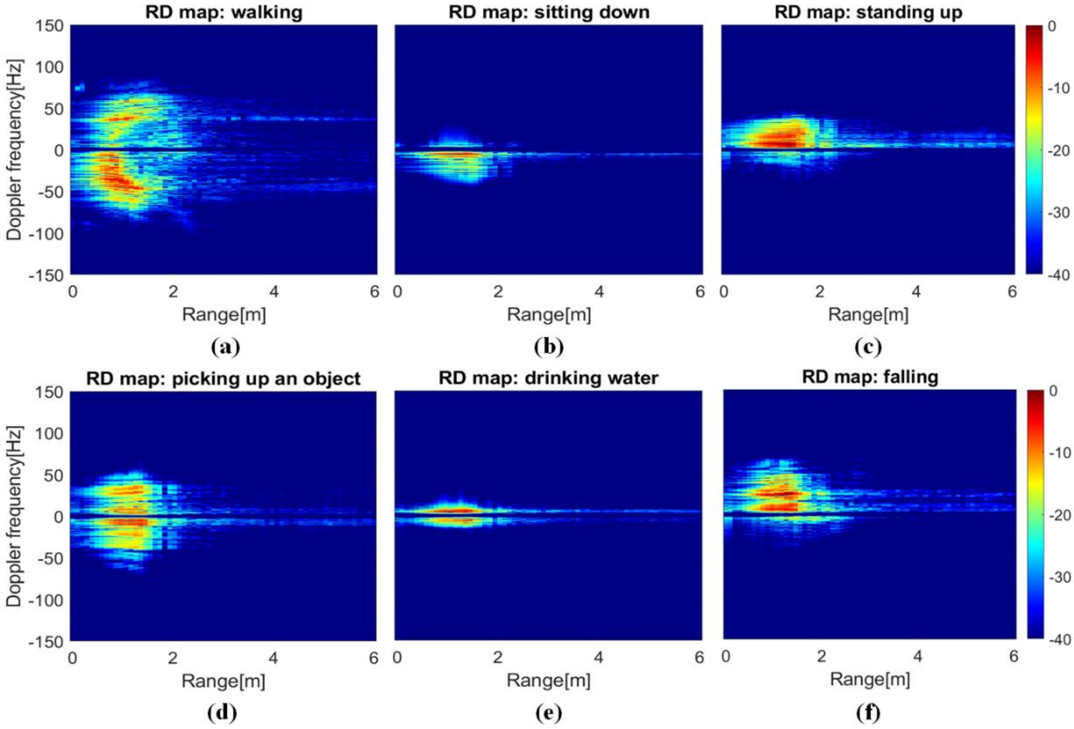


Figure 3.4 Integrated time range-Doppler maps [19] for six different activities: (a) walking with normal speed 10s (b) sitting down on a chair and still 5s (c) standing from a chair and still 5s (d) picking up a pen and drop 5s (e) drinking water two times from a cup and put it back when finishing (f) simulating a frontal fall on a mattress

The integrated time range-Doppler maps in figure 3.4 are constructed by coherently summing successive range-Doppler frames, where each frame is the output of the second FFT process. In our case, N_c is set as 400 and the duration of the chirp equates to 1 ms, so each frame contains range and Doppler information within 0.4 s. For the integrated-time range-Doppler maps, each row denotes a Doppler bin, whereas each column denotes a range bin. It should be noticed that the Doppler component at 0 Hz has been filtered out due to the aforementioned MTI process. Compared to range-time maps illustrated in figure 3.3, it is interesting to notice that pairs of similar activities appear to yield more difference within the range-Doppler domain, mainly on the shape (for picking up an object from the ground and drinking water with a paper cup) and power amplitude (for sitting down and standing up) of signature. The Maximum Doppler Shift (MDS) of the proposed six daily activities is about ±75 Hz. Thus there is no ambiguity in the Doppler measurements.

3.1.3 Summary of FMCW Radar Parameters

- Range Resolution: as depicted in equation 3.1, the range resolution ΔR of the FMCW radar system is correlated with radar chirp bandwidth. It is observed that the wider bandwidth *B* will lead to the finer range resolution. In some cases, radar needs very fine range resolution to separate very close targets, thus, UWB systems have gained lots of interest.

- Angular/Cross-range Resolution: angular resolution of the radar system is orthogonal to the radar range measurement; hence, it is also referred as cross-range resolution. The equation 3.24 shows the approximation of angular resolution, where λ denotes the wavelength, R denotes the range of radar system to the target, and L denotes the length of antenna array used. However, since the size and weight of the radar system are important factors for some applications, it is impossible to obtain very fine angular resolution by keep increasing the length of antenna array.

$$\Delta A = \lambda R / L \tag{3.24}$$

- SNR: SNR is an important parameter to measure the quality of received echo signal, it could be derived as equation 3.25, where P_t is the peak transmit power, τ is pulse duration, G_t and G_r denote the transmitter and receiver gain, respectively, λ is the wavelength corresponded to the operating frequency, σ is the target RCS, k is the Boltzmann constant, L refers to the global loss factor that takes both system and signal propagation loss into account, T_s is the noise temperature, R_t and R_r denote the range from the radar transmitter and receiver to the moving target, respectively.

$$SNR = \frac{P_t \tau G_t G_r \lambda^2 \sigma}{(4\pi)^3 k L T_s R_t^2 R_r^2} \tag{3.25}$$

- Maximum Detectable Range: the MDR could be obtained by setting a minimum SNR for the radar system.
- Number of Samples in Each Chirp: it should be noted that the range measurement of FMCW radar is related to the beat-note frequency of the signal after matched filtering, as depicted in equation 3.18, therefore, the maximum expected beat-note frequency could be derived when the MDR is known. In FMCW radar system, the signal needs to be digitized through an ADC eventually, and due to the requirements of Nyquist sampling law, the sampling frequency of that ADC should be at least twice as the beat-note frequency. Combining the Nyquist criterion and equation 3.15, it is possible to find a relationship between the number of samples per digitized chirp and maximum detectable range. The relationship is given in equation 3.26 below:

$$f_s \geq 2 f_B \equiv Z \geq MDR / \Delta R \tag{3.26}$$

  Where f_s is the ADC sampling frequency, f_B is the beat-note frequency, Z denotes the number of samples per chirp.
- Maximum Doppler Shift: since Doppler shift is proportional with target velocity, the MDS could be computed as shown in equation 3.27, where f_c the operating frequency and c is the speed of light.

$$MDS = f_c V_{\max} / c \tag{3.27}$$

In order to avoid Doppler ambiguity, the chirp duration needs to follow the constraint in equation 3.28. When the target is moving in a high velocity, the chirp duration T_{PR} should be shorter.

$$T_{PR} < 1/2MDS \tag{3.28}$$

- Maximum Unambiguous Range: the MUR refers to the longest distance from which a transmitted pulse could be received before the next pulse is emitted. Any detection over MUR will have an ambiguity as radar can't identify whether the echo signal belongs to the first or second pulse. The calculation of MUR is shown in equation 3.29, since chirp duration is the reciprocal of PRF, the higher PRF, the shorter MUR is.

$$MUR = cT_{PR} / 2 \tag{3.29}$$

- Doppler Resolution: compared to radar range resolution, Doppler resolution depends on the integration time (number of successive radar chirps processed times pulse duration), as shown in equation below, where N_c is the number of chirps analysed. It is obvious that higher N_c will lead to finer frequency resolution, however, there is a compromise between the value of N_c and data processing time of the radar system, and for the convenience of users, the processing time should be as short as possible.

$$\Delta f_D = 1 / N_c T_{PR} \tag{3.30}$$

3.1.4 Micro-Doppler Signature and its Time-Frequency Transformations

The raw radar data is comprised of IQ complex samples with the real and imaginary part as digitized by the ADC module of the radar (The IQ block is not shown in figure 3.1 for simplification). To obtain the range information of moving target, the first FFT is performed on the raw complex data and a 4th order Butterworth Notch filter with cut-off frequency equal to ±0.075 Hz is utilized to remove the non-of-interest objects (such as furniture, wall), which are able to create confusion when characterizing the different activities and movements performed by the testing subjects.

However, instead of performing the second FFT on the range data to enter the range-Doppler domain, Time-Frequency (TF) analysis [114], [115] can be applied on those range bins containing target signatures in order to generate spectrograms, also referred as micro-Doppler signatures [8], [114], [116], [117]. There are several algorithms for the TF analysis, notably, Short Time Fourier Transform (STFT) [8], [20], [118] that performs FFT in a sliding window manner and combines the resulting vectors, Discrete Wavelet Transform (DWT) [119]–[121] that uses a multi-level low-high pass filter structure to extract the approximate and detailed coefficients of the input signal, and Hilbert-Huang Transform [122], [123] that decomposes the signal into Intrinsic Mode Functions (IMF) and applies Hilbert Transform to obtain their instantaneous frequency/energy. In this thesis, STFT is chosen as the main

TF analysis method as it is widely used in the field of AAL and less computational intensive than the others. Besides that, DWT is ideal for characterizing movements with high frequency change in short duration (e.g. drones and vehicles) [124], whereas Hilbert-Huang Transform prefers non-stationary or nonlinear dataset. However, STFT, given in equation 3.31 has to compromise between the time and frequency resolution and the factor to affect them is the length of the discrete window. In details, longer discrete window will lead to finer frequency resolution with the loss of time resolution, and vice versa. The discrete window function *w* used in this thesis is a Hamming window, whose length equates to *m* and the overlap between two nearest windows equates to R_o, in our case, *m* is typically 128 samples and R_o is 95%. The discrete frequency set *p* could be derived from Doppler frequency f_d with the help of equation 3.32, where f_s denotes the radar sampling frequency (equivalent to PRF after data processing) and *N* denotes the number of samples within one hamming window.

$$STFT(\omega, n) = \sum_{n=-\infty}^{\infty} x(n)w(n - mR_o)e^{-j2\pi np/N} \tag{3.31}$$

$$p = N\omega / f_s \tag{3.32}$$

In addition to the main Doppler frequency shift induced by the target movement, there are often some sidebands regarding the main shift, which are caused by minute oscillatory or translational movements of different parts of the target (e.g. blade rotation of a helicopter), and this is known as micro-Doppler effect [21], [36], [114], [125]. Figure 3.5 illustrates the micro-Doppler signatures for same activities depicted in figure 3.3 and 3.4. It can be seen that movements of different body parts have distinctive patterns due to micro-Doppler effect, for instance in (a) the central mass of the signature represents torso movement and these spikes upon the central mass represent the swings of limbs and rotations of head. Features extracted from micro-Doppler signatures are able to characterize the small difference among similar activity pairs since they are correlated with micro-motion dynamics that the testing subject undergoes.

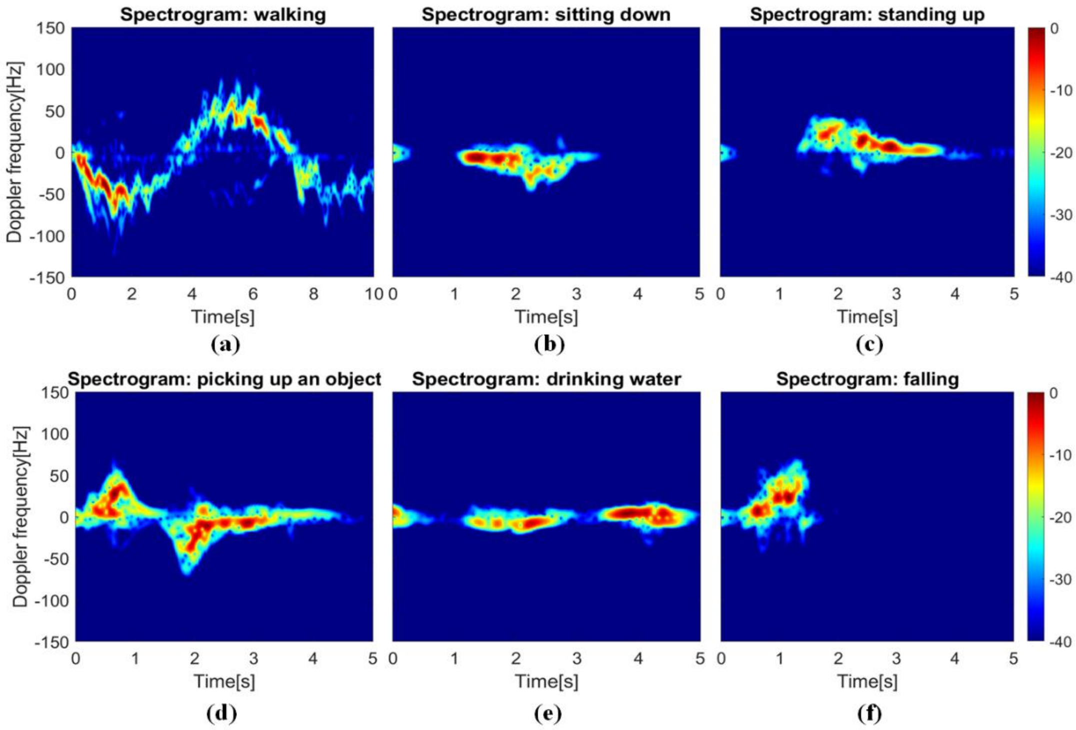


Figure 3.5 Spectrograms for six different activities: (a) walking with normal speed 10s (b) sitting down on a chair and still 5s (c) standing from a chair and still 5s (d) picking up a pen and drop 5s (e) drinking water two times from a cup and put it back when finishing (f) simulating a frontal fall on a mattress

Due to that STFT performs FFT in a sliding window manner, the periodicity of the backscattered signal will remain in its micro-Doppler signature, with every Doppler bin yielding the same periodicity. Thus, compared to signal depiction in time-domain, the periodic properties of the cyclic human activities such as walking are persistent within the TF domain, with a more sparse representation and higher power content. By taking a FFT along the time-vector of the micro-Doppler signature, as shown in equation 3.33 it is easy to extract the appearance frequency of certain Doppler shift over a time duration. The FFT results are known as Cadence Velocity Diagram (CVD) [8], [21] [126]–[128], in which the target velocity is linear with the Doppler shift. CVD is an interesting tool in the analysis of human gaits since the most significant cadence frequency denotes human stride rate.

$$CVD(\omega,\omega_{cad}) = \left| \sum_{n=0}^{N-1} STFT(\omega,n) e^{-j2\pi n p_{cad}/N} \right| \tag{3.33}$$

$$p_{cad} = N\omega_{cad} / f_s \tag{3.34}$$

$$v_0 = \lambda\omega_{cad} / 2 \tag{3.35}$$

where *STFT* (ω,n) denotes the spectrogram matrix, the discrete cadence set p_{cad} could be used to compute the cadence frequency set ω_{cad} through equation 3.34 as well as target velocity v_0 through 3.35. Figure 3.6 illustrates the CVD for six daily activities, where the spread, shape and periodicity of the Doppler signatures are exhibited. Compared to pure micro-Doppler signatures in figure 3.5, (a) and (e) show that how often certain Doppler frequency arises within walking and drinking water cycle, separately. For walking, it has been confirmed in our data that ±50 Hz are the fundamental frequency components, whereas for drinking water, the numbers are reduced to nearly ±5 Hz.

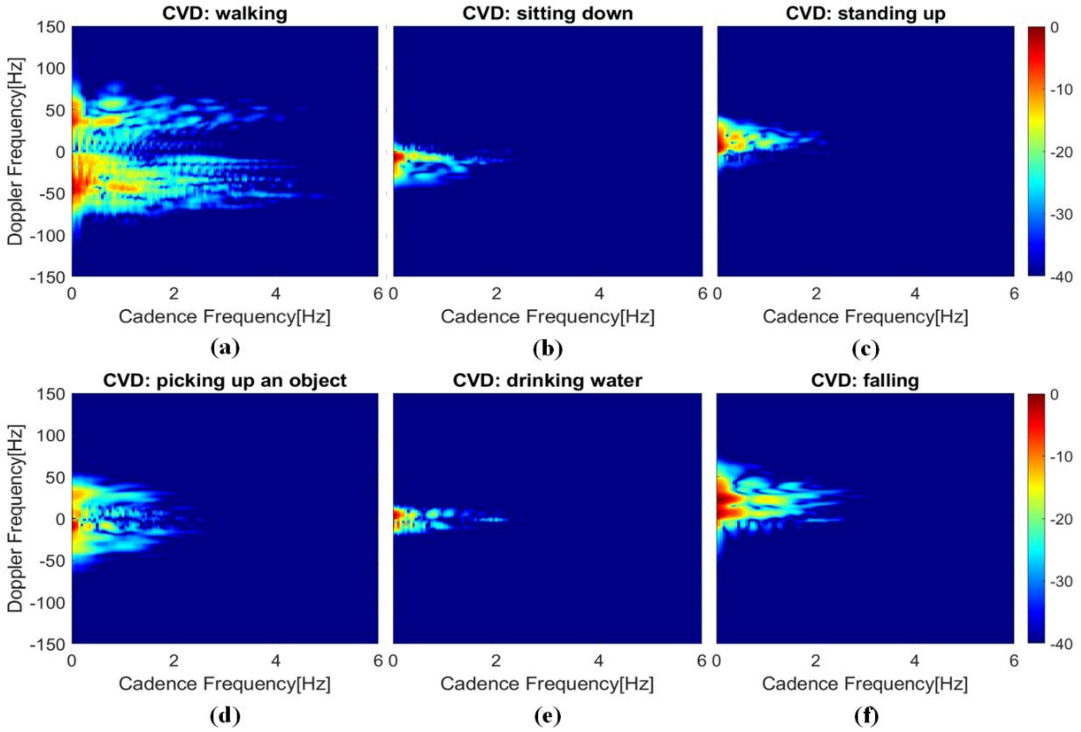


Figure 3.6 CVD for six different activities: (a) walking with normal speed 10s (b) sitting down on a chair and still 5s (c) standing from a chair and still 5s (d) picking up a pen and drop 5s (e) drinking water two times from a cup and put it back when finishing (f) simulating a frontal fall on a mattress

Similar to CVD, cepstrum is used to convert signals mixed by convolution into sum of their cepstra, for the purpose of linear separation. Radar cepstrum [8], [129]–[131], whose information content is highly related with the energy of the spectrogram, is given in equation 3.36 by taking the Inverse Discrete Fourier Transform (IDFT) of the logarithm of the absolute energy within the spectrogram, where the Quefrequency axis is in the dimension of pseudo units (millisecond), representing inverse of frequency. The cepstrum of six different activities are drawn in figure 3.7, the signal energy involved in certain Doppler frequency are shown in a more prominent manner by applying IDFT to the log scale of the spectral energy within each short hamming window.

$$Ceps(\omega_q, n) = \left| F^{-1}[\log(\left|STFT(\omega, n)\right|^2 + 1]\right|^2 \tag{3.36}$$

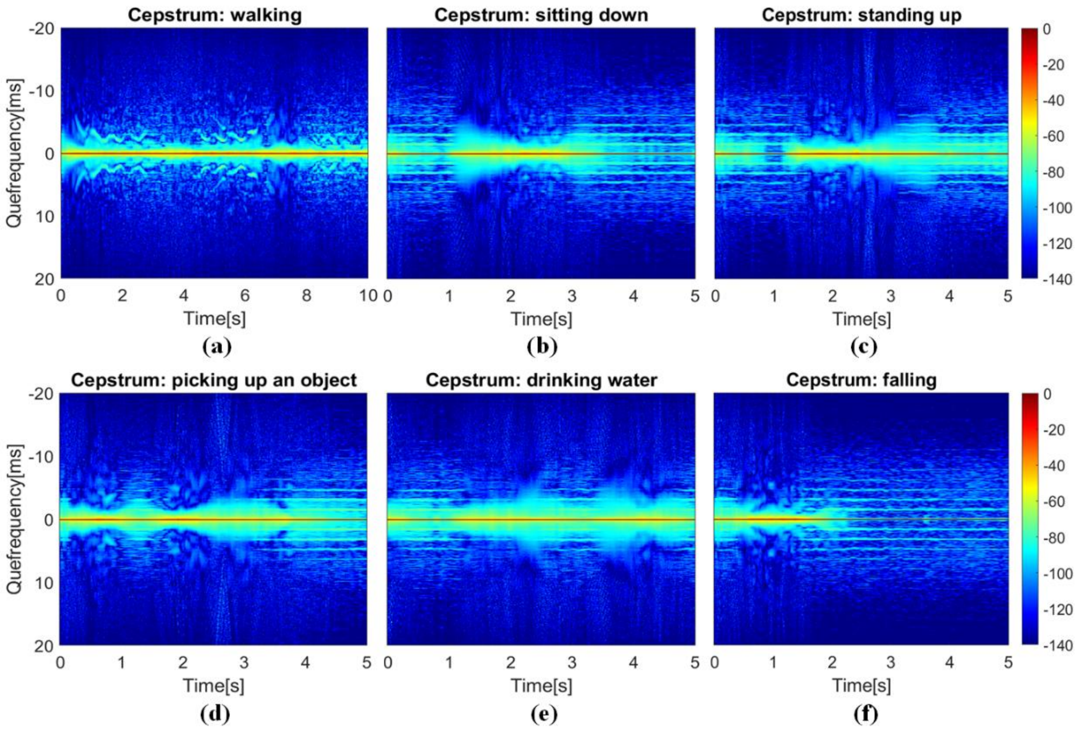


Figure 3.7 Cepstrum for six different activities: (a) walking with normal speed 10s (b) sitting down on a chair and still 5s (c) standing from a chair and still 5s (d) picking up a pen and drop 5s (e) drinking water two times from a cup and put it back

3.2 Wearable IMU Sensor

Wearable IMU data are more straightforward to process than the radar data. First, the raw wearable data needs to be pre-processed via a band pass filter prior to the feature extraction and classification. The high-pass part removes the DC component, whereas the low-pass part filters the white noise caused by tilting the IMU sensor and other high-frequency electronic noise. The cut-off frequency for high and low band are chosen empirically from [76] as 0.4 Hz and 10 Hz, respectively, therefore the non-of-interest activities and movements, whose frequency bands are out of range, are eliminated from the signal profile. The figure 3.8 below compares the raw acceleration of walking 10s with the filtered data. Obviously, the filtered one contains less noise content (DC removal and signal becomes smoother) and its value are more approaching to the actual acceleration that testing subject undergoes. In addition, the signal pattern indicates regular peaks with similar amplitude, since the wearable IMU is placed on the wrist of testing subject and the testing subject is swinging his/her arm in an almost-fixed frequency during the walking.

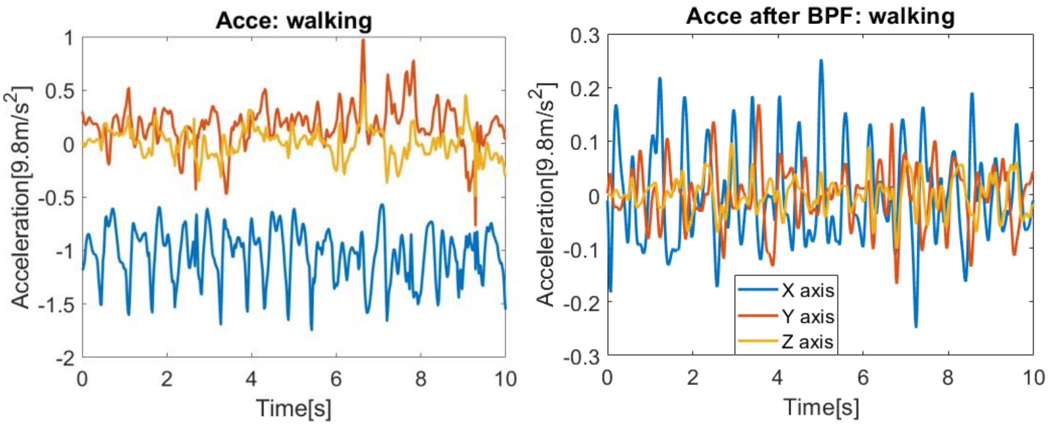


Figure 3.8 Acceleration of 10s walking: raw data (left); data filtered by a band-pass filter with cut-off frequency equal to 0.4 Hz and 10 Hz. The unit of acceleration is $9.8 m/s^2$.

3.3 Summary of the Chapter

In this chapter, the pre-processing of radar and wearable IMU data are discussed. For the radar, the transmitted and received chirp signal are analysed mathematically and the double FFT processing to extract range and Doppler information is derived in details. Micro-Doppler signatures for classifying ‘micro-motions’ of individual body parts and other useful information domains involving CVD and cepstrum are also shown.

For the wearable IMU, the signal filtering process to remove both DC component and noise due to sensor tilting are presented. In the next chapter, we are going to discuss the machine learning related techniques (feature extraction, classification, fusion) used in this thesis.

4 Machine Learning and Artificial Intelligence in AAL

This chapter introduces the machine learning-based techniques used in this thesis. The techniques involves extraction of handcrafted features (both radar and wearable IMU sensor), classification algorithms (both conventional classifiers and deep neural networks), and data fusion methods to leverage the strengths of different sensors.

4.1 Handcrafted Feature Extraction

4.1.1 Radar Features

For the conventional classifier, prior to the classification, handcrafted features are extracted from the raw data to characterize each class in specific manner. In general, feature extraction reduces the dimension of classifier inputs, however, it may lose significant information due to compressing the raw data into one numerical number.

Radar features are divided into physical features of Doppler spectrogram, transform-based features and range-Doppler features, as shown in table 4.1, where physical features of Doppler spectrogram includes Doppler centroid, bandwidth, upper and lower envelope as well as higher order statistical parameters (e.g. skewness and kurtosis) of the spectrogram. Doppler centroid and bandwidth [132] are the most salient features, where the centroid represents the translational position of the central mass of the human torso and limbs, the bandwidth indicates the energy spread around the central mass and this is highly related with the range of human motions especially from the swing of limbs. They can be derived from equation 4.1 [133] and 4.2 [133] respectively.

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

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

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

The upper and lower envelopes [22], [134] express the peak velocity of the human arms and legs movements towards/away from the radar separately. They are computed from the spectrogram using a percentile-based method, the percentile value as a function of time is defined:

$$P(E(i),i)=\frac{\sum_{E(i)=f_{\min}}^{E(i)} S(E(i),i)}{\sum_{E(i)=f_{\min}}^{f_{\max}} S(E(i),i)} \tag{4.3}$$

Where $E(i)$ denotes the envelope index at the i^{th} time bin, f is a vector that storing the Doppler frequency values ($f_{\min}$ and $f_{\max}$ are the minimum and maximum Doppler frequency) and $S(E(i),i)$ represents the spectrogram element at row $E(i)$ and column i. $P(E(i),i)$ is set to 0.98 and 0.02 for the computation of upper and lower envelope. The difference between the upper and lower envelope is also calculated for characterizing the variance of peak velocity at two opposite directions. Energy curve of the Doppler and time bins of the spectrogram takes the coefficients of the bins and their subsequent moments to enumerate the energy within a given frequency band. The equation of energy curve could be derived as below:

$$EC(i)=\sum_{j=f_1}^{f_2}|S(j,i)|^2+\sum_{j=-f_2}^{-f_1}|S(j,i)|^2 \tag{4.4}$$

Where $S(j,i)$ denotes the spectrogram element at the i^{th} time bin and the j^{th} Doppler bin, the frequency band in our case is selected between 70 Hz (f_1) and 100 Hz (f_2) to detect events with high energy content, for instance, falls.

Transform-based features performs a mathematical transformation including SVD (Singular Value Decomposition) [20], [135], DCT (Discrete Cosine Transform) [22], [109] and LPC (Linear Predictive Coding) [22], [136] on the Doppler spectrogram to search useful hidden information. Singular Value Decomposition based spectral and temporal projections reduce the information within the spectrogram to the first few vectors of U and V matrices. The statistical moments of these indicate the amount of information/motion in the overall time and frequency bins of the spectrogram. LPC has been generally utilized in the speech signal compression since it is capable of representing the original signal as a linear combination of previous values, similarly, those coefficients from a linear predictor could be also employed on the time-dependent Doppler spectrogram. Furthermore, the received echo signal in equation 3.2 only involves cosine transforms, therefore, DCT is more efficient for expressing the Doppler spectrogram in a few DCT coefficients (in our case, first 10 coefficients).

Features from other radar domains such as CVD [21], [41] and cepstrum [22], [130], [137] are also categorized as transform-based features since additional transform function (a FFT along the spectrogram time axis for the CVD and an IFT of the logarithm of the spectrogram for the cepstrum) is essential prior to the feature extraction. Step repetition frequency and its amplitude, along with central component (the most significant Doppler frequency), upper and lower bands of the main peaks are extracted from the CVD profile. The step repetition frequency denotes the cadence frequency index with respect to the second peak (first is at 0 Hz) in the average CVD profile, as shown in figure 4.1 (middle), whereas the amplitude of Doppler bins that corresponded to this cadence frequency index are sketched in figure 4.1 (right). The Doppler frequency with respect to the maximum of main peaks is considered as the most significant Doppler frequency. For the upper and lower bands of main peaks, a threshold-based algorithm is used to determine their positions, as depicted in equation 4.5:

$$T = 0.05 * diff \tag{4.5}$$

Where T denotes the threshold and $diff$ indicates the amplitude difference between Doppler bins yielding maximal and minimal value. The Doppler bins closest to the threshold at both sides of the main peaks are output as the upper and lower band, respectively. Beyond that, the area of main peaks, also referred as energy of main peaks, is computed through cumulative trapezoidal numerical integration as part of the CVD features.

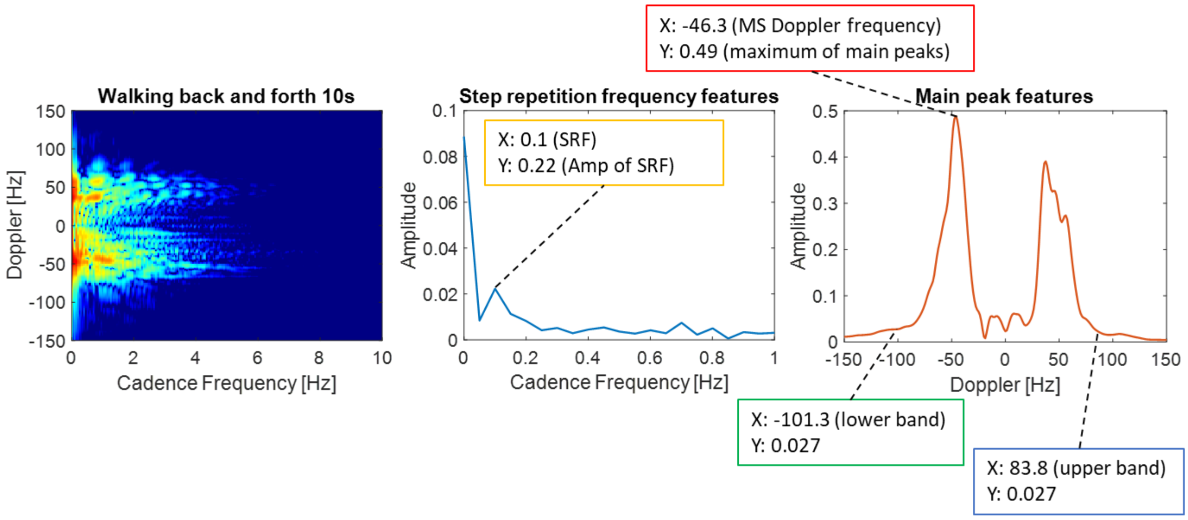


Figure 4.1 CVD features from 10s walking (left: CVD profile for walking; middle: average CVD profile; right: main peaks of the spectrum)

Furthermore, cepstral features including two-dimensional mean, standard deviation and minimum of the radar cepstrum are added as an additional source of information to characterize the periodicity of movements. Three features are extracted from the integrated time range-Doppler map, namely, range-Doppler velocity, range-Doppler displacement and range-Doppler dispersion to complement information from other domains. Range-Doppler velocity represents the maximum Doppler frequency along the range axis, whereas range-Doppler displacement measures the average distance between radar and testing subject with respect to each Doppler bin, and range-Doppler dispersion denotes the 'extension' of the average distance due to swing of limbs. Those two features (range-Doppler displacement and dispersion) are computed in the same way as Doppler centroid and bandwidth from the spectrogram, where the time and Doppler bins in equation 4.1 and 4.2 are substituted to the Doppler and range bins, respectively.

All the features in table 4.1 are selected from different works in the literature. Through this, it is expected to increase the feature diversity and the overall relevant information for human activity and gait classification.

Feature ID	**Physical features**	**No. of features**
1-4	Mean, standard deviation, skewness, and kurtosis of the centroid of the Doppler spectrogram	4
5-8	Mean, standard deviation, skewness, and kurtosis of the bandwidth of the Doppler spectrogram	4
9-12	Two-dimensional mean, standard deviation, skewness and kurtosis of the whole segment of the spectrogram	4
13	Entropy of the Doppler spectrogram	1
14-16	Mean, standard deviation and range of energy curve of the Doppler spctrogram	3
17-19	Mean, maximum and minimum of the upper envelope	3
20-22	Mean, maximum and minimum of the lower envelope	3
23	Difference between the mean of the upper and lower envelope	1
	Transform-based features	**No. of features**
24-35	Mean and standard deviation of the first three left and right eigenvectors of the SVD decomposition of the spectrogram	12
36-37	Sum of pixels of the entire left and right matrices	2
38-39	Mean of the diagonal of the left and right matrices	2
40-49	Discrete Cosine Transform of the spectrogram	10
50-59	First 10 coefficients of the LPC of the spectrogram	10
60	Step repetition frequency	1
61-62	Upper and lower bands of the main peaks	2
63	Amplitude of the step repetition frequency	1
64	Maximum of the main peaks	1
65	Energy of the main peaks	1
66	Most significant Doppler frequency in CVD	1
67-69	Two-dimensional mean, standard deviation and minimum of the cepstrum	3
	Range-Doppler features	**No. of features**
70	Mean of Range-Doppler velocity	1
71	Mean of Range-Doppler displacement	1
72	Mean of Range-Doppler dispersion	1
	Total number of features	**72**

Table 4.1 List of the radar features

4.1.2 Wearable IMU features

Totally 64 features from wearable sensor can be extracted from the filtered signal of wearable Inertial Measurement Unit (IMU) sensor along the 3 axes X, Y and Z, those are listed in table 4.2, where temporal features [66], [76] and spectral features [66], [138] are generated from time and frequency domain separately. Temporal features such as mean, variance, skewness and kurtosis are utilized to characterize the deviation level of the data, whereas correlation-based features, especially the cross-correlation between two different axes of the signal is used in classifying human motions that containing the changes of signal amplitude along two kinetic dimensions (e.g. rotation of the body). Spectral features denotes the power content of the wearable signal and its distribution, involving the sum of the amplitude of the Power Spectral Density (PSD) at three selected frequency bands, notably, 0.5-1 Hz, 1-5 Hz and 5-10 Hz, the accumulation of Fourier Transform coefficients, and the entropy of the normalized PSD. The derivation of spectral entropy is given following in equation 4.6 and 4.7:

$$P(k) = \frac{S(k)}{\sum_j S(j)} \tag{4.6}$$

$$H = -\frac{\sum_{k=1}^{N} P(k)\log_2 P(k)}{\log_2 N} \tag{4.7}$$

Where $S(k)$ is the power spectrum of one channel (e.g. X axis of accelerometer) within IMU signals, the probability distribution $P(k)$ of the power spectrum is obtained by normalizing the $S(k)$ with the sum of power spectrum and N is the total number of frequency points.

Temporal features	#	Spectral features	#
Norm of XYZ	1	Spectral Power	9
Mean	3	FFT 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**

Table 4.2 List of wearable IMU features

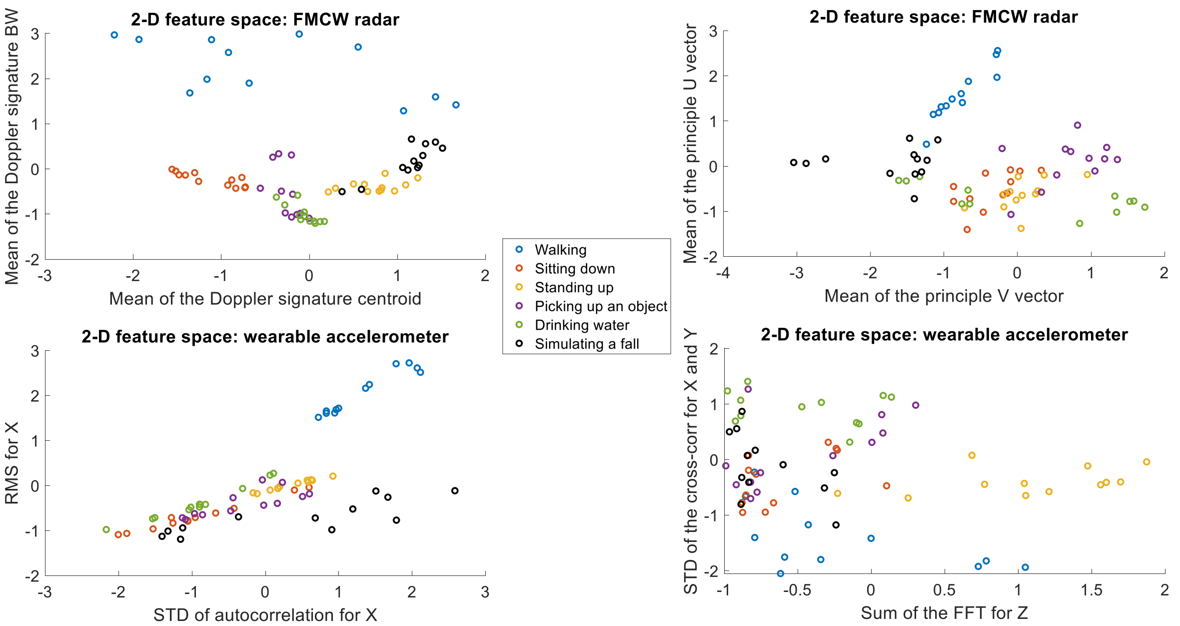


Figure 4.2 The feature spaces for FMCW radar (top) and accelerometer in the wrist-IMU (bottom)

Figure 4.2 illustrates the feature spaces of FMCW radar and accelerometer within the wrist-IMU with respect to six activities, where both sensors contribute four significant features. For the radar, mean of the centroid and bandwidth of the Doppler signature are selected, along with the mean of the principle U and V vectors after SVD, whereas for the wearable accelerometer, Root Mean Square (RMS) of the X axis data and standard deviation of X axis autocorrelation are picked from the feature set, as well as standard deviation of the cross-correlation between X and Y axes and sum of the FFT coefficients for Z axis. It is reported that centroid and bandwidth can distinguish the six activities well except 'picking up an object' and 'drinking water'. Those two activities could be identified by SVD-based features, however, the data points of 'sitting down' and 'standing up' are hard to separate in this case. In terms of RMS and autocorrelation features from accelerometer, the data points of 'drinking water' are mixed with other classes and some fall events are closer to 'sitting down', whereas the cross-correlation feature and sum of the FFT coefficients are able to recognize 'walking', 'standing up' and 'drinking water', but the rest of the classes, especially 'sitting down' and 'simulating a fall', are seriously confused. To conclude, using only one feature may not be enough to separate all the classes in a good manner, thus different kinds of features should be combined to form a feature pool (feature matrix) for training a classifer with higher robustness in classifiying human acitivities.

4.2 Machine Learning Algorithm

4.2.1 Support Vector Machine

Support Vector Machine (SVM) [94], [139] is known as a robust classifier in the field of indoor human activity recognition (HAR), it intends to build a hyperplane to separate the feature points of different classes based on the distribution of the features, as shown in figure 4.3. The support vectors are the feature points close to the decision boundary and they are able to control the position and orientation of the hyperplane, whereas the margin between the positive and negative hyperplane needs to be maximized through those support vectors.

The mathematical representation of a linear SVM hyperplane and its objective function are given as follows:

$$h : x'W + b = 0 \tag{4.8}$$

$$\min_{W,b,\xi_i} \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{n}\xi_i (C > 0, \xi_i \geq 0) \tag{4.9}$$

Where W denotes the normal vector to the hyperplane and b is the bias value. C refers to the regularization parameter, also known as penalty factor, which is highly correlated with the tolerance of misclassification. The penalty factor is always greater than zero and the larger factor will create a hard margin, and vice versa (soft margin), its value needs to be determined carefully since hard margin may result in overfitting of the classifier. In our case, the penalty factor is set as one in the training of the classification model. ξ_i represents the slack variable related to the classification error, the SVM algorithm automatically allocates a slack variable for the feature points between the hyperplane and its margin, whereas the value of slack variable ($0 \leq \xi_i \leq 1$) is proportional to the distance of feature points to the hyperplane. In the circumstance that the feature points beyond the hyperplane (misclassification), the slack variable is larger than one.

If a linear hyperplane is not able to separate the feature points, the features can be mapped to a higher-dimensional space through a kernel function, where a linear boundary is available. The conventional kernel function includes higher order polynomial (quadratic, cubic) and Gaussian function, whereas the choice of the kernel function depends on the data distribution and the optimal hyperplane to separate them. SVM algorithm is suitable to implement on a multi-class problem by utilizing multiple binary classifiers via 'one vs one' approach, for instance, if there are N classes to distinguish, $N(N-1)/2$ times binary SVM will be computed to construct hyperplanes between each individual class.

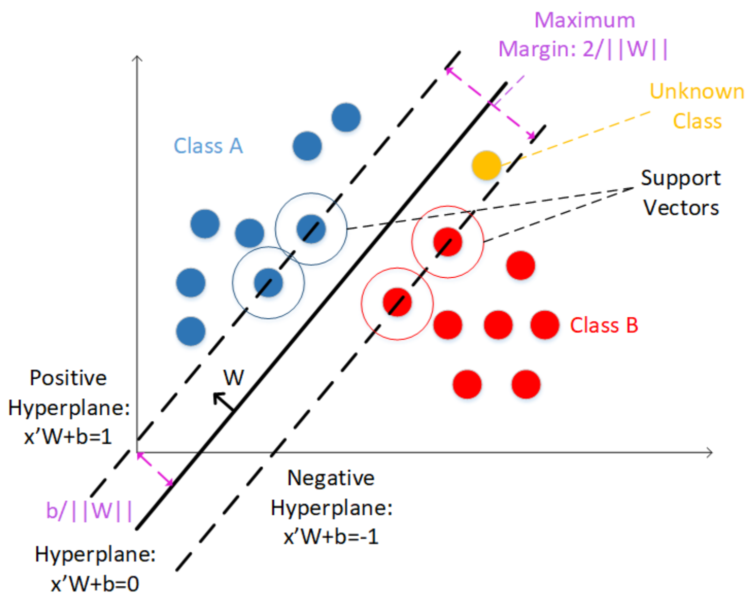


Figure 4.3 The hyperplane constructed by liner SVM to separate class A and B

4.2.2 K Nearest Neighbors

KNN (K Nearest Neighbors) [21], [47], [134] is a simple classifier with relatively low computational cost. Figure 4.4 illustrates the basic principle of KNN when the value of K equates to three, five and seven. Once the classifier is used to predict an unknown sample of the dataset, the majority class of K nearest points will determine the class of the unknown sample. For example, in figure 4.4, when K is equals to three, the unknown class is Class B because Class B takes the majority, whereas the unknown class is Class A when K is equals to seven for the same reason. Therefore, it is significant to select an appropriate K for different types of problems. Additionally, in order to avoid the decision clash, the value of K is better to be odd rather than an even number.

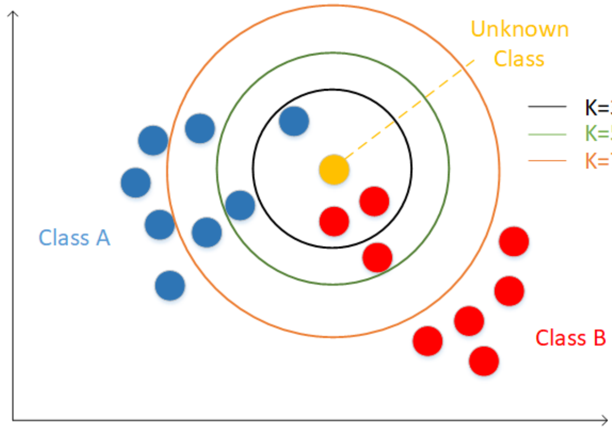


Figure 4.4 The classification boundary of KNN with K equates to 3, 5 and 7

4.2.3 Random Forest Decision Trees

Decision Tree (DT) [76] is a classic classification algorithm proposed by J. R. Quinlan, it contains root node (top node of one tree in figure 4.5), internal node (node can split) and leaf node (end node). Assume that the dataset contains M features, when the internal node starts splitting, one out of M features is chosen as the 'node feature' based on the 'node splitting' rules (e.g. information entropy, information entropy ratio or Gini impurity), the node will stop splitting in the condition that no feature could be used as the 'node feature' or the metric of node splitting is very low. Most common DT algorithms involve ID3 (Iterative Dichotomiser 3), C4.5 and CART (Classification and Regression Tree).

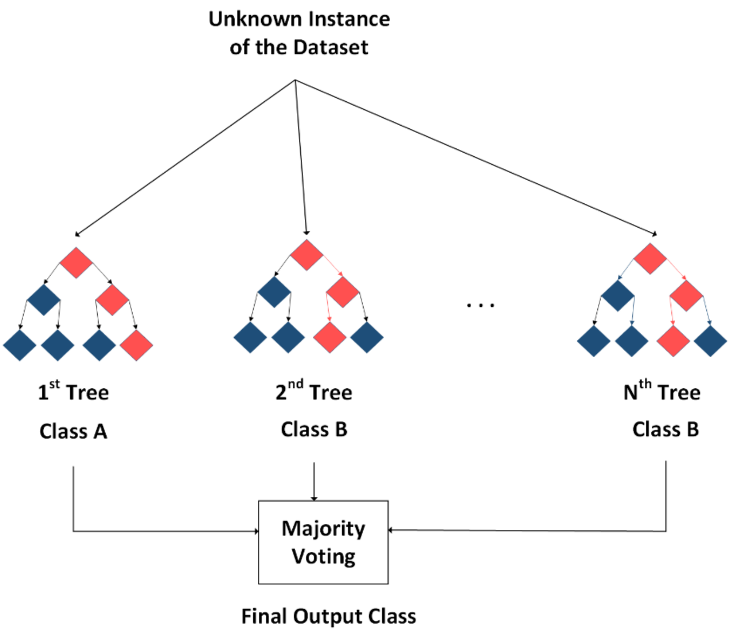


Figure 4.5 The structure of RFBT and its majority voting process

RFDT (Random Forest Decision Trees) [8], also referred as Random Forest Bagging, is considered as an ensemble of multiple decision trees, where each tree is independent. Prior to the classifier training, Bagging algorithm, also known as 'Bootstrap' is applied to construct the training set. It randomly selects one of the training samples to join the training set and this repeats N times, while at each time the selected sample needs to be repositioned to the training samples. Thus, a training set containing N samples is prepared. In our case, N equates to the number of trees. When the classifier is utilized to predict unknown sample, root node of each tree receives this unknown instance and then starts judging and classifying. Eventually, each tree generates a label for the new sample and the majority class among results of all the trees is output as the final class. RFDT is good at processing high dimensional dataset containing a large number of features and preliminary feature selection has been embedded in the algorithm, beyond that, it can be used to evaluate the importance of the individual feature and the mutual interference between pairs of features. For our work, CART is chosen as DT algorithm due to the fastest ‘node splitting’ rule (Gini impurity) and the number of decision trees in the RFDT is set as 200.

4.2.4 Artificial Neural Network

Artificial Neural Network (ANN) [49], [76] is a type of feedforward neural network using Multi-Layer Perceptron (MLP). It is comprised of one input layer, one or multiple hidden layers and then fully connected to one output layer as depicted in figure 4.6.

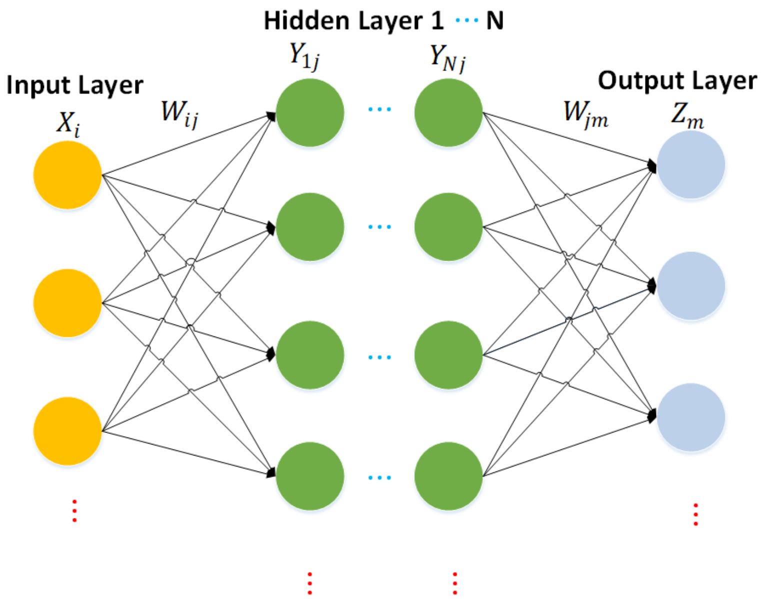


Figure 4.6 The ANN structure and network weight transfer between the layers

The dimension of the network inputs and outputs equates to the number of input features and number of classes respectively, whereas the number of neurons on the hidden layer is linear with the computational cost. The sigmoid hyperbolic-tangent activation function is utilized to multiply nonlinear components in this architecture while the initial weights and bias between each internal layer are randomly set from 0 to 1. To minimize the loss function of validation data in a feed-forward net, the backpropagation function called ‘scaled conjugate gradient’ (SCG) is introduced to update the weight and bias parameters within every iteration. Each connection between different stages is associated with a weight parameter and each neuron has an associated bias. The output of a generic layer is therefore defined as in equation 4.10, summing all his inputs multiplied by the relative weights and the bias.

$$Y_j = f(\sum_{n=1}^{i} W_{ij} X_i + b) \tag{4.10}$$

Where, W_{ij} is the weight of connection from node i on the input layer to node j on the hidden layer, b denotes the bias at every internal layer. As the number of samples is not big, we have used a shallow network design with only one to three hidden layers and a number of neurons up to 50.

4.2.5 Bi-directional Long Short-Term Memory Networks

LSTM (Long Short Term Memory) [140], [141] is a variant of RNN (Recurrent Neural Network), which is capable of learning backward long-term dependencies between time steps in the time-series data. It uses the internal state to characterize and store the potential correlations within the sequence of previous data, whereas the labels of new samples are predicted based on the previous knowledge. Therefore, LSTM is widely used in the classification task related to temporal sequence, for instance, continuous, unsegmented speech recognition. Initial LSTM was proposed by S. Hochreiter and J. Schmidhuber in 1997 [142], and after that the network was refined by lots of researchers. K. Cho et al. simplified the original architecture of a LSTM cell to a GRU (Gated Recurrent Unit) at 2004 and A. Graves et al. extended the standard LSTM to a bidirectional structure (Bi-LSTM) at 2005. In our work, a dual Bi-LSTM layers network is used to classify the continuous activities and gaits. Compared with original LSTM, Bi-LSTM [29], [143]–[145] can simultaneously search the backward and forward long-term dependencies from the data sequence. This allows the output states of the network to correlate with both previous and future information.

The proposed Bi-LSTM network involves an input layer, two Bi-LSTM layers and one output layer. The connections and weight transfer between each layers are shown in different colour arrows are shown in figure 4.7, while the information propagation between the gates of a Bi-LSTM cell are also sketched.

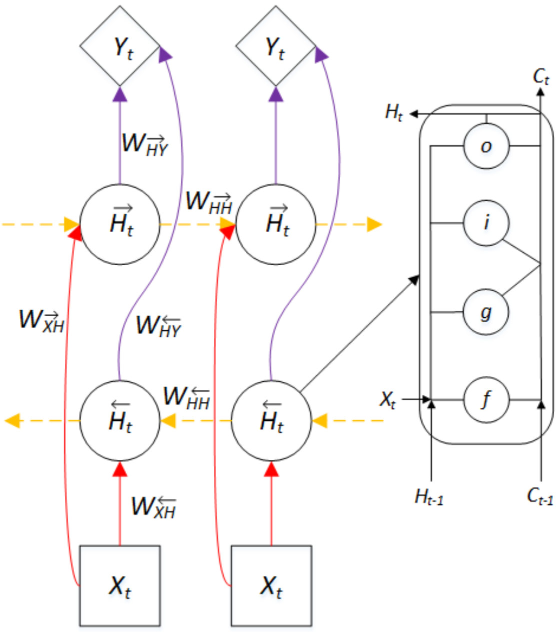


Figure 4.7 The structure of the Bi-LSTM layer and sketch of a single LSTM cell (***i***: input gate; ***f***: forgot gate; ***g***: cell candidate; ***o***: output gate)

The forward hidden state of a Bi-LSTM cell is governed by,

$$\overrightarrow{H_t} = \tanh(W_{X\overrightarrow{H}} X_t + W_{\overrightarrow{H}\overrightarrow{H}} \overrightarrow{H_{t+1}} + b_{\overrightarrow{H}}) \tag{4.11}$$

The backward hidden state of the cell is determined by:

$$\overleftarrow{H_t} = \tanh(W_{X\overleftarrow{H}} X_t + W_{\overleftarrow{H}\overleftarrow{H}} \overleftarrow{H_{t+1}} + b_{\overleftarrow{H}}) \tag{4.12}$$

Where *tanh* is the hyperbolic tangent activation function, X_t is the input state of the Bi-LSTM network, W_{ij} denotes the weight coefficient regarding to the states i and j, b_k represents the bias element of the state k.

The output state of the proposed network is obtained by adding the product of the weight and its corresponding hidden state with the bias element, notably,

$$Y_t = W_{\overrightarrow{H}Y} \overrightarrow{H_t} + W_{\overleftarrow{H}Y} \overleftarrow{H_t} + b_Y \tag{4.13}$$

The block diagram at the right hand side of figure 4.7 describes the information links among the gates with four different functions, those gates are used to process the input information inside one cell and their operational states are controlled by equations 4.14-4.17. The updating weights of one Bi-LSTM cell are divided into the input weights W, the recurrent weights R and the bias weights b.

- Input gate: $i_t = \sigma_{gate}(W_i X_t + R_i H_{t-1} + b_i)$ (4.14)

Input gate attempts to control the level of input in the cell state computation.

- Forgot gate $f_t = \sigma_{gate}(W_f X_t + R_f H_{t-1} + b_f)$ (4.15)

Forget gate aims to reset the cell state by forgetting the redundant information learned from previous time step.

- Cell candidate $g_t = \sigma_{state}(W_g X_t + R_g H_{t-1} + b_g)$ (4.16)

Cell candidate intends to generate new knowledge and add them to the current cell state.

- Output gate $o_t = \sigma_{gate}(W_o X_t + R_o H_{t-1} + b_o)$ (4.17)

Output gate decides the level of current cell state added to the output hidden state.

Where *i*, *f*, *g* and *o* refer to the input gate, forgot gate, cell candidate and output gate, the σ_{gate} and σ_{state} denotes the gate and the state activation function respectively. In our case, the state activation function is computed by the *tanh* function, while a sigmoid function is used as the gate activation function, given by below:

$$\sigma_{gate}(x) = \frac{1}{(e^{-x} + 1)} \tag{4.18}$$

The state of the Bi-LSTM cell at time bin *t* contains the knowledge learned from time bin *t*-1, it is calculated as following:

$$C_t = C_{t-1} \odot f_t + i_t \odot g_t \tag{4.19}$$

And the output hidden state is given by the product of output gate and cell state:

$$H_t = \sigma_{state}(c_t) \odot o_t \tag{4.20}$$

Where $\odot$ is the Hadamard product between two matrices with same dimension.

4.2.6 Transfer Learning using a Pre-trained Network

Transfer learning [146], [147] is an extension to the conventional concept of machine learning, it uses the information gained from an old task on a different but related new task. For instance, train the classifier by the images of cats and then apply it on the images of birds. VGG-16 is a deep neural network pre-trained on ImageNet, where ImageNet is a database containing more than 15 million high quality figures and approximately 22 thousands classes. Figure 4.8 shows the process of constructing a transfer net adaptive to the classification of radar images based on VGG-16. Transfer net utilizes the early convolutional and pooling layers of VGG-16 as the first stage, which allows it to 'inherit' the output weights related to the ability to find common features among the edges, curves and other properties of the image patterns. The final fully-connected layers of VGG-16 are replaced to several custom layers by the transfer net, this makes it capable of adapting to a new dataset through re-training with a small amount of the new labeled data, radar data in our case, and fine-tuning the original weights. For the other transfer nets (e.g. AlexNet, GoogleNet, ResNet), their network structures are different from VGG-16 in terms of number of layers and size of convolution filters, more details could be found in [148].

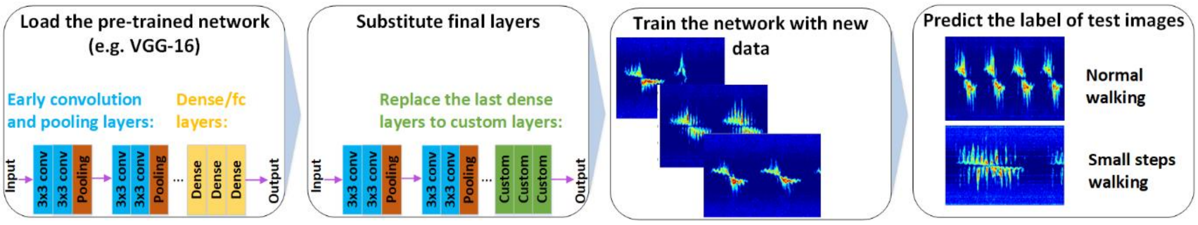


Figure 4.8 The adaptive processes of a pre-trained net (VGG-16 as an example)

Transfer learning has two significant advantages compared to the traditional deep neural network, in particular, the dataset for re-training is not necessary to be large, this solves the problems that image data is hard to acquire (e.g. iceberg) and as a result of using a pre-trained net, it saves a lot of computational power and training time.

4.2.7 Metrics for Classification Performance

General metrics [66], [94] related to the classification performance of a specific class are shown above in equation 4.21 to 4.25, and those metrics are simplified on a binary classification problem represented in table 4.3, where the row and column denote output and target classes separately. Sensitivity (equation 4.21), also referred as Recall, measures the correctly classified rate for the class of interest (A in our case). On the contrary, Specificity (equation 4.22) is correlated with the number of opposite class that are recognized successfully. Besides that, Precision (equation 4.23), also known as Positive Predictive Value (PPV), is the number of true positives divided by the total number of samples belonging to the class of interest.

Output\Target	**A**	**B**
A	True Positive	False Positive
B	False Negative	True Negative

Table 4.3 Binary confusion matrix

$$Sensitivity = \frac{TP}{TP + FN} \tag{4.21}$$

$$Specificity = \frac{TN}{TP + TN} \tag{4.22}$$

$$Precision = \frac{TP}{TP + FP} \tag{4.23}$$

Furthermore, another two metrics are introduced to evaluate the overall performance of the classifier. The F-measure (equation 4.24) utilizes the harmonic mean of both sensitivity and precision to show the overall 'missing positives' and 'false alarms', whereas the classification accuracy (equation 4.25) is the correctly classified rate over all the classes, and it is usually served as the most significant feature for evaluating a classifier due to the balance. Beyond that, the classification accuracy turns into sensitivity if it is applied on the single class. In the rest of the thesis, those metrics have been re-calculated considering the multi-class effect.

$$F - measure = \frac{2 * (Precision * Sensitivity)}{Precision + Sensitivity} \tag{4.24}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.25}$$

4.3 Data Fusion Approaches

4.3.1 Signal Level Fusion

Signal level fusion methods mainly include:

- **Weighted Average (WA)** [19] is the most represenative signal fusion model for combining the information through averaging all the sensor measurements. The influence of the largest measurement error will be signficantly reduced after averaging, however, instead of considering single sensor equally, each sensor could be assigned with a weight index based on its contribution to the classification accuracy. For instance, the weighted average of accelerzation readings combines the information content from multiple wearable accelerometers at different postions of the human body (e.g. ankle, wrist, waist) , it can be used to better estimate the accelerzation related to human motions.
- **Kalman Filter (KF)** [102], [149] is a well-known approach to fuse various sensor sources for the optimal estimation of the parameter of interest. It consists of three processes, namely, prediction, measurement and update. For instance, to better estimate the position of a car, firstly the position of a car at time step t+1 is predicted based on the previous knowledge of its original position and kinematic state, secondly find the GPS readings at time t+1, finally the previous knowledge is updated by comparing the GPS readings (measurement) and the prediction. By

running the algorithm exhaustively, the variance between the actual postion of the car and prediction will be decreased to a very low level. KF is tyically used to combine accelerometer and gyroscope readings to improve the estimation accuracy of a navigation system.

- **Trilateration**: We proposed a novel signal fusion method based on trilateration algorithm, which combines the range readings from three radar sensors at different positions and turns into the real-time localization of the testing subject.

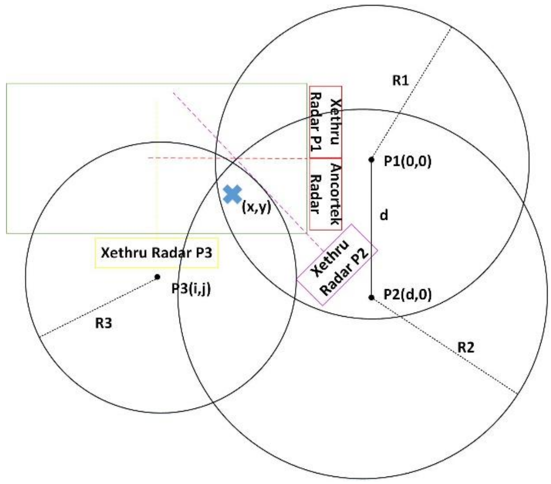


Figure 4.9 Concept figure of trilateration algorithm: Xethru P1(0, 0); Xethru P2 (d, 0); Xethru P3 (i, j)); R1(distance from Xethru P1 to participant); R2(distance from Xethru P2 to participant); R3(distance from Xethru P3 to participant)

Figure 4.9 above illurstrates the geometry of the radar sensors with respect to the subject and their range measurments (R_1, R_2 and R_3), where X and Y denote the position of the subject in terms of hortional and vertical coordinates. R_1, R_2 and R_3 can be represented as a function of X and Y:

$$R_1^2 = X^2 + Y^2 \tag{4.26}$$

$$R_2^2 = (X-d)^2 + Y^2 \tag{4.27}$$

$$R_3^2 = (X-i)^2 + (Y-j)^2 \tag{4.28}$$

where X and Y could be derived as following:

$$X = (R_1^2 - R_2^2 + d^2)/2d \tag{4.29}$$

$$Y = (R_1^2 - R_3^2 - x^2 + (x-i)^2 + j^2)/2j \tag{4.30}$$

It should be noticed that R_1, R_2 and R_3 are also correlated with Z (height) in a 3-D space, however, in our case, we only interest in X and Y as the difference between each gait styles is not too much relevant to height. Hence Z is set to 0 and automatically ignored in equations 4.26 to 4.30. As shown in figure 4.10, each radar sensor has a range resolution ΔR, thus the target location would be between measurement plus ΔR and measurement minus ΔR, since we don't consider angular resolution, the target could appear at anywhere within the radar beamwidth (65 degrees in our case). By using the range information of two radar sensors, namely, Xethru P1 and P2 (UWB radar in front of the participants and on the ceiling), the target location can be narrowed to one small area (marked in black dash line), whereas this small area can be subsequently narrowed by using trilateration (marked in green solid line). Compared to using range measurements of single radar, trilateration-based signal level fusion algorithm significantly increases the precision of localization, which is beneficial to the following training and testing of the proposed Bi-LSTM network.

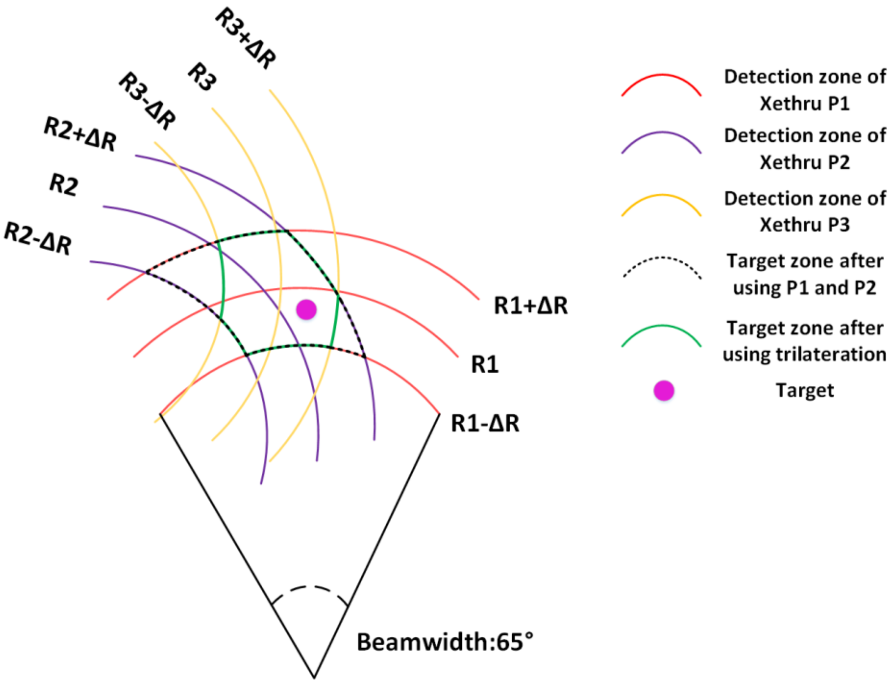


Figure 4.10 The geometry of trilateration

4.3.2 Feature Level Fusion

$$F_{Sensor} = \begin{bmatrix} F_{11} & F_{12} & F_{12} & \cdots & F_{1j} \\ F_{21} & F_{22} & F_{23} & \cdots & F_{2j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ F_{i1} & F_{i2} & F_{i3} & \cdots & F_{ij} \end{bmatrix} \tag{4.31}$$

The feature matrix of an individual sensor is shown in equation 4.31, where i denotes the index of data sample collected by that sensor and j denotes the number of feature vectors. In the procedure of feature fusion, a 'wider' matrix is constructed by concatenating the matrix of each sensor horizontally, as shwon in eqaution 4.32.

$$F_{Fusion} = [F_{S_1} \quad F_{S_2} \quad F_{S_3} \quad \cdots \quad F_{S_N}] \tag{4.32}$$

Where N is the number of sensors involved in the fusion.

The **feature selection process** aims to reduce the computational intensity and achieve subsequent performance improvement through removing the redundant or high relevance features with negative/repetitive information. Those methods involves filter-based methods, wrapper methods and embedded methods.

- **Filter-based methods** [32], [109], [150] calculate the statistical moments of features (e.g. entropy, correlation, Euclidean distance) as a metric to sort them. Filter-based methods can be used without the engagement of classifiers.
- **Wrapper methods** [29], [109], [118] use a particular classifier to evaluate the feature combinations based on the classification performance and the feature subset yielding the highest accuracy is chosen as the output features. Wrapper methods are usually computation intensive especially running on a high dimensional feature set.
- **Embedded methods** [151] combine the feature selection with classification through a feedback loop, during the classifier training, the algorithm is able to compute the weight factor for each feature and rank them accordingly. Most representative embedded method is SVM-RFE (Support Vector Machine-Recursive Feature Elimination).

In the following part, two filter-based methods, notably, Fisher score (F-score), Relief-F, and two wrapper methods: SBS (Sequential Backward Selection) and SFS (Sequential Forward Selection) are introduced. Fisher score sorts the feature set according to two abilities, namely, clustering the samples with same class by a minimal spatial distance (L2 norm) and separating the samples with different classes by a maximal spatial distance. The F-score of feature X_i is derived below:

$$F(X_i) = \frac{\sum_{j=1}^{c} m^j (\mu_i^j - \mu_i)^2}{\sum_{j=1}^{c} m^j (\sigma_i^j)^2} \tag{4.33}$$

Where i denotes the index of features and j denotes the index of classes, c is the total number of classes in the task. m^j represents the number of samples with class j, μ_i is the mean value of i^{th} feature among all the samples, μ_i^j and σ_i^j indicate the mean and standard deviation of the i^{th} feature with j^{th} class, respectively. The Relief-F was proposed by I. Kononenko et al. to overcome the limitation of original Relief algorithm especially for the multi-class problems. Similar to the F-score, Relief-F uses the distance measurement (L1 norm) as a metric to grant each feature a normalized weight between -1 and 1. In the feature space, one specific feature vector is selected from a random sample of the dataset, whereas the feature weight is generated based on its proximity to feature vectors of other samples. The N closest feature vectors with same class and with different classes are denoted as 'N nearest hits' and 'N nearest misses', respectively, and their contributions to the feature weight are averaged in Relief-F to increase the robustness of the algorithm. The algorithm will run exhaustively for *K* times, where *K* equates to the number of samples in our case.

The weight update of the selected feature at iteration *k* ($k \leq K$) is shown as equation 4.34 below:

$$w_k = w_k - \sum_{p=1}^{N} \frac{\left| f_k - H_k^p \right|}{N} + \sum_{C \neq S} \frac{P(C)}{1 - P(S)} \sum_{p=1}^{N} \frac{\left| f_k - M_k^p \right|}{N} \tag{4.34}$$

Where f_k denotes to the value of the selected feature at iteration *k*, H_p^k and M_p^k refer to the value of the p^{th} nearest hit and the p^{th} nearest miss at same iteration. $P(C)$ is the ratio of class C ($C \neq S$) in the dataset, whereas S is the class of the selected sample and $P(S)$ denotes the ratio of Class S. Those probabilities are considered to mitigate the effect of class imbalance and missing features. N is the total number of the nearest hits/misses, in our case, N equates to 10.

Sequential feature selection searches the best combination of features rather than separating them and scoring, it uses a classifier (e.g. SVM, KNN, and RFBT) and the classification accuracy is utilized as a basis to sort the features. This technique can be implemented at two directions, notably, sequential forward selection (SFS) by keep adding features from an empty feature set until the classification performance stops increasing and sequential backward selection (SBS) by progressively removing features from a full feature set. SFS takes more time to finish since it is hard for the data fitting of classifier in the situation that number of features is very low. The algorithm of SFS is given below:

Algorithm 1: Sequential Forward Selection

```
Input: Original feature matrix Z (size: n X m, n: number of samples; m: number of features)
Output feature matrix T←0;  %%initialize the output matrix with an empty set
k←m;                          %%Initialize k with the value of m
for j = 1,…m do
    for i = 1,…,k do
        T'←cat(T, Z(j));      %% constructing a new matrix T' by concatenating jth feature from Z
        A(i) ←classifier (T');  %% test T' with a classifier, store the accuracy into a new variable
        T'←T;                 %% clear T'
    end
    if max(A) > H(j-1)
      break;                  %% until the accuracy stops increasing
    end
    H(j)←max(A);              %% store the previous maximum accuracy in a variable
    choose the feature vector F from Z with max(A);   %% best feature from one iteration
```

```
    T←cat(T, F);              %% update T by concatenating the best feature
    k←k-1;                    %% the number of available features-1
    A←0;                      %% clear the accuracy
end
```

Output: Output feature matrix T

In the case of backward selection, the output feature matrix starts with original feature matrix and its content is updated by dropping the worst feature iteratively.

4.3.3 Decision Level Fusion

- **Soft fusion**

The confidence level of the classifier is a probability matrix with its size eqaul to *n*X*m*, *n* is the number of samples and *m* is the number of classes. It is used to measure the certainty of classifier decision making, whereas for each sample, the class yielding the highest confidence level will be chosen as the output class. The value of the confidence level is converted from the unnormlized classifier output through a softmax function, as shown below:

$$P_c = \frac{e_c}{\sum_{k=1}^{K} e_k} \tag{4.35}$$

Where class c is the class of interest, P_c the confidence level of class c, e_c and e_k denote the unnormlized classifier output of class c and class k ($k \leq K$) , repsectively, K is the number of classes.

Decision level fusion based on confidence level is known as soft fusion, which mainly includes linear soft fusion (e.g. equal weight soft fusion, weighted soft fusion) and nonlinear soft fusion (e.g. LOGP [105], fuzzy logic [32], [101]). Equal weight soft fusion considers the contribution of each classifier as equally, it is basically a linear, accumulative sum of the confidence level matrices, whereas the weighted soft fusion allocates one unique weight index for each classifier based on the performance and it can be derived as below:

$$P_{Fusion}(s,c) = \sum_{n=1}^{N} W_n \cdot P_n(s,c) \tag{4.36}$$

Where P_n and W_n denote the confidence level matrix and weight index for classifier n respectively, P_{Fusion} is the confidence level matrix after fusion. s refers the index of samples and c indicates the class number.

On the contrary, nonlinear soft fusion intends to add some nonlinear content to the fusion process, for instance, LOGP (Logarithmic Opinion Pool) algorithm introduces an exponential function to combine the confidence level and summarized in equation 4.37 below:

$$P_{Fusion}(s,c) = \prod_{n=1}^{N} e^{-P_n(s,c)^d} \tag{4.37}$$

Where $P_n(s, c)$ is the confidence level of sample s, class c. n denotes the index of classifier and N is the total number of classifier participating in the fusion, d is a distribution factor equal to $1/N$. The fusion probability matrix is constructed by multiplying the exponential confidence level of different classifiers, and the output will be the class with the highest fusion probability. The second nonlinear approach uses Fuzzy logic, where the confidence level matrix of each classifier is utilized as a Fuzzy set as depicted in equation 4.38:

$$P_{Fusion}(s) = \max\{\min\{P_1(s)\}, \min\{P_2(s)\}, \ldots \min\{P_n(s)\}\} \tag{4.38}$$

Where $P_1(s)$ to $P_n(s)$ denote the confidence level vector for sample s with respect to classifier 1 to n, the final fusion probability of sample s is equal to the maximum out of the minimal confidence level among all the classifier. After that, the class yielding this maximum value is chosen as the output class. In other words, the Fuzzy logic is opposite to other methods in terms of selecting the least errors from the worst cases.

- **Hard Fusion**

Hard fusion uses the prediction results of classifiers, typical hard fusion approaches include Majority Voting (MV) [32], [92], [132], Weighted Majority Voting (WMV) [101], Recall Combiner (RC) [91], [152] and Naïve Bayes Combiner (NBC) [91], [152]. MV is the most representative hard fusion methods and it takes the majority class of the classification results as the output label. The number of classifiers is better to be odd than even, otherwise an additional function needs to be embedded into the voting machine for solving the possible decision clash. WMV assigns a weight index to each classifier depending on its contribution to the fusion, the classifier with better performance is associated with higher weight. Recall Combiner (RC) is a type of optimal combiners proposed by [152], it uses the recall/sensitivity of the class of interest from the confusion matrix and the rest of classes are treated as a 'joint class'. Hence, the misclassification could be considered to distribute equally among the rest of classes. The formula of RC can be derived as equation 4.39.

$$P(C_k \mid d) = P(C_k) \cdot \prod_{m \in M_+^k} p_{m,k} \cdot \prod_{m \in M_-^k} \frac{1 - p_{m,k}}{C-1} \tag{4.39}$$

Where $P(C_k \mid d)$ denotes the possibility of C_k (class of interest) being chosen as the output class out of d and d is a class set containing all the available classes. In the RC algorithm, N different classifiers are stored in a classifier ensemble M, m is the index of classifier and the total number of classes to distinguish is equal to C, whereas the classifier ensemble is grouped into M_+^k and M_-^k, which represents the classifier group that supporting class C_k and other classes, respectively. In the case that the classifier supports class C_k, $p_{m,k}$ refers to the confusion matrix element (row and column k) of classifier m in the ensemble, oppositely, $\frac{1-p_{m,k}}{C-1}$ denotes the misclassification shared by C-1 classes.

The performance of RC is limited by the fact that the classification errors are divided equally to the classes of non-interest, however, in the real case scenarios, the misclassification is a variable for each class and this indicates that the contribution of each class to the total misclassification is not the same. Therefore, the performance of RC is highly relevant with the number of classes and the number of classifier participating in the fusion. NBC is introduced to address the problem that RC has poor performance with a small number of classifiers. It replaces the equal division of classification errors to a real misclassification probability between the class of interest and the classifier output class, the possibility of class C_k being the final output is derived as equation 4.40.

$$P(C_k \mid d) = P(C_k) \cdot \prod_{m=1}^{N} p_{m,R_m,k} \tag{4.40}$$

R_m is the prediction result of classifer m in the enseblem, whereas the output probablilty is a fucntion of the classifer supporting rate and the confusion matrix component (classifier m, row R_m and column k). However, the gain of the NBC is not the same level as RC in terms of the augmentation of classifier numbers. Beyond that, NBC has less compatibility with data involving high noise content and the computational cost regarding the volume of parameters per sample for RC (N*C+N) is much lower than NBC (N*C^2+C).

4.4 Summary of the Chapter

In this chapter, all the data processing procedures related to machine learning and artificial intelligence used in this thesis are presented, involving the introduction and derivation of both radar and wearable IMU handcrafted features, the discussion and comparison of machine learning-based classification algorithms ranging from conventional classifiers to deep neural networks, and the interpretation of data fusion methods from three different perspectives (signal, feature and decision).

5 Individual Human Activity Recognition and Fall Detection

This chapter analyses the “snapshot” (individual, time-limited) activities. The classification performance of wearable IMU sensor and FMCW radar fusion as well as magnetic sensor and FMCW radar fusion are evaluated and compared with single sensor case. The information fusion methods include feature fusion, linear and nonlinear soft decision fusion and a novel voting system, whereas the classifiers include SVM, KNN and ANN and feature selection methods comprise Fisher score, Relief-F and SFS.

5.1 Sensor Fusion with Inertial Sensors and Radar

This section utilizes two types of sensing technologies, in particular, an IMU within the smartphone and a FMCW radar sensor, to distinguish daily human activities and identify the critical events like falls. Information fusion is implemented to combine the advantages of two sensors and the gain of distinct fusion schemes are compared, where feature level fusion indicates an accuracy improvement of 12% with both SVM and KNN classifiers. Meanwhile, a decision fusion method based on the voting machine is proposed to combine the confidence level of classifiers and the classification accuracy raises to 97.4%. Beyond that, different feature selection methods, namely, F-score, Relief-F and Sequential Forward Selection (SFS) are evaluated in terms of subsequent accuracy improvement and computational cost.

5.1.1 Experimental Setup

In this section, an FMCW radar (Ancortek 580B) operating at 5.8 GHz and a nine Degrees of Freedom (DOF) IMU within a Huawei Honour 6 smartphone were simultaneously used to collect the data of human motions from 9 subjects, where the IMU module is comprised of a triple axial accelerometer, gyroscope and magnetometer, and it is able to measure the acceleration, angular speed and magnetic field strength relative to the human activities at roughly 100 samples/s. The FMCW radar has 400 MHz instantaneous bandwidth and the Pulse Repetition Frequency (PRF) equates to 1000 Hz. The radar transmitted power is approximately 18 dBm, whereas the maximal gain of the transmitting and receiving Yagi antenna is about 17 dB. Additionally, the -3dB beam-width of the transmitting antenna is about 60 degrees and vertical polarization is used for the activity measurements.

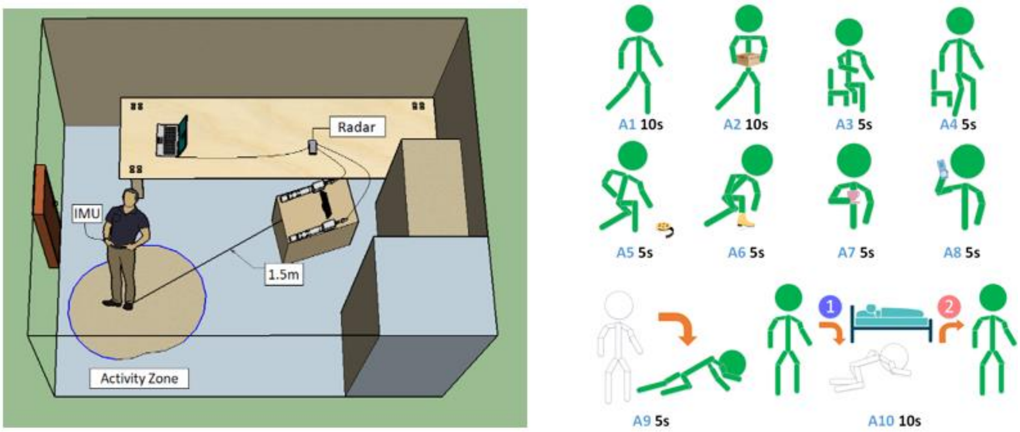


Figure 5.1 The experimental setup (left) and list of the activities (right)

The data was collected in the laboratory (room size about 2.7m x 3.5m) of CSI (Communication, Sensing and Imaging) group at the University of Glasgow and figure 5.1 sketches a 3-D vision of the measurement environment. The antennas were placed in parallel on a wooden box at approximately 0.8 m height, facing to the activity zone where the participant was required to perform different human motions. The transmitter and receiver were separated with a distance equal to 0.3 m (quasi-monostatic configuration), and their range to the central of the activity zone was roughly 1.5 m. The radar system was connected to a laptop via a USB cable for data transferring and powering, whereas the smartphone was fixed on the wrist of the participant's dominant hand through a flexible strap while starting the experiments. Additionally, a 5s time delay is set after triggering the data recording of the smartphone to make sure that the user has enough time to operate the radar system through a MATLAB script.

A1	Walking back and forth
A2	Walking while carrying an object with both hands
A3	Sitting on a chair and still
A4	Standing from a chair and still
A5	Bending down to pick up an object
A6	Bending and staying down to tie shoelaces
A7	Drinking a glass of water while standing
A8	Picking up a phone call while st;p64anding
A9	Simulating a frontal fall on the mattress
A10	Checking under an imaginary bed and coming back up

Table 5.1 List of the activities

The participants are aged from 23-31 with diversity in terms of gender (2 female, 7 male), height (1.6 to 1.8 m), body shape and dominant hand. The recorded activities are illustrated in the right part of figure 5.1 and summarized in table 5.1, where some of the activities are designed to be similar in pairs (e.g. 'A1' and 'A2', 'A3' and 'A5', 'A7' and 'A8', 'A9' and 'A10). Those similar activities are deliberately added to create more classification challenge, for instance, 'A3' and 'A5', both yielding strong acceleration changes towards the ground. Beyond that, fall events are particularly crucial to be recognized correctly, with both high sensitivity and low false alarms. The data collection mode is 'snapshot', therefore, single activity is recorded each time without transition to others, whereas the duration of data recording are 5s and 10s for the short and long activities respectively. For each participant, three repetitions of each activity are taken and this generates a database containing 270 samples (number of participants* number of activities* number of repetitions).

5.1.2 Classification Results of using Inertial Sensors and Radar Individually

Prior to the classification, for the FMCW radar, 28 numerical features were extracted (ID: 1-2, 5-6, 11, 13, 14-16, 24-35, 38, 60-62, and 70-72 in table 4.1) as suggested by [133], whereas for the IMU case, totally 177 features were calculated from the pre-processed signals: 63 for accelerometer (all the features in table 4.2 except for the Norm XYZ), 57 for gyroscope and 57 for magnetometer. Those features extracted for each inertial sensor are the same apart from the skewness and kurtosis, which are positive to the accelerometer and redundant to other two sensors.

For the classification, a SVM classifier with a quadratic kernel and a KNN classifier with the number of nearest neighbours equal to 9 were trained to recognize different human activities listed in table 5.1. The cross-validation method is ‘K-fold’ for the classifier training and testing, which can be described as steps following:

- The dataset is randomly divided into K different parts (folds).
- Nine folds are used to train the classifier, while data in the rest fold is used as a test set to validate the classification model.
- This training and testing scheme will run K times until each fold has been tested upon.
- The final classification accuracy is the mean value of all K iterations.

Where in our case, K is chosen to be 10 as suggested by [37]. 90% of the dataset is used to train the classifier and the rest 10% is used to test the classifier performance.

Table 5.2 shows the classification accuracy of each single sensor and feature fusion results between all the inertial sensors, where accelerometer and gyroscope yield similar results, and their performance outperform than other single sensors. For magnetometer and radar, the accuracy are roughly 4% and 6.1% lower with SVM, whereas the accuracy drop is nearly 10% in the case of using KNN. Beyond that, compared to the best single sensor case, the internal fusion between inertial sensors improves the classification accuracy with approximately 4.1% and 15.6% for SVM and KNN.

Classification Accuracy (%)	**SVM**	**KNN**
Accelerometer	85.2	79.6
Gyroscope	84.1	79.6
Magnetometer	80.4	69.6
Inertial Sensor Fusion	89.3	85.2
Radar	77.9	70.7

Table 5.2 The classification accuracy for individual sensor and inertial sensor fusion

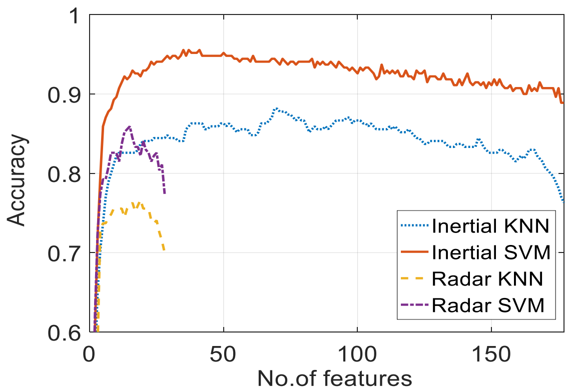


Figure 5.2 SFS classification accuracy for inertial sensors (long curve) and radar (short curve) with different classifiers

Inertial Sensors

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

Radar

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	15
SFS(KNN)	79.8	32	19

Table 5.3 The accuracy, time cost and number of selected features through different feature selection methods

Figure 5.2 compares the SFS results of inertial sensor fusion (concatenating the features from accelerometer, gyroscope and magnetometer) and radar. The highest accuracy reaches 95.6% for inertial sensors fusion using SVM, whereas the single radar yields 85.6%. It is reported that switching to a KNN classifier will lead an accuracy drop between 6 and 7%. Table 5.3 summarizes the results from different feature selection methods, in particular, F-score, Relief-F and SFS, for inertial sensor fusion and radar. The accuracy gain of F-score and Relief-F for both inertial sensor fusion and radar are relatively small, whereas using the wrapper method SFS with SVM yields subsequent improvement of 5-7% in classification accuracy for both inertial sensor fusion and radar. In the case of KNN, the enhancement for the inertial sensor fusion is approximately 3% while radar has a significant accuracy boost of 9%. In the best case scenario, filter-based method reduce the dimension of feature matrix by 60% and 35% for inertial sensor fusion and radar. However, the optimal features suggested by SFS are only 20% of the overall features for inertial sensor fusion. The computational cost is proportional to the number of features involved and the time taken is generally shorter for filter-based methods, especially used with a simple classifier like KNN, whereas the computational load for SFS using a SVM classifier is several times greater than F-score and Relief-F under same condition. In conclude, in the condition that computation power is not limited, SFS with SVM is the best choice for both sensors, otherwise choosing Relief-F with KNN for inertial sensor and F-score with SVM for radar.

5.1.3 Classification Results of Information Fusion

- **Feature level fusion**

For the feature level fusion, the feature sets from radar and inertial sensors are cascaded before using them to train a classifier. SFS is chosen as the optimal feature selection method based on its ability of boosting accuracy and an overview of the SFS results with a SVM classifier are presented in figure 5.3. The feature fusion yields the maximal accuracy of 97.4% when 31 out of 205 features are used, it improves the approximately 12% and 2% in classification accuracy with respect to the highest points of radar and inertial sensors. For the optimal feature combinations, inertial sensors contribute correlation-based features, particularly cross-correlation and spectral features, whereas for the radar sensor, physical features including centroid, bandwidth and energy curves and SVD-based features are selected.

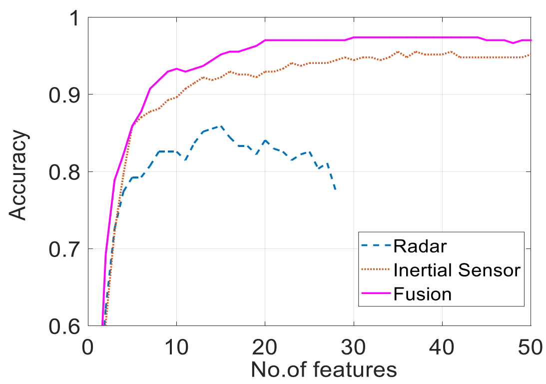


Figure 5.3 SFS classification accuracy for individual sensor and feature fusion

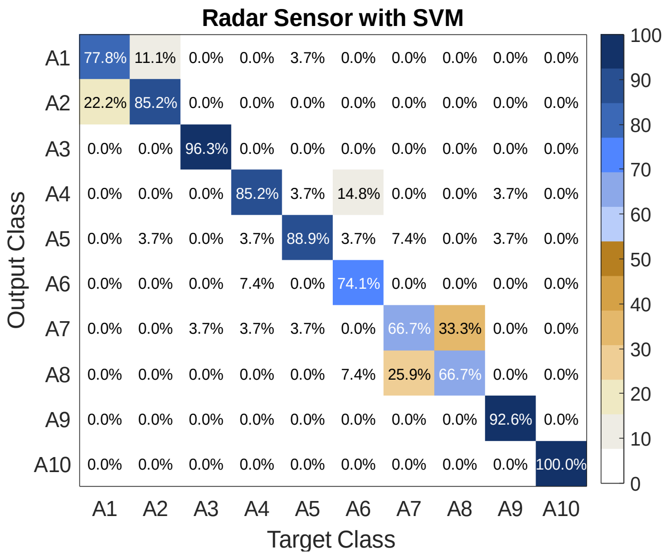


Figure 5.4 Confusion matrix for optimal feature combination selected from the radar feature set

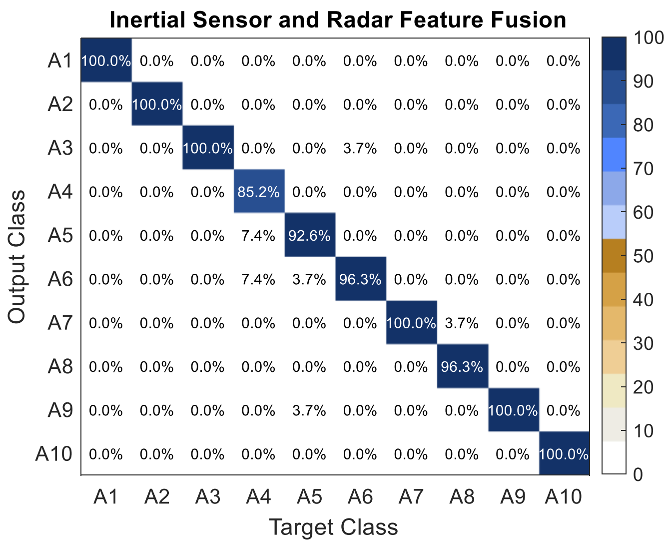


Figure 5.5 Confusion matrix for optimal feature combination selected from the fusion feature set

Figure 5.4 shows the classification results of using FMCW radar only, where the row and column of the confusion matrix denotes for the target and output class respectively. The sum of column elements should be equal to 100%, the diagonal elements represent the sensitivity/classification accuracy for each class, whereas the non-diagonal elements yield the misclassification rate across two different classes. It is obvious that there are lots of misclassifications between A1 'walking' and A2 'walking with an object', A4 'standing up' and A6 'picking up an object', A7 'drinking water' and A8 'taking a phone call', espcailly 'A7' and 'A8', this is due to the limitation of radar range resolution (about 30cm in our case and this number is not enough to recognize the difference between drinking water and phone call). For the most ciritical 'A9' falling, the radar indicates moderate performance with around 7.4% falls misclassified to other acitivities. Figure 5.5 summarizes the correctly classified events and misclassifications of feature fusion between inertial sensors and radar. Compared to radar-only results in figure 5.4, most of the activities are identified accurately, especially for A1 'walking', A2 'walking with an object', A3 'sitting down', A7 'drinking water', A9 'simulating a fall' and A10 'checking under the bed'. However, there are several classification errors between A4 'standing up', A5 'picking up an object', A6 'bending and tying shoelaces'. Apart from that, one sample of A5 has been misclassified to A9, leading to a false alarm in fall detection. It is very interesting that the classification accuracy of some classes are very high, even close to perfect. However, since the training and testing dataset are relatively small (totally 270 samples), the classification results may have some coincidence. Thus the proposed fusion method needs to be validated through a dataset including more subjects, and that will be discussed in details in future work.

- **Decision level fusion**

In this part, two soft fusion and one hard fusion approaches, notably LOGP, Fuzzy logic and voting system are utilized to combine the classification results of inertial sensors and radar. The voting system is inspired by F. Fioranelli's work [20], where majority voting is used to combine the results of multiple radar nodes. Differ from their work, our proposed voting system is based on the prediction labels of SVM and KNN classifier with respect to inertial sensors and radar. Figure 5.6 depicts the whole voting procedure, when there is no decision clash, the majority of classifier outputs is used as the new prediction. However, in the circumstances that two classes are supported by same votes (2 vs 2 decision clash), the confidence level of SVM classifier regarding inertial sensors and radar are fused by LOGP algorithm to create the final prediction.

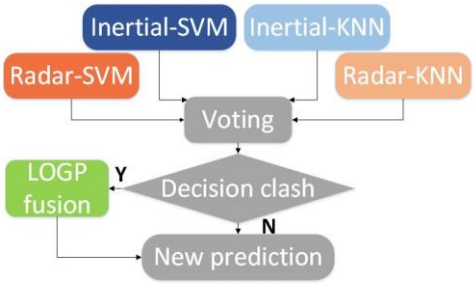


Figure 5.6 Block diagram of the proposed voting system

Method	Average error*
LOGP	9
Fuzzy logic	14
Voting system	6

Table 5.4 The average number of errors through different fusion approaches

As table 5.4 shows, the average number of misclassification events and average classification accuracy are compared for the three decision fusion methods, where the voting system-based method outperforms than others. Average error over the ’10-fold’ iterations is used as a metric to quantify the accuracy difference, compared to LOGP and Fuzzy logic, voting system-based method decreases the average classification errors by 3 and 8 respectively.

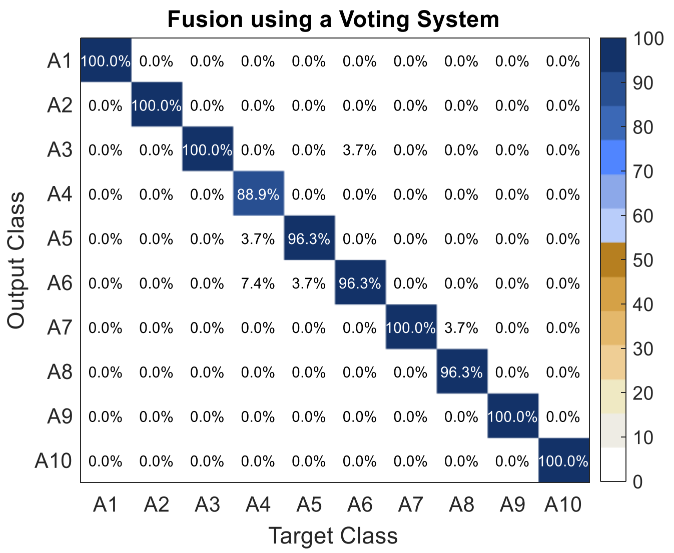


Figure 5.7 Confusion matrix for voting system

The summary of classification results based on the voting system are shown in figure 5.7. Compared to the feature level fusion in figure 5.5, voting system removes the false alarms for the most crucial fall events and the sensitivity of A4 'standing up' increases approximately 3.7%.

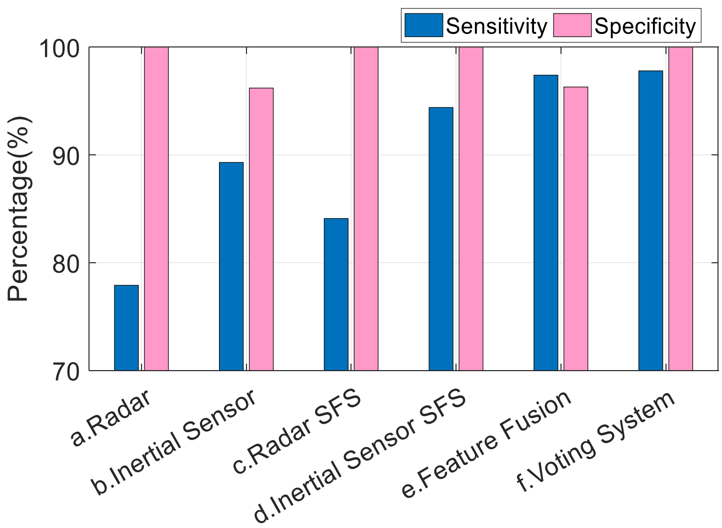


Figure 5.8 Sensitivity and specificity for different approaches

The overall sensitivity across the 10 human activities and specificity for 'simulating a fall' are compared in figure 5.8 for all the aforementioned approaches. Significant improvement can be found in the sensitivity of activity recognition after using the optimal feature set generated by SFS, beyond that, the feature fusion between inertial sensors and radar provides subsequent accuracy gain, whereas the fall specificity drops roughly 3.7%. This is due to that sequential feature selection only picks the feature combinations yielding the best overall performance (feature fusion yields better average sensitivity than radar SFS and inertial sensor SFS), whereas the selected feature combinations of sensor fusion may not be as optimal as selected feature combinations of a single sensor in detecting falls. The proposed voting system-based method fixes the problem and slightly improves the average sensitivity on top of feature fusion.

5.2 Sensor Fusion with Magnetic Sensor and Radar

In this section, we expand our work in previous section to a larger dataset containing 20 participants (600 samples totally, 60 samples per class). Typically, magnetic sensor is jointly utilized with accelerometer and gyroscope in the field of activity recognition, or ignored in the scenario that using data from those two inertial sensors can satisfy the requirements. From previous section, it is reported that radar and magnetic sensor yield similar performance in the activity classification, whereas the accelerometer and gyroscope outperform a lot by using individually. Therefore, magnetic sensor is picked out to apply the information fusion with the FMCW radar sensor. Statistical features are extracted from both sensors and SFS, which has been proved as the best feature selection method in terms of accuracy boost, is used to pick the most relevant features. A SVM classifier with quadratic kernel and an ANN with multi-hidden layers are employed to test the sensor fusion performance via a more challenging cross-validation method called ‘leaving one participant out’ (L1O). Using fusion along with SFS, almost 96% classification accuracy is achieved by both classifiers, whereas the testing of samples from unknown participant yields an accuracy of nearly 93% in the best case.

5.2.1 Experimental Setup

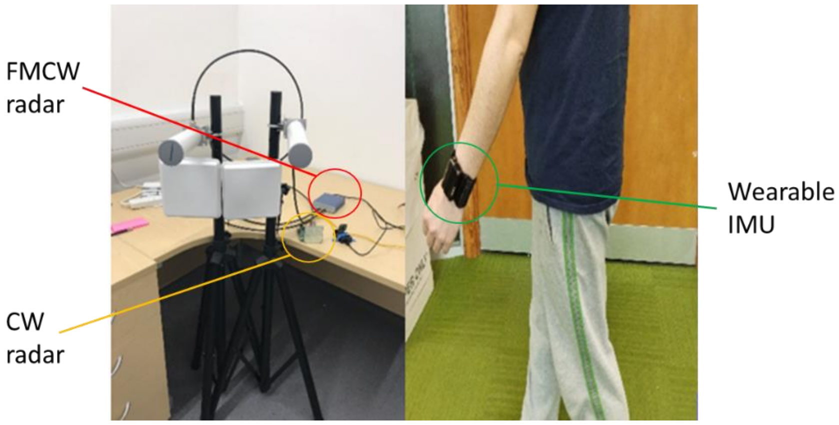


Figure 5.9 Radar setup (left): FMCW radar (red circle), CW radar (orange circle); the wrist-worn IMU (green circle) on the dominant hand of the participant (right)

The experimental area and activity list in this section are the same as previous data collection (shown in figure 5.9), whereas a commercial wearable IMU from X-IO technologies is used to replace the smart phone due to its small size and high resolution. The Ancortek FMCW radar and an additional CW radar produced by White Horse are also included in the measurement of human activities. New database contains 600 simultaneous samples for each sensor, where the number of participants extends to 20 and each of activity in table 5.1 is repeated by three times for everyone.

The FMCW radar operates at 5.8 GHz with 400 MHz bandwidth and 1 KHz PRF, whereas the CW radar works at 24 GHz, thus there is no inter-frequency interference between two radar sensors. The magnetic sensor embedded in the new IMU is a Hall-effect based magnetometer (BMM150) from Bosch, which is capable of sampling the data at 20 Hz with a super high resolution close to 0.3 µT and roughly ±1300 µT measurement range.

Previously, the magnetometer samples at 100 Hz, whereas in this experiment the sampling frequency reduces to 20 Hz. Figure 5.10 below explains the reason by comparing the SFS classification accuracy for the two different sampling rates, where the SFS algorithm is implemented on the normal (100 Hz) and down-sampled (20 Hz) magnetometer data from previous data collection separately. It is observed that the performance of 20 Hz surpasses the 100 Hz with a nearly 5% improvement of the highest accuracy. Additionally, CW radar is not considered in the sensor fusion for the similar function of Doppler measurement as FMCW radar.

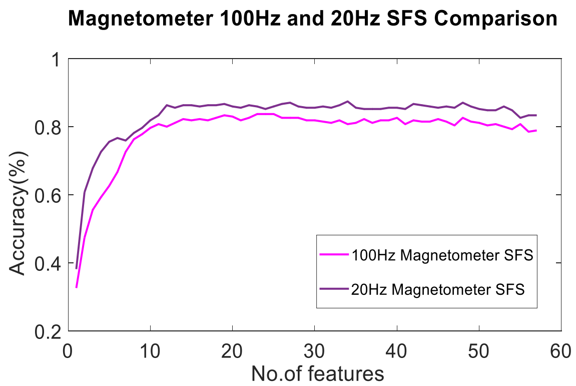


Figure 5.10 Comparison of SFS performance when sampling the magnetometer data at 100 Hz and 20 Hz

5.2.2 Sensor Fusion using SVM Classifier

Similarly, statistical features were extracted from both radar and magnetometer data before the classification process. The list of radar features is the same as described in section 5.1.2 except three range-Doppler features (ID: 70-72 in table 4.1) and one of the SVD-based features (ID: 39 in table 4.1) which are dropped based on the negative influence on the classification accuracy. For the magnetic sensor, all the 64 features in table 4.2 are used, where skewness and kurtosis are added to validate their performance on a larger dataset.

For classification, the performance of a robust SVM classifier with quadratic kernel and a multi-hidden layer ANN are evaluated with the help of SFS. Besides that, ‘Holdout’ cross-validation method is used to separate the training and testing set, which is depicted as following:

- The dataset is randomly divided into two parts depending on the training and testing ratio P, in the condition that selecting equal number of samples for each class.
- This training and testing scheme will run K times and the results of each test are saved.
- The final classification accuracy is the mean value of all K iterations.

Where in our case, 70% of data is used for training and 30% of data for testing, and K equals to 10. Beyond that, ‘Holdout’ is able to avoid class imbalance, which occurs in the ‘K-fold’ cross-validation because the test set has unequal number of samples for each class.

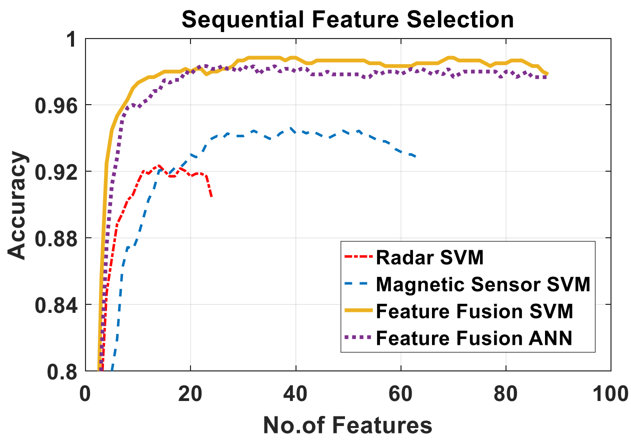


Figure 5.11 SFS classification accuracy for different sensors and classifiers

Figure 5.11 above compares the classification accuracy for radar, magnetic sensor and feature fusion when different classifiers are used to select features by SFS algorithm. Magnetic sensor using SVM yields the highest accuracy after using 34 out of 64 features, and for the radar sensor, 14 out of 24 features are selected. With the optimal feature combinations by SFS, the mean classification accuracy is approximately 94% for the magnetic sensor and in the order of 92-93% for the radar. In terms of using feature fusion along with SFS, the accuracy profiles for SVM and ANN outperform than using single sensor, where the highest classification accuracy for SVM reaches 97% when using 40 out of 88 features and this number is slightly higher than ANN. Additionally, ANN used for fusion includes one hidden layer and 50 neurons, the details of its performance will be discussed later in this section.

SFS selects the most appropriate feature combinations from magnetic sensor and radar in the fusion through SVM, which are summarized in table 5.5 below. Compared with previous SFS results in section 5.12, there are some overlapping features, magnetic sensor includes cross-correlation based features, minimum value of each axis, spectral power features and FFT coefficients, whereas radar sensor has Doppler centroid, bandwidth as well as energy curve. It is also interesting that features related to cadence velocity are added to the optimal feature set by the algorithm.

Magnetic Sensor	Radar
STD of auto-correlation of axis	Mean of centroid
RMS of axis	Mean of principle V vector
Mean of cross-correlation between axis	Mean Doppler bandwidth
STD of cross-correlation between axis	STD of Centroid
Gated Spectral Power Y, Z	Entropy of the image texture
Min of axis	STD of Doppler bandwidth
STD of auto-correlation Z	STD of energy curve
Mean of auto-correlation X,Z	Step repetition frequency
Spectral Entropy Y	Minimum cadence velocity
75th percentiles Z	Maximum cadence velocity
Norm of X, Y, Z	
Mean X and Mean Z	
STD Y and STD Z	
Inter-quadrature Range Z	
Sum of FFT Z	
25th percentile X	
Variance Y	
Range Y	
Total 30	**Total 10**

Table 5.5 Significant features from the fusion feature set selected by SFS

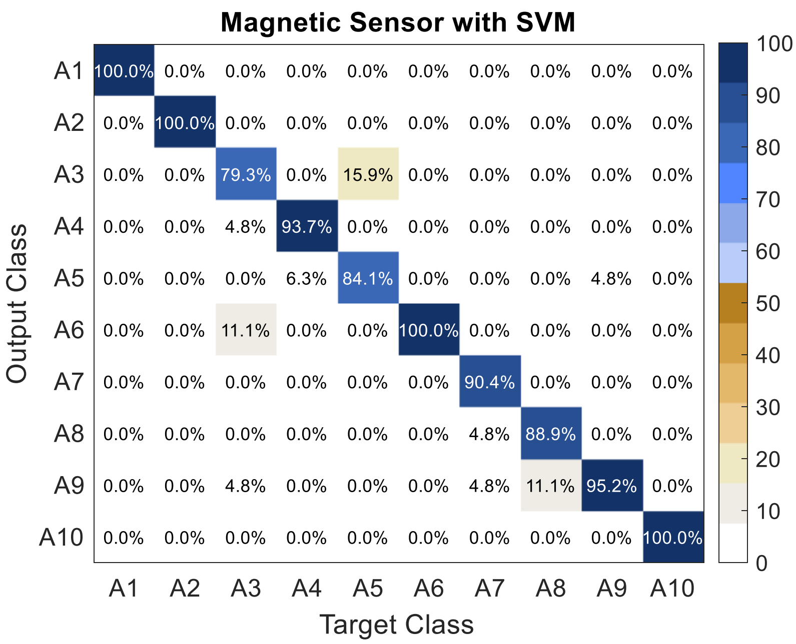


Figure 5.12 Confusion matrix of magnetic sensor with SVM

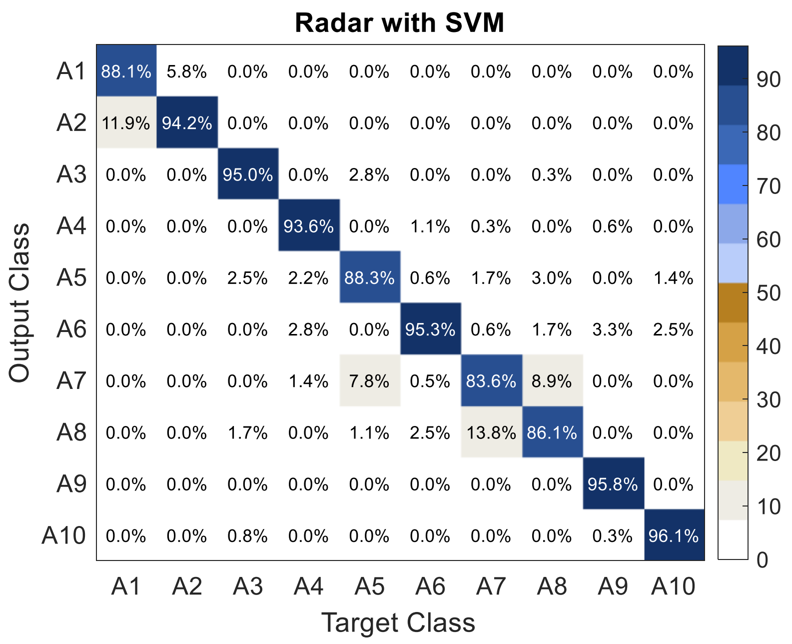


Figure 5.13 Confusion matrix of FMCW radar with SVM

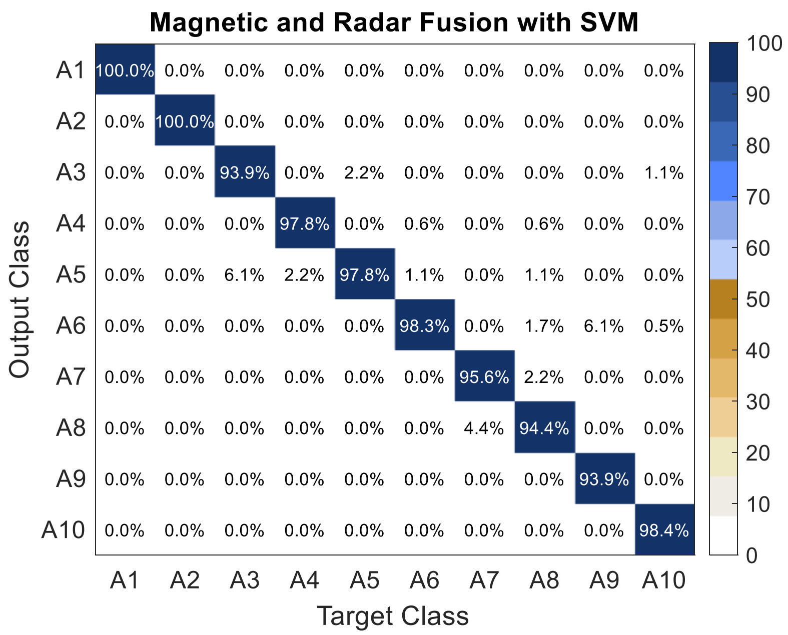


Figure 5.14 Confusion matrix of feature fusion with SVM

Figure 5.12 and 5.13 summarizes the classification results using the best feature combinations for magnetic sensor and radar respectively, where magnetic sensor indicates higher sensitivity and specificity for almost all the activities except A3 'sitting down', A5 'picking up an object' and A9 'falling'. For the fall detection, a low specificity will lead to lots of false alarms and confusions for the whole health monitoring system, and as a result of that, this may reduce the confidence level of the users and their health care providers about its reliability. In figure 5.12, the prediction results of magnetic sensor involves many false alarms for A9 'falling', whereas in the radar case, there is no activity that misclassified to the fall event. This suggest that radar as a contactless sensing approach, is more suitable for fall identification. However, radar is not able to distinguish A1 and A2 very well due to the similarity between walking styles and the limitation of radar range resolution. For the sensor fusion using SVM presented in figure 5.14, the specificity for the fall detection is improved to the same level of radar, with a little accuracy compromising to the A6 ‘tying the shoelaces’. Beyond that, most of the classes are correctly classified, yielding an average accuracy over 96% and the highest misclassification rate is decreased to 6.1%. The classification results of magnetic and radar fusion show that each sensor prefers different human activities, and feature combination of different sensors could be more adaptive in terms of activity recognition and fall detection.

5.2.3 Sensor Fusion using ANN Classifier

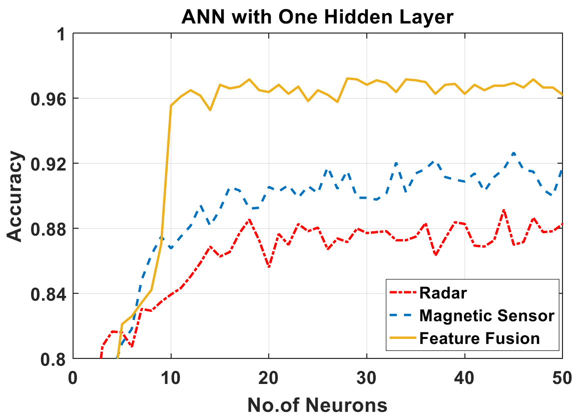


Figure 5.15 The classification accuracy of one hidden layer-ANN with different sensors and number of neurons

Figure 5.15 presents the classification results as a function of neuron numbers with respect to different sensors and fusion. Magnetic sensor still outperforms the radar and it reaches the highest accuracy point when 45 neurons are used, whereas in the radar case, the maximal accuracy appears at 18 neurons. Compared to the number of features selected by SFS for both case (14 for radar and 34 for magnetic sensor), it is interesting that the number of neurons on the hidden layer follows the same manner. This may indicate that ANN is capable of internally selecting the most relevant information from the feature spaces. For the sensor fusion, it is better than using radar and magnetic sensor individually, in terms of both classification accuracy and number of neurons used. The accuracy converges much quicker and the amplitude of accuracy change is fairly small after using 10 neurons. The optimal accuracy point of feature fusion is over 96%, which is very similar to the SVM classifier with the help of SFS.

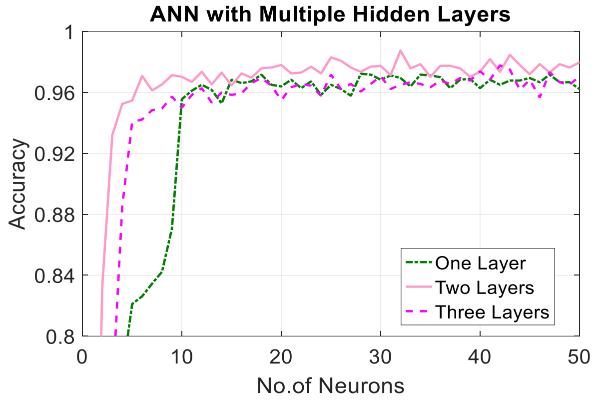


Figure 5.16 The classification accuracy of ANN with different numbers of hidden layer

Figure 5.16 compares the difference in classification performance when using a multi-hidden layer structure, where the number of neurons is varied on the last hidden layer with a constant at 50 neurons on the other layers. There is no significant accuracy difference between using one and two hidden layers, whereas the increasing number of layers will lead to the growth in the computational power and time cost. In terms of the number of neurons, the accuracy curve starts to converge after adding 15 neurons on the single hidden layer network, for the multi-hidden layer architecture, it takes less neurons to reach the highest point. This is due to that the first hidden layer filters the noise and transfers the useful information to the coming layers.

Epochs	Training Accuracy (%)	Testing Accuracy (%)
10	94.6	82.5
20	100	92.9
30	100	96
40	100	98.4
50	100	94.4

Table 5.6 Training and testing accuracy versus number of epochs

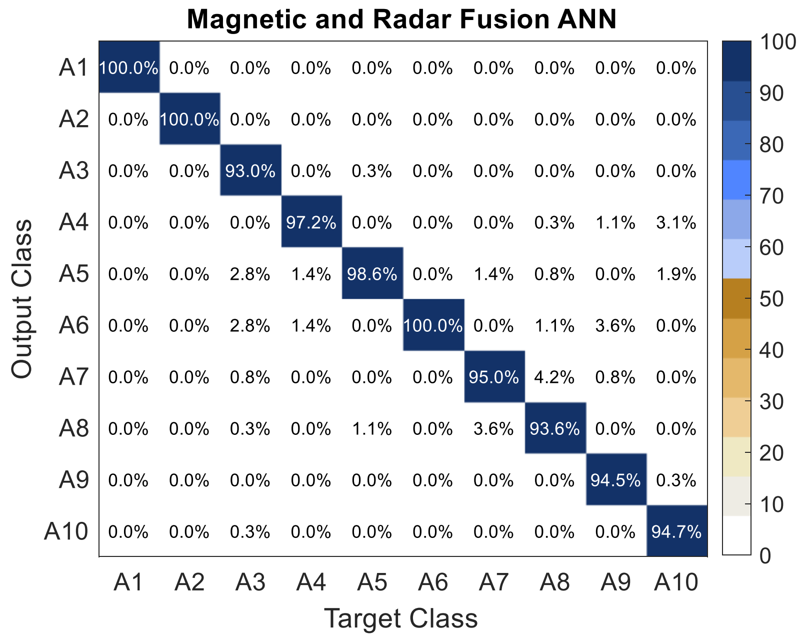


Figure 5.17 Confusion matrix of feature fusion with ANN

The change of training and testing accuracy over the training epochs of proposed ANN are summarized in table 5.6, where the training accuracy reaches 100% after 20 epochs and the testing accuracy increases in a gradual manner until 40 epochs, over that, the neural network turns to be overfitted. Therefore, 40 epochs is set as the maximum time for training the ANN. Figure 5.17 illustrates the classification performance of magnetic sensor and radar fusion with ANN, the overall performance is close to SVM in figure 5.14 aside from that the sensitivity of fall event is slightly improved and there are more misclassifications in the A10 ‘checking under the bed’.

Classifier/Sensor	***Precision (%)***	***Sensitivity (%)***	***Specificity (%)***	***F-measure (%)***
SVM (Radar) *	91.4	91.7	99.1	91.4
SVM (Mag) *	92.3	93.2	99.1	92.0
SVM (Fusion)	96.8	96.6	99.6	96.7
SVM(Fusion) *	97.8	98.3	99.9	98.1
ANN (Radar)	89.6	89	98.8	89.3
ANN (Mag)	92.3	92.1	99.1	92.2
ANN (Fusion)	97.4	97.2	99.7	97.3
ANN(Fusion) *	97.4	97.9	99.9	97.7

Table 5.7 Performance comparison of different sensor methods: Average of all 10 classes - ***** Indicates use of SFS

The summary of different classification metrics are shown in table 5.7 with respect to all the aforementioned sensor combinations and classifiers. Compared to using radar and magnetic sensor separately, fusion can produce significant improvement in terms of both sensitivity and specificity, where the overall sensitivity increases about 6.6% for the SVM classifier and in the order of 5.7%-8.9% for the ANN. The best case for SVM and ANN share similar performance, whereas the cost of ANN is few times higher than SVM considering the training time. Additionally, SFS algorithm is more suitable to apply on the SVM since the level of improvement could be ignored on the ANN.

5.2.4 Classification Results of ‘L1O’ Cross-Validation

A more realistic and challenging cross-validation method, ‘Leaving One Participant Out’ (L1O) is used to evaluate all the aforementioned methods, compared to the ‘K-fold’ and ‘Holdout’, the classifier is not allowed to see the data from unknown participants before the actual testing. The breakdown of ‘L1O’ is shown below:

- Data from one participant is picked out for testing while the rest of participants are used for training the classifier.
- This training and testing scheme will run K times until each participant has been treated as ‘testing subject’.
- The final classification accuracy is the mean value of all K iterations.

Where K denotes the number of participants in the dataset, in this case, K equate to 20.

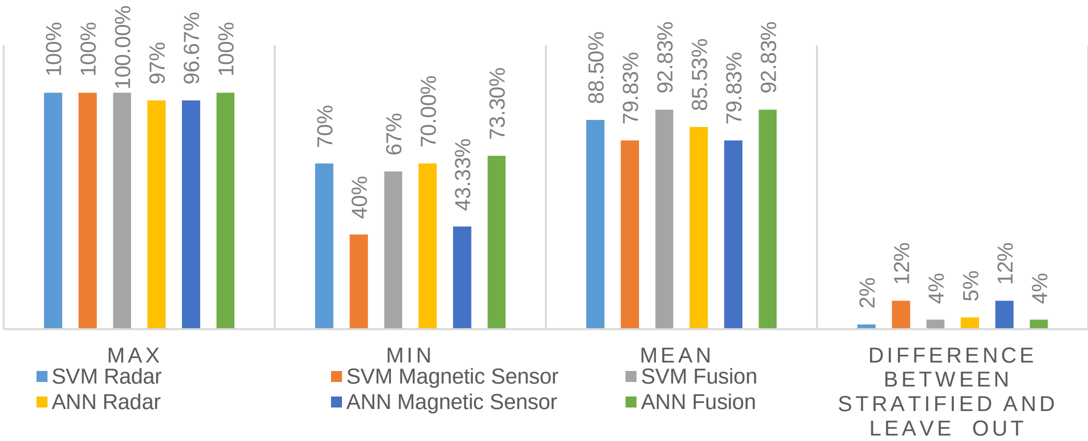


Figure 5.18 Comparison of statistical metrics when the classifier is trained and validated using 'L1O' approach.

The results of using 'L1O' cross-validation are presented in figure 5.18 in terms of different sensor combinations and classifiers, where the maximum, minimum and mean accuracy for each case are calculated along with the difference between 'L1O' approach and the stratified test in table 5.7. The maximum and minimum accuracy corresponds to the best and worst individual participant under test, whereas the mean is the average accuracy of testing all the participants. 'L1O' results shows that the classification accuracy is varied by the participants, especially for the magnetic sensor with both classifiers, the lowest accuracy reaches nearly 40% for one specific subject. In the radar case, the minimum accuracy is a lot better, beyond that, the difference between the 'Holdout' tests and 'L1O' is limited between 2 and 4%, which suggests that radar outperforms magnetic sensor in terms of robustness. Most of the metrics related to the magnetic sensor improves thanks to the feature level fusion, it is reported that the performance of the worst participant is increased to the same level as radar and gains approximately 12% growth in the average accuracy. There is prominent difference between the minimum value and the ideal numbers in table 5.7, which could be more than 50% for some extreme cases. This is due to that everyone has unique performing styles for different activities and 'L1O' reflects those realities on the training and testing scheme. With the help of sensor fusion, the difference between previous stratified test and 'L1O' approach is narrowed to approximately 4% for both SVM and ANN.

5.3 Summary of the Chapter

In this chapter we discuss the classification results for snapshot activities. Different information fusion methods are utilized to combine the strengths of wearable IMU sensor and FMCW radar (LOGP soft fusion, fuzzy logic and voting system) and their results under different classifiers are compared.

Furthermore, different feature selection techniques are applied on the handcrafted features to drop the redundant information and wrapper method SFS yields the best performance. The weakest sensor when used individually, namely, magnetic sensor, is then picked out to fuse with radar at feature level and their fusion results are verified under a stricter 'L1O' cross-validation scheme. Some of the classification results are summarized in table 5.8 below.

Dataset	Sensor/Fusion method	Classification method	Classification accuracy/Cross-validation
inertial sensors and radar (270 samples)	inertial sensor	SVM	95.6%/10 fold
inertial sensors and radar (270 samples)	FMCW radar	SVM	85.6%/10 fold
inertial sensors and radar (270 samples)	both sensors/feature fusion	SVM	97.4%/10 fold
inertial sensors and radar (270 samples)	both sensors/LOGP fusion	SVM	96.7%/10 fold
inertial sensors and radar (270 samples)	both sensors/majority voting +LOGP fusion	SVM+KNN	97.8%/10 fold
magnetic sensor and radar (600 samples)	magnetic sensor	SVM	79.8%/L1O
magnetic sensor and radar (600 samples)	magnetic sensor	ANN	79.8%/L1O
magnetic sensor and radar (600 samples)	FMCW radar	SVM	88.5%/L1O
magnetic sensor and radar (600 samples)	both sensors/feature fusion	SVM	92.8%/L1O
magnetic sensor and radar (600 samples)	both sensors/feature fusion	ANN	92.8%/L1O

Table 5.8 The summary of classification results presented in Chapter 5

6 Continuous Human Activity and Sequential Gait Recognition

This chapter focuses on the classification of continuous human activities and gaits. A dual Bi-LSTM layers neural network is proposed to process the continuous data stream and its performance is compared with conventional sliding window-based method, whereas the sensor fusion takes places between both wearable IMU sensor and FMCW radar (for activity recognition) and between distributed radars (for gait analysis).

Additionally, different fusion methods involving soft fusion, hard fusion (RC and NBC), and hybrid fusion (combining soft and hard fusion) are compared in terms of the accuracy gain, whereas the performance of range information fusion using the proposed trilateration-based method are compared with micro-Doppler fusion using the best hard combiner at two different perspectives (average accuracy and computational cost).

6.1 Continuous Human Activity Recognition and Fall Detection

The data analysed in this section are continuous activity streams from FMCW radar sensor and three wearable IMUs on the different parts of human body. Opposite to the traditional snapshot data, continuous data stream contains a list of unsegmented activities with natural transitions between them. In our case, the time length of each single activity is not fixed, which means that the transition parts take place at any possible times during the data stream.

In terms of the classification for continuous data sequence, a conventional sliding window-based method is first used to segment the data into smaller frames and then process the data frame the same way as the snapshot, whereas a dual Bi-LSTM layer network is then proposed to learn the long-term time dependency within the continuous data. Beyond that, soft fusion based on the confidence level is applied between wearable sensor and FMCW radar through the proposed Bi-LSTM network, along with two hard fusion approaches, namely, RC (Recall Combiner) and NBC (Naïve Bayes Combiner). For gaining the advantages from both soft and hard fusion, a novel hybrid fusion approach is then raised to fuse the weighted soft fusion results via the two hard combiners. All the fusion and classification methods are tested with ‘L1O’ (Leaving One Participant Out) cross-validation, where hybrid fusion boosts the average classification accuracy for continuous activities to 96%. Compared to using the single sensor, it is also reported that the accuracy variance across different participants decreases by 18.1% and the minimum accuracy for the worst case rises up to 16.2%.

6.1.1 Experimental Setup

The experiment collects human motions data from 1 female and 15 male subjects from one FMCW radar sensor and three wearable IMU sensors placed on the human wrist, waist and ankle. The experimental setup is shown in figure 6.1, the transmitter and receiver of the FMCW radar points to an activity zone (3m x 2.2m), where the daily activities are performed and the fall accidents are simulated by the participants.

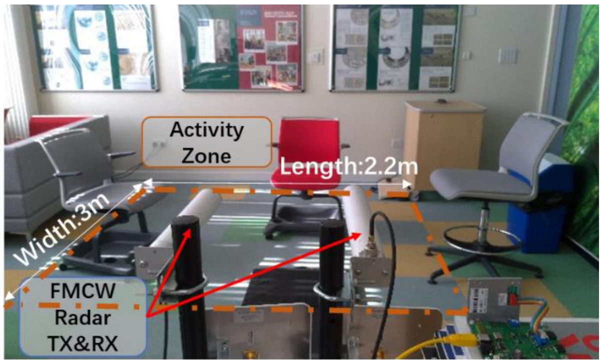


Figure 6.1 View of the experimental setup for recording data: common room at the University of Glasgow, with furniture and clutter nearby.

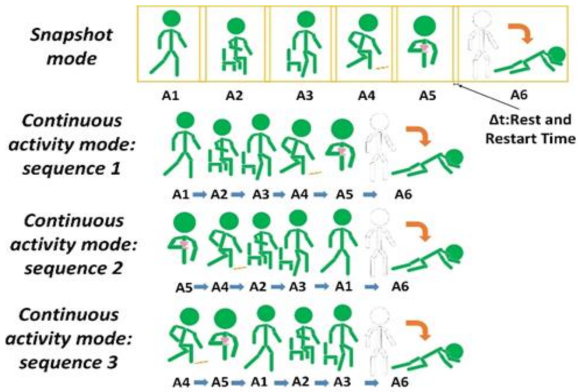


Figure 6.2 Sketch of the human activities recorded - top: snapshot mode, bottom: continuous activity mode from sequence 1 to 3.

Three IMUs are connected and synchronized by a wireless router, whereas the radar and IMU are also operated simultaneously with a MATLAB script to compensate the offset between the two data collection triggers. The collected activities are sketched in figure 6.2, including A1 ‘walking back and forth’, A2 ‘sitting on the red chair’, A3 ‘standing up from the red chair’, A4, ‘bending to picking up a pen’, A5 ‘drinking water from a glass bottle’ and A6 ‘simulating a fall on the mattress’. The data collection mode in the top part of the figure is ‘snapshot’, where the activities are recorded individually with fixed time length (10s for A1 and 5s for the rest) and there is an interval for rest between two ‘snapshot’ activities. The ‘snapshot’ activities are designed to simulate a simple scenario that the users will only perform one specific activity within a short period of time (no transfer between two activities, no repetition). The rest of the figure shows the continuous activity streams with respect to three different orders of six activities (each activity sequence ends with a fall), whereas the whole time length for a single sequence was limited to 35s. These activities are connected by the transitions (e.g. the participant needs to put the bottle back before falling in sequence 1) without further constraint on their durations. Moreover, the participants have the freedom to decide the transfer time between two activities for simulating the realistic scenario that users will change their activity patterns without reminding the classification system. In terms of application, the analysis of these ‘snapshot’ activities could be used in where long-term data collection is not available or simply no condition to process/store a large amount of data, whereas continuous activity classification could be employed on real-time health monitoring systems in care homes and hospitals with the help of high-performance computing units and cloud storage.

The dataset contains 48 continuous sequences, where each participant was asked to perform sequence 1-3 in figure 6.2. In terms of Degrees of Freedom (DOF), each sequence has 28 DOFs including FMCW radar and three IMUs with tri-axial accelerometer, gyroscope and magnetometer. Regarding the number of data points/time bins in the continuous activity stream, each sequence has 35s duration, and the sample rate for wearable sensor data collection is 50 samples/s, thus the total number of data points within one continuous sequence is 1750 (35x50) for the wearable sensor. For the radar sensor, this number is almost double (the spectrogram has more time bins than wearable data). Therefore, the radar spectrogram is downsampled to 1750 to keep consistency with the wearable sensor.

6.1.2 Conventional Sliding Window-based Approach

For the IMU, all the features in table 4.2 were extracted from the raw data. There are totally 174 (58 x 3) features considering the number of sensors within one IMU.

For the FMCW radar, 12 physical features involving Doppler centroid (ID.1-12 in table 4.1) and bandwidth as well as 8 features from SVD transform (ID. 24-27 and 36-39 in table 4.1) are used to train the classifier.

The features from IMU and radar are concatenated, yielding a feature matrix with 194 columns (58x 3+20). Sequential Backward Selection (SBS) with a quartic kernel SVM classifier is used to choose the most powerful feature combinations and reduce the input dimension of classifier, compared to the SFS in previous sections, SBS starts with the entire feature set and exhaustively removes redundant features till there is no significant accuracy improvement. The best advantages of SBS is the time cost, in particular dealing with a high-dimensional feature set like our case, whereas the performance for those two methods are generally very close. To make SBS more efficiently, 50 ‘surviving’ features are set as the threshold to stop the algorithm since the accuracy reaches its plateau after that. Figure 6.3 shows the relationship between the classification accuracy and the number of features dropped, where the maximal accuracy of fusion is approximately 93% by using 57 out 194 features, for the 57 features, 11 features are from FMCW radar and the rest of them is contributed by IMU. Compared to using all the available features, SBS improves the classification accuracy for roughly 3.1% and drop nearly 75% features of the full fused feature set, leading to more than 60% saving in classification time. It is also interesting to notice that dropping too many features will not improve the classification accuracy significantly (dropping 125 features only improves about 1.5% on top of dropping 25 features).

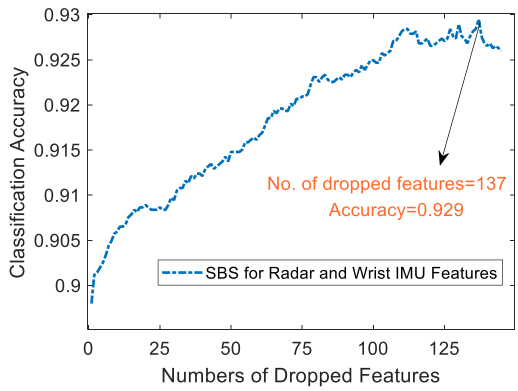


Figure 6.3 SBS classification accuracy as a function of numbers of features dropped

Continuous data could be considered as a combination of snapshot, to process it through conventional classifier (because conventional classifier like SVM can't be used for sequential classification directly), a sliding window is necessary to divide the data stream into smaller segments for extracting the statistical features and training the classifier. In this case, the label of data segment is decided by the majority class of the data samples within that segment (e.g. 60% of the data samples are 'walking', 40% is 'sitting down', and then the label for the data segment is 'walking'). However, the performance of continuous activity recognition is varied by the window size and overlapping percentage between two nearest window. For finding the optimal parameters, window sizes from 2s to 5s with an interval of 0.5s are tested along with overlapping factor between 0% and 90%. Meanwhile, in terms of feature fusion, different combinations of radar and single IMU sensor are verified and compared to find the best partner for radar. Additionally, 'L1O' cross-validation is used as the training and testing scheme, where the SVM classification model is trained by data from 15 participants and tested with the remaining 16th participant, at the same time, the whole process will repeat 16 times until each participant has been used as a test set and their results are averaged.

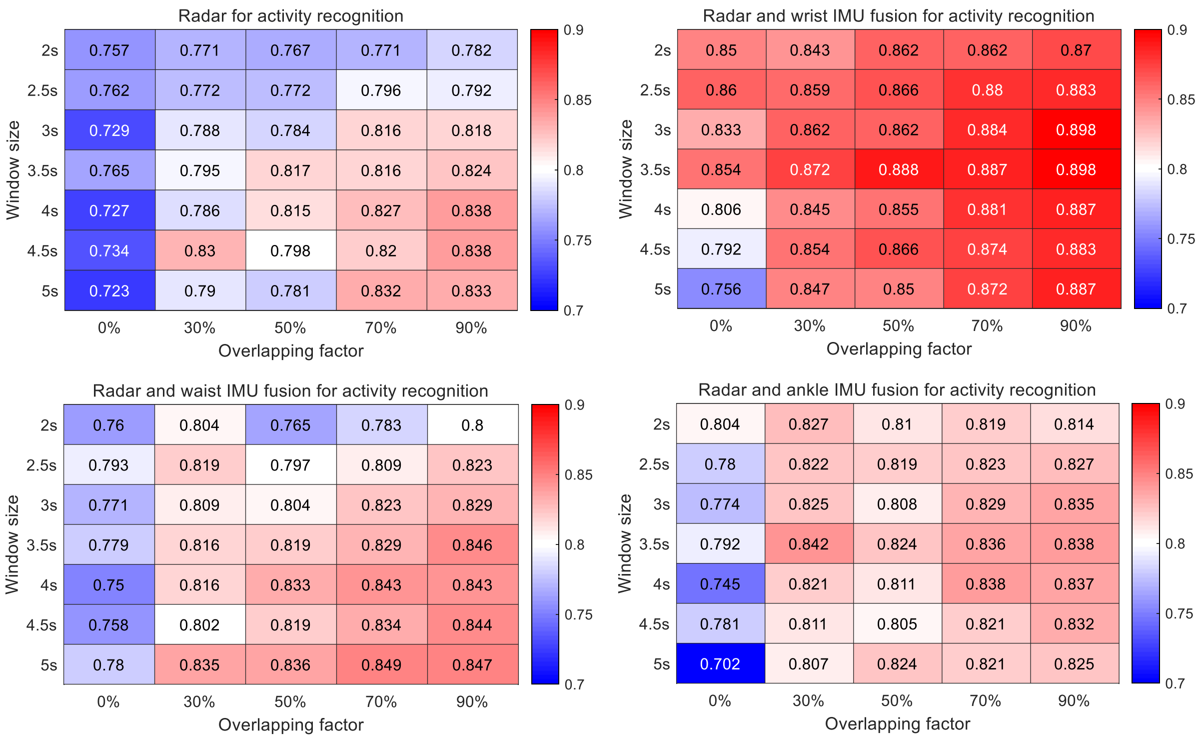


Figure 6.4 The heatmap of the relationship between sliding window size, overlapping factor and classification accuracy (left above: radar-only, right above: radar and wrist IMU fusion, left below: radar and waist IMU fusion, right below: radar and ankle IMU fusion)

The classification results based on sliding window method are summarized in figure 6.4 as a format of heat map. It is reported that radar yields an accuracy of 83.8% when using a 4s window with 90% overlap, while the best sensor combination (radar and wrist IMU) improves about 6% on top of using radar alone. The classification accuracy reaches approximately 89.8% with a window size equal to 3.5 s and the highest overlap. From the results of all sensor combinations, there is a trend that the accuracy improves with the rise of overlapping percentage, whereas the optimal window size appears to be in the range of 3s to 4.5s.

The confusion matrices for using radar individually and jointly used with wrist IMU are presented in figure 6.5 and 6.6, respectively. In the radar case, for the most critical fall events, the sensitivity is not very high and there are few minor false alarms. Class A2 and A3 share similar sensitivity with nearly 7% of the activity ‘sitting down’ have been misclassified to ‘standing up’. Additionally, main confusion takes place between A4 and A5, where the classifier misclassifies more than 15% of both activities. The feature fusion in figure 6.6 improves the sensitivity of identify fall event by 7.9% along with a boost of approximately 21% for both A4 and A5. Apart from that, there are still lots of minor classification errors across different classes.

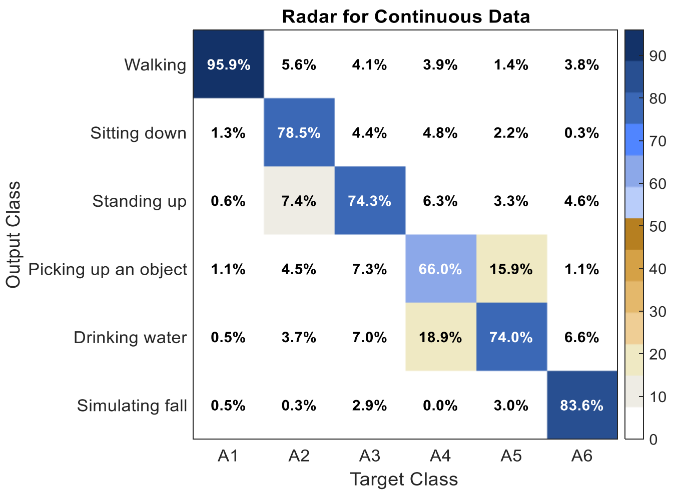


Figure 6.5 Confusion matrix of FMCW radar for continuous activity recognition using a sliding window

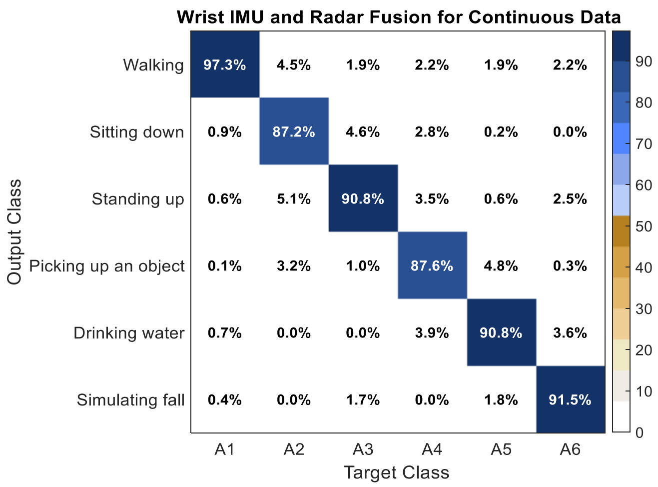


Figure 6.6 Confusion matrix of wrist IMU and radar fusion for continuous activity stream using a sliding window.

6.1.3 Bi-LSTM network-based Approach

Conventional sliding window method is very computational intensive especially used with a small window size and high overlapping factor. A dual Bi-LSTM layer neural network is proposed to tackle the problems from sliding window method and provide subsequent accuracy improvement. The wrist IMU sensor is selected to combine with FMCW radar according to the SVM performance in figure 6.4. Furthermore, the proposed Bi-LSTM network is trained and tested under the 'L1O' method, which is slightly different with the process on SVM classifier. The data from the 16th subject is utilized to construct a validation dataset for monitoring the training progress, whereas the rest of subjects follows the classic 'L1O', and in that case, the final classification accuracy is an average of 15 rounds of test.

Table 6.1 below summarizes the training hyper-parameters, where the training, validation and testing ratio is 14: 1: 1. For the learning rate, the initial value is set as 1e-3, whereas it will drop 90% after 200 and 400 training epochs with respect to wrist IMU sensor and FMCW radar. A drop-out layer with 50% dropping probability is connected with each Bi-LSTM layer to avoid overfitting problems. Furthermore, the validation dataset is used to test the incomplete network once per epoch to provide the validation accuracy.

Hyper-parameters	Radar	Wearable Sensors
Training: Validation: Test	14:1:1	14:1:1
SGD Optimizer	Adam	Adam
Decay	0.9	0.9
Initial Learning rate	1e-3	1e-3
Learning rate drop period	200	100
Number of input dimension	8	9
Number of bi-LSTM layers	2	2
Number of dropout layers	2	2
Dropout probability	0.5	0.5
Training epochs	400	200
Validation frequency	Once per epoch	Once per epoch

Table 6.1 The hyper-parameter of the proposed Bi-LSTM network

Doppler centroid and bandwidth, upper and lower envelope, along with the statistical moments including mean, standard deviation, skewness and kurtosis of the spectrogram are extracted to serve as input features for the proposed Bi-LSTM network. All those input features are generated as a function of time, therefore, the network can learn the forward and backward time dependency between each time bin. In the case of IMU, nine features are used involving tri-axial information from accelerometer, gyroscope and magnetometer. The training and validation accuracy as well as the loss based on cross entropy are plotted in figure 6.7, where the wrist IMU data processing reaches its plateau more quickly (100 epochs) than the FMCW radar case (200 epochs). In terms of the validation accuracy, both sensors stay at 90%, whereas the difference between peaks of their training accuracy is only 2%.

Figure 6.8 and 6.9 show the Bi-LSTM classification results for radar and wrist IMU respectively, where the radar yields an average classification accuracy across the 15 participants about 88.9% and the wrist IMU surpasses radar for only 0.2%. Similar to the sliding window method, the main classification errors for FMCW radar occurs between A4 and A5, whereas the remaining major misclassifications are between A1, A2 and A3. Compared to figure 6.5, the proposed network increases the sensitivity of fall detection by 8%. For the wrist IMU case, 15.7% of the most significant fall events are misclassified with nearly 10% false alarm rate. For the Bi-LSTM network, the new prediction is based on the knowledge learned from previous and following activity patterns, thus two closest activities within the continuous sequence are usually confused. In terms of the location of the activity transitions, there is a 'shift' between the predicted sequence of activities and the ground truth.

Soft fusion is utilized to combine the confidence level of different sensors, where radar and wrist IMU share the same weight in the fusion process. The classification results are summarized as a confusion matrix in figure 6.10. The average classification accuracy of soft fusion improves about 5.7%, with respect to using wrist IMU and radar separately. Fusion increases the sensitivity of each class, in particular, the detection of falls, where the enhancement is more than 10% on top of the wrist-IMU results. Beyond that, the misclassifications between two nearest classes are reduced, which suggests that soft fusion improves the ability of correctly identifying the transition parts between activities.

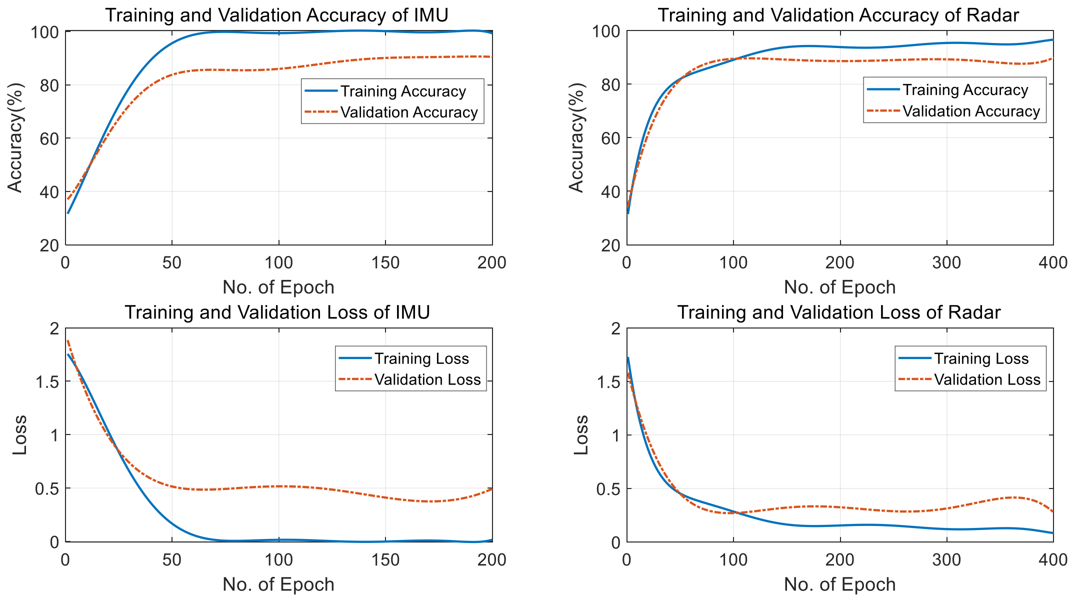


Figure 6.7 Training progress of IMU and radar using double-layer bi-LSTM (top-left: training and validation accuracy of inertial sensor, top-right: training and validation accuracy of radar, bottom-left: training and validation loss calculated by cross entropy for radar)

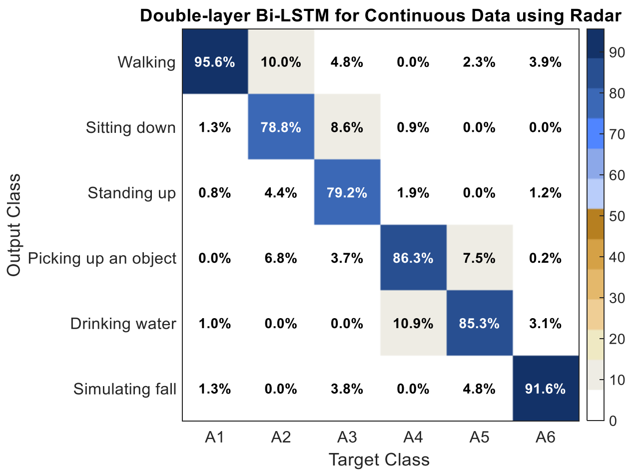


Figure 6.8 Confusion matrix for the proposed Bi-LSTM network using radar data

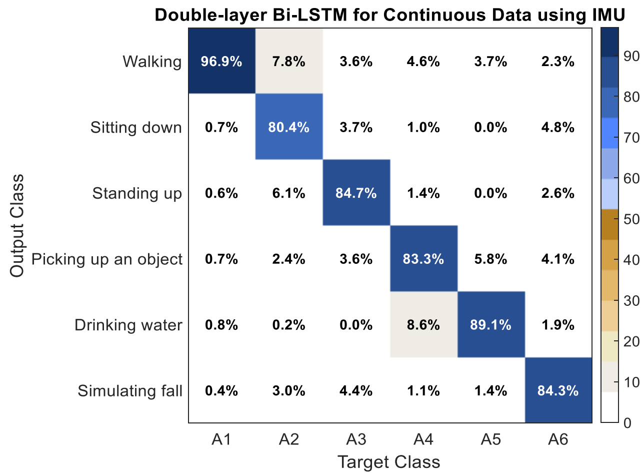


Figure 6.9 Confusion matrix for the proposed Bi-LSTM network using wrist IMU data

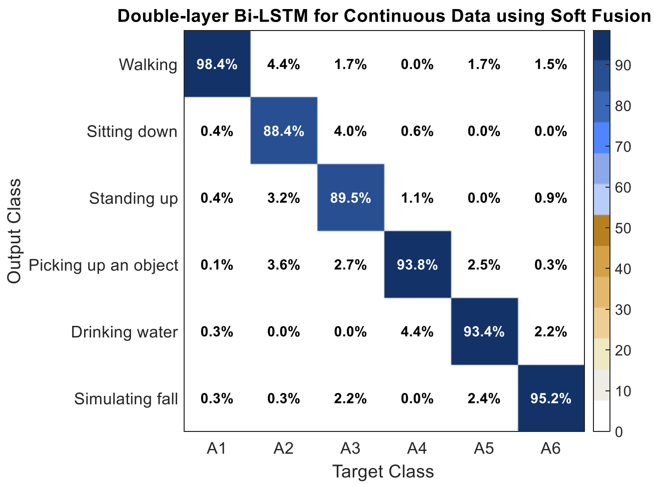


Figure 6.10 Confusion matrix for the proposed Bi-LSTM network with soft fusion

Two hard fusion approaches, namely Recall Combiner (RC) and Naïve Bayes Combiner (NBC), are used to fuse the prediction results from radar and wrist IMU, instead of focusing on their confidence levels. Both RC and NBC are based on the confusion matrix of each sensor, which is computed by the number of corrtely classfied and misclassified events. Additionally, a novel hybrid fusion method is proposed to incorpoate the soft and hard fusion schemes, where the soft fusion results with different sensor weights are considered as 'virtual classifiers' for the inputs to the RC and NBC. The ratio of weights in the soft fusion process is varied from 10:1 and 1:10 for the radar and wrist IMU, whereas the number of 'virutal classifiers' is determined by the step of changing the ratio. This saves computational power and time for training the realistic classifers, especially with a deep neural network.

The classifiers used as inputs of the combiner are stored in an ensemble, whereas table 6.2 shows the variation of the classifier ensemble in terms of its length and content. For the classic hard fusion, the prediction results from radar and wrist IMU are used; thus the length of classifier ensemble L equates to 2. Equal weight soft fusion is included in the proposed method A, and as a consequence of that, the length L increases to 3. The proposed methods B and C use multiple ‘virtual classifiers’ as the inputs of the combiners, where method B changes the weight ratio between radar and wrist IMU as a step of 2, therefore the number of ‘virtual classifiers’ is equal to 10 and the total length of classifier ensemble is 13 (10+3). However, in the case of method C, the step is reduced to 1 for seeking more resolution; thus, the number of ‘virtual classifiers’ is eight more than proposed method B, leading to a total length of 21 (13+8).

In the ideal case, the hard combiner leverages information from all the possible weight ratios between the radar and wrist IMU, whereas the number of redundant classifiers will significantly increase under a small changing step. Hence it is important to select an appropriate changing step to cover most of the useful information without creating too many useless ‘virutal classifiers’. The steps in our case are fixed as ‘0.1’ and ‘0.2’ through a list of attempts.

Figure 6.11 compares the average classification performance between RC and NBC, where the accuracy is plotted against the number of classifiers used in the hybrid fusion process. Additionally, the number of classifiers is in a logarithmic scale and the proposed hybrid fusion is cross validated through ‘L1O’ method. It is observed that NBC surpasses RC for all the cases, whereas the accuracy difference is fairly small (<1%). Proposed method B with NBC yields the optimal performance, producing about 95.8% accuracy. However, the most significant accuracy gain takes place when including the equal weight soft fusion to the classifier ensemble (proposed method A), yielding nearly 1.2% for RC and 0.95% for NBC. The classification performance is not always proportional to the number of classifiers used, since there is no subsequent improvement in adding more ‘virtual classifiers’ over 10, for the RC case, the accuracy even appears a slight drop. This suggestes that adding more ‘virtual classifiers’ will not benefit the overall classificaiton performance.

Classifier ensemble length	Inputs of the combiner
L=2 (Normal hard fusion)	Radar, wrist IMU
L=3 (Proposed method A)	Radar, wrist IMU, normal soft fusion
L=13 (Proposed method B)	Radar, wrist IMU, normal soft fusion, weighted soft fusion with 10 different ratios
L=21 (Proposed method C)	Radar, wrist IMU, normal soft fusion, weighted soft fusion with 18 different ratios

Table 6.2 Number of classifiers used as input of the proposed hard fusion scheme

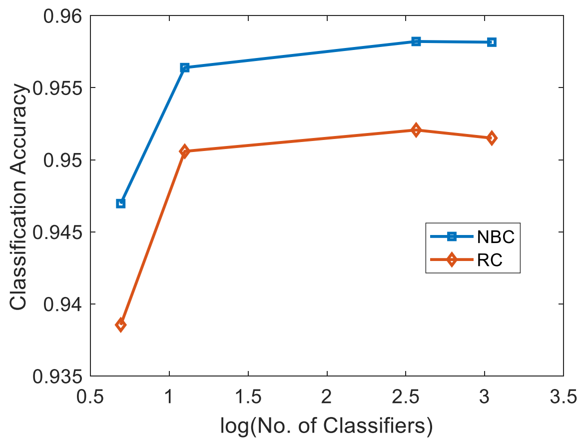


Figure 6.11 Number of classifiers versus classification accuracy for the proposed hybrid fusion scheme

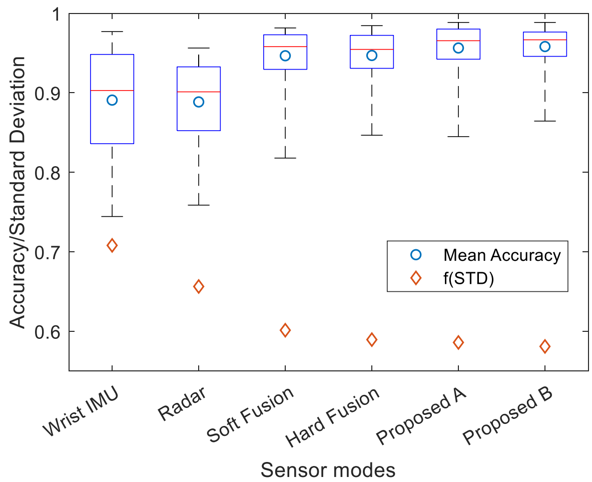


Figure 6.12 Statistical parameters of individual sensor and different fusion methods

The statistical analyse of the classification results for different methods is shown in figure 6.12 to explore more, where the maximum, minimum, mean, median, 25th and 75th percentiles, along with the standard deviation are calculated across all the participants in the 'L1O' cross validation. The standard deviation values in figure 6.12 are amplified by a linear transform (f (STD) multiplies the original value by 3 then add with 0.55) to reach the same scale as other metrics. Compared to using wrist IMU and radar separately, the average accuracy for all sensor fusion methods increase about 5-6.8%, whereas the worst case scenario among all the participants are raised approximately 12% through proposed method B. The gap between 25th and 75th percentiles, as well as standard deviation are highly correlated with the stability and robustness of the classification methods. In the ideal case, they should be as less as possible, the hybrid fusion method B reduces the difference between the two percentiles to 0.13 and the transformed standard deviation to 0.58. This is due to that weighted soft fusion is able to identify different types of activities based on the sensor information ratio, whereas the hybrid fusion leverages those soft fusion results under a hard fusion framework.

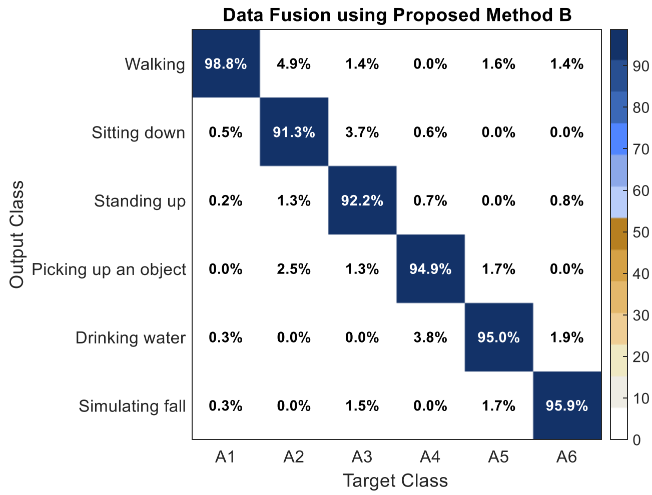


Figure 6.13 Confusion matrix for the proposed hybrid fusion method B (best case)

The overall classfication results for the optimal hybird fusion method (proposed method B) are indicated in Figure 6.13, where the proposed method B increase the sensitivity of most classes for about 0.4-2.9% compared to the soft fusion results in figure 6.10. Meanwhile, the minor misclassiciations between two neighbour classes are reduced to a very low level, whereas the fall detection rate raises about 0.7% with the false alarm ratio dropping to 3.5%.

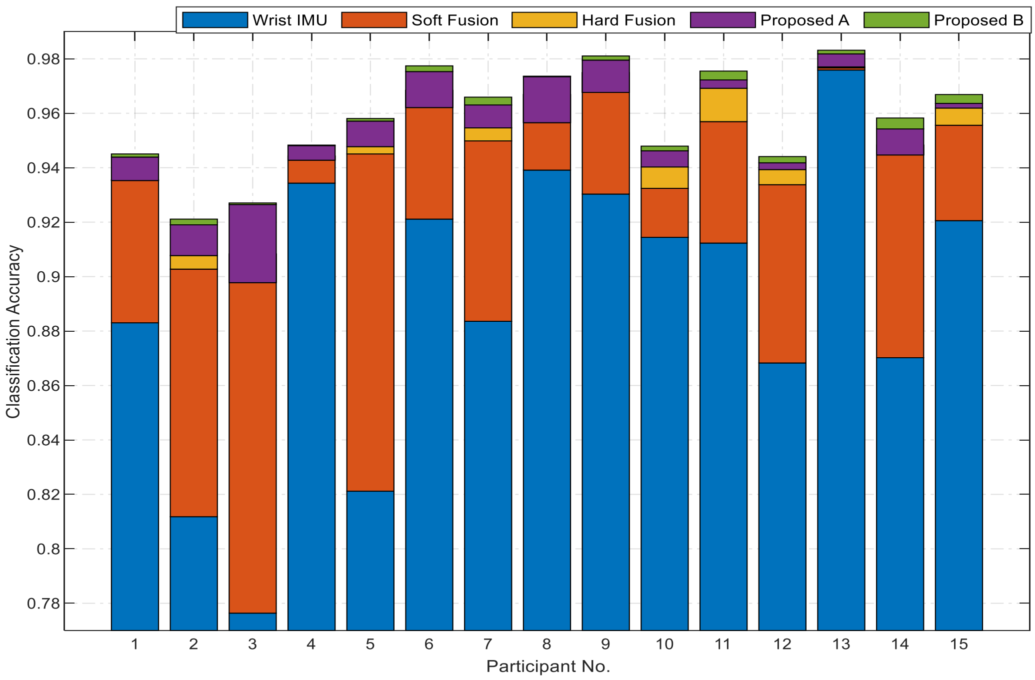


Figure 6.14 The accuracy gain for each subject with different fusion methods (baseline: wrist-IMU results)

The enhancement in classification accuracy for each subject are illustrated in figure 6.14, where the baseline is that using the wrist IMU sensor (blue bar) individually and the improvements contributed by different fusion method are highlighted with different colours. In terms of the volume of improvement, the soft fusion between radar and wrist IMU outperforms others with an average accuracy gain of 5.6%. However, the proposed hybrid fusion method B yields the overall optimal results as it still slightly improves the accuracy across most of the subjects on a very high baseline. Beyond that, all the fusion schemes raises the stability of the classification by reducing the accuracy variance between different subjects (about 1.79% when compares proposed method B and wrist IMU-only). It is also interesting to see that participant No. 2, 3 and 5 share a low classification accuracy when using wrist IMU only, whereas their accuracy improvement via the help of radar sensor is top three among all the 14 participants.

For testing the ability of the proposed Bi-LSTM network about classify the continuous activity sequence from unknown participants, three sequences collected from one random participant and their 'activity tracking' results with two different classification methods are shown in figure 6.15. The first row of the figure indicates the radar Doppler spectrogram with respect to the three selected sequences. Note that the selected sequences are with different orders of activities. The absolute value out of the tri-axial (X, Y, Z) information for the accelerometer, gyroscope and magnetometer inside the wrist IMU are drawn in the second row, with the same continuous activity sequences as radar. It is observed that there is a spike in the wearable data caused by the fall events. The third/bottom row tracks the continuous activity through a wrist IMU and proposed hybrid fusion method B, where the ground truth is emphasized in yellow dash line, eventually, the proposed fusion scheme (red line) rectify most of the classification errors, in particular, between 15-17s in the first sequence (a), between 9-13s in the second (b) and between 22-27s in the third (c).

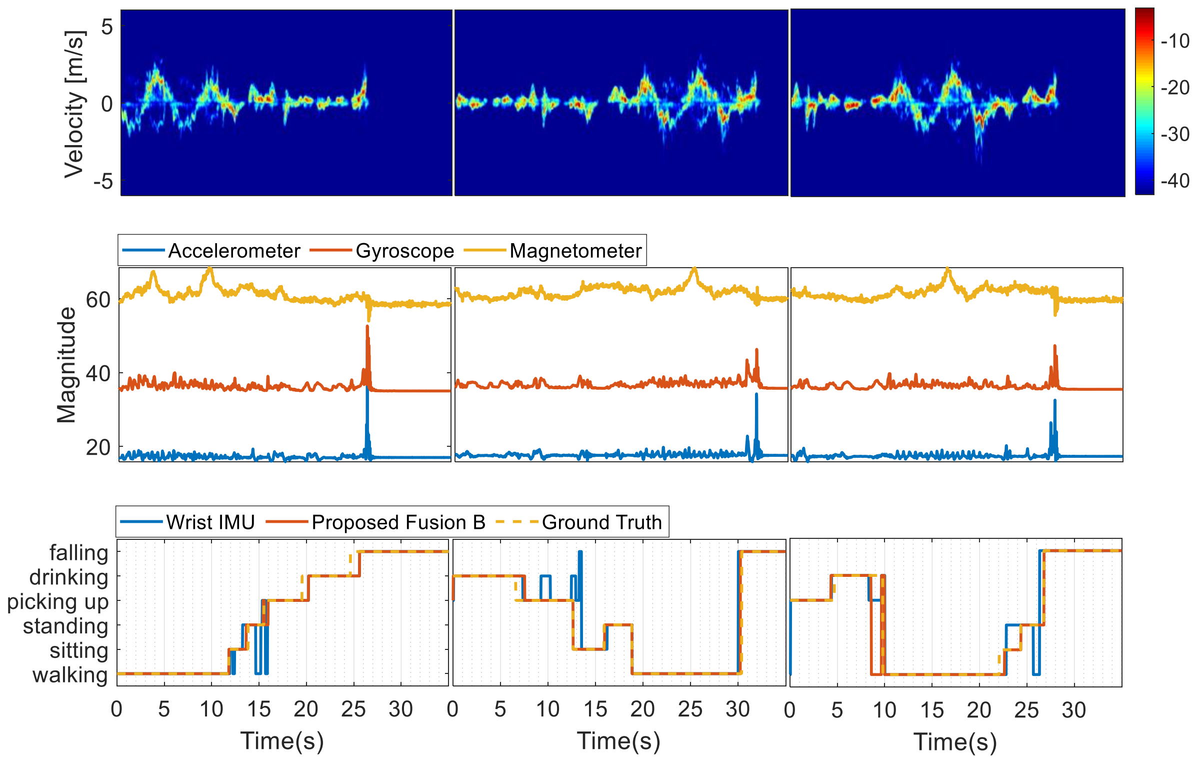


Figure 6.15 The activity pattern 'tracking' (first row: radar Doppler spectrogram for activity sequence 1, 2 and 3 in (a), (b) and (c), second row: absolute value of X, Y and Z axis of the inertial sensor data, third row: prediction label of using IMU-only and proposed method B (best fusion approach)

6.2 Sequential Gait Recognition

In this section, different information fusion approaches, in particular, signal and decision level fusion to recognize human gait patterns and detect fall events. Both individual and sequential gaits are collected through a radar sensor network, including a FMCW radar and three UWB impulse radars placed at different spatial positions. Similar to the continuous activities, sequential gaits are comprised of multiple, different gait styles and the natural transitions between them, beyond that, fall events are performed in conjunction with walking in some cases. For the signal level fusion, the range information from three UWB Doppler radar sensors are combined by a trilateration algorithm and tested with the proposed dual Bi-LSTM layers network, yielding comparable classification accuracy as conventional micro-Doppler inputs, along with over 90% saving in computational cost. For the decision level fusion, the classification results using Doppler information with the proposed network are combined by a Naïve Bayes Combiner (NBC), and the fusion results indicate an accuracy gain of 5.8% and 7.3% with respect to the best single radar case in individual and sequential gaits. Compared to the conventional classifier (SVM and Random Forest Bagging Trees), the proposed range trilateration and NBC decision fusion approaches improve the average classification accuracy of the individual gaits for about 20% and 17%, respectively. Furthermore, the fusion accuracy for the two proposed signal and decision methods achieve 91% and 93% with validation by a ‘Leaving One Participant Out’ (L1O) scheme.

6.2.1 Experimental Setup

The gait measurements were recorded from a distributed radar sensor network in the Computer Intelligence for Radar (CI4R) Lab at the University of Alabama, involving 14 participants (3 female and 11 male) ageing from 20 to 45. The radar sensor network is comprised of one 25 GHz FMCW radar with 2 GHz bandwidth (Ancortek 2500B) and three 7.3 GHz Ultra Wide-Band (UWB) impulse Doppler radars with approximately 1.5 GHz bandwidth (Novelda Xethru X4M300), whereas the PRF for the two different types of radar are set to 1 KHz and 0.5 KHz, respectively. The FMCW radar utilizes a monostatic structure with two horn antenna as the transmitter and receiver, the transmitted power is about 19 dBm, for the UWB Doppler radar, the microstrip antennas are integrated with the processor and signal generator on one chip. The radars within the network characterize the human gait patterns in three different spatial perspectives, as shown in figure 6.16, where the Ancortek radar and one of the Xethru radar (red box in figure 6.16) are placed in front of the participants, another Xethru (purple box in figure 6.16) is fixed on the celling with about 2.2 m height to the ground and an elevation angle of roughly 45° to the centre of the experimental zone, and the last Xethru (yellow box in figure 6.16) is positioned at the right hand side of the participant. Since Xethru radar has a beamwidth of 65 degrees, the locations of radar sensors are carefully chosen to avoid the potential interference between each other (e.g. prevent radar A from receiving the RF signal from radar B).

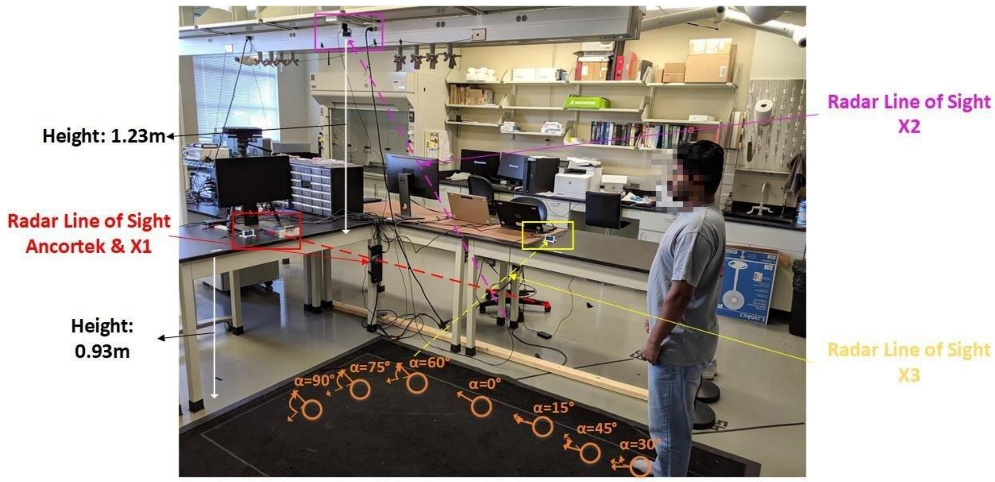


Figure 6.16 Experimental setup including line-of-sights of different radar systems and walking trajectory (Red: radar in front of participants, Purple: radar on the ceiling, Yellow: radar on the right hand side)

All the radar sensors are connected to a laptop for powering and data transferring via USB cables, whereas the sensor network is synchronized by adding a time delay as compensation since Ancortek and Xethru radar have different waking up time (the waking up time has been tested for multiple times and the variance between each test is small enough to ignore). Beyond that, the pressure data from the intelligence mattress (GaitRite) on the floor in figure 6.16 are used as ground truth. The radar data collection platform is programmed by C++ with all the MATLAB scripts complied in, as shown below in figure 6.17.

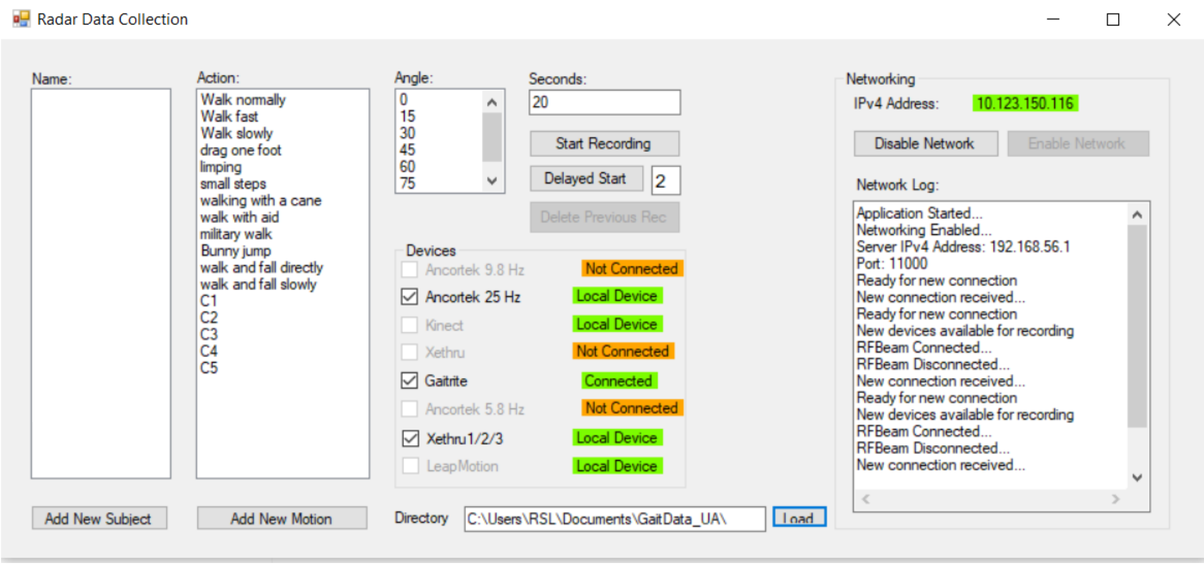


Figure 6.17 The radar data collection GUI (Ancortek FMCW radar, three Xethru radar sensors and pressure mattress connected)

The gait styles measured in the experiments are listed below in table 6.3, where individual gaits include walking with different speed (the participant is asked to walk about 0.5 m/s in slow speed, double in normal, and then double again in fast), dragging one foot while walking, moving with small steps, walking with aids, jumping back and forth, as well as some joint gaits (gaits followed by a fall event). In the experiment, the participants are asked to perform 20s elliptical loops in different gait styles (with the trajectory shown in figure 6.16), whereas in 'G11' and 'G12', two kinds of falling are following a short period of walking (12s approximately) to attempt to simulate the sudden loss of consciousness and progressive exhaustion and fall of elderly people, respectively.

Individual gaits:
1.Walking normally
2.Walking quickly
3.Walking slowly
4.Dragging one foot
5.Limping with an orthopedic cast
6.Small steps
7.Walking with a cane
8.Walking with a walker
9.Military walking
10.Bunny jump
11.Walking and direct fall
12.Walking and controlled fall
Sequential gaits:
A:20s walking normally; 20s walking slowly; 5s controlled fall
B:20s walking normally; 20s dragging one foot; 5s controlled fall
C:20s walking with a cane; 5s controlled fall
D:20s bunny jump; 20s walking normally; 20s dragging one foot
E:20s walking with a walker; 20s walking slowly; 20s walking normally

Table 6.3 List of the 12 individual gaits and 5 (A-E) sequential gaits

Additionally, five different sequential gaits (A-E) are listed in table 6.3, where the content and length of each are different. Compared to individual gaits, sequential gaits were recorded in an uninterrupted continuous manner, producing the natural transitions between two different gait styles. The purpose of sequential data is to simulate a gradual process, for instance, sequence A for the increased falling risk related to the change of gait patterns and sequence E for the rehabilitation from a fall accident.

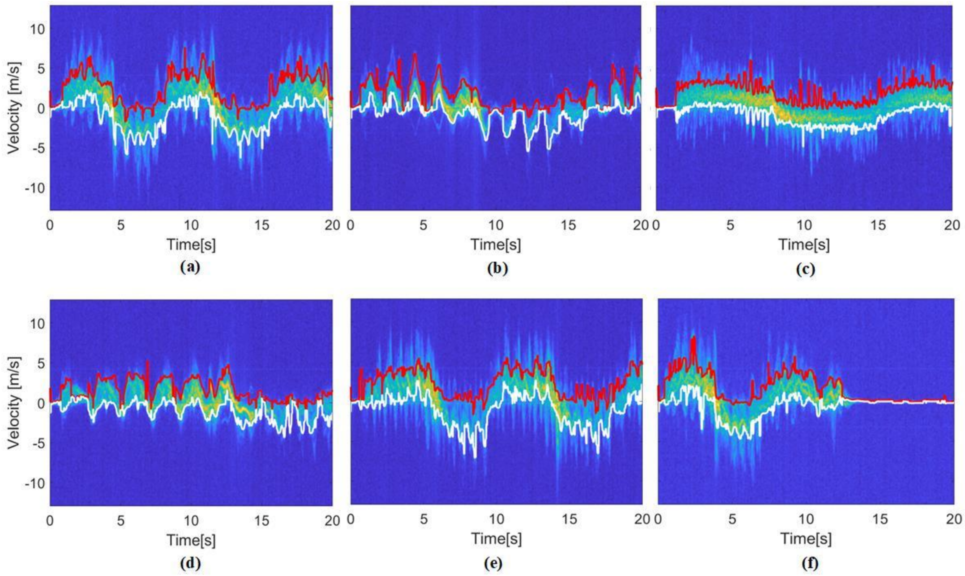


Figure 6.18 Ancortek FMCW radar spectrograms. The spectrogram from (a) to (l) correspond to the gait from 'G1' to 'G12' in Table 6.3. Red line: upper envelope, white line: lower envelope.

For the individual gaits, each participant was asked to perform each gaits for 3 times, so the dataset contains 504 (14x12x3) samples, whereas for the sequential gaits, totally 71 (2x3x1+7x5x1+3x5x2) samples are collected from 12 participants.

Figure 6.18 shows the radar spectrograms of different gaits, where the positive Doppler shift represents the stride towards the radar and vice versa. The upper (red line in figure 6.18) and lower envelope (white line in figure 6.18) denote the maximum positive and negative Doppler as a function of time, respectively. The testing subject will lie on the mattress to simulate loss of consciousness after falling down, expressed as the flat part at the end of last two spectrograms (k and l in figure 6.18).

6.2.2 Conventional Classifier Results for Individual Gaits

The relationship between the classification accuracy and the number of features dropped by SBS algorithm is presented in figure 6.19, with respect to different radar sensors. In terms of using a SVM classifier with SBS, the FMCW Ancortek radar yields the highest performances with approximately 69% accuracy based on 20 out of 57 features (37 features are removed from the initial feature set by SBS). Compared to using the entire feature set, Xethru P1 to P3 achieve an accuracy gain of 4.5% to 7% with the help of SBS. However, it is reported by figure 6.19 (b) that combining the SBS with a RFB classifier is less powerful than SVM in the perspective of accuracy improvement. The accuracy gain with respect to the case without SBS is only 2-3%. RFB is considered as a cluster of decision tree which embeds the preliminary ranking for features during the training of classifier, and as a consequence of that, SBS repeats the feature selection process and this reduces the robustness of the selected feature set. Figure 6.19 (c) indicates the accuracy profile through the feature fusion between all the radar sensors, where the volume of improvement is 12% and 3.3% for SVM and RFB classifier based on 64 and 109 features respectively (out of 228 features). Through the feature fusion and selection, the optimal result of conventional classifier yields about 80.6% in classification accuracy using 'L1O' cross-validation method.

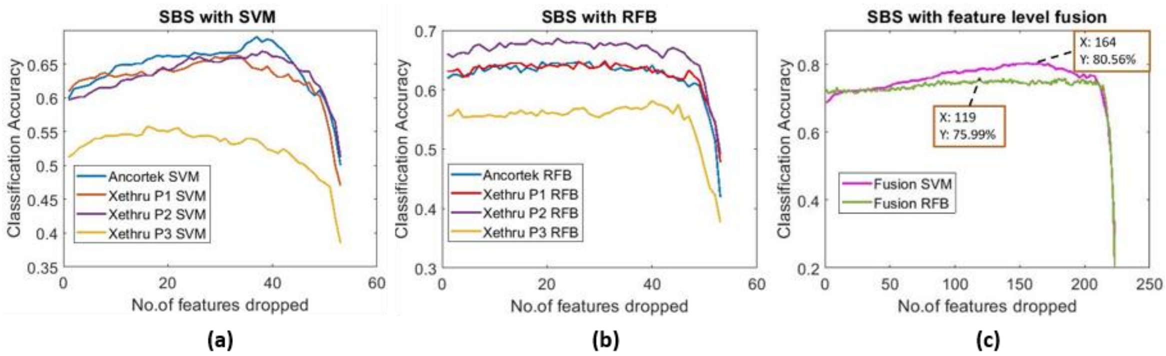


Figure 6.19 The classification accuracy as a function of the number of features dropped via SBS algorithm: (a) with SVM classifier on individual radar data, (b) with RFB classifier on individual radar data, and (c) with both SVM and RFB on feature fusion

6.2.3 Classification Results of Individual Gaits

The Bi-LSTM network used in this work has the same architecture as described in previous Section 6. It contains an input layer, dual Bi-LSTM layers, a softmax layer and an output layer. Typically, Bi-LSTM based network is usually used for sequential classification problems, whereas the individual gait is composed by repetitions of same motion patterns (for instance swing of limbs). Thus, the proposed network can treat the individual gaits as simplified sequential gaits without considering other gait styles and transition parts between them.

In this work, three different types of information are used as the inputs of the proposed network, notably, micro-Doppler signature, range-time matrix and 2-D position of testing subjects calculated by trilateration.

- For the micro-Doppler signature, Doppler centroid and bandwidth along with the upper and lower envelope of the spectrogram are extracted as a function of time. Doppler centroid and bandwidth express the centre of mass of the human torso and the energy spread around this respectively, while the envelope of the micro-Doppler signature captures the velocity change related to the movement of human limbs. Similar features have been extracted and applied in the field of arms motions recognition as well as gesture classification. Additionally, those four features are utilized as parallel input channels for the proposed Bi-LSTM network, referred to as micro-Doppler Bi-LSTM. The size of an input channel is equal to the total number of time steps in one individual gait.

- For the range-time matrix, the average distance between the radar and testing subject (red line in figure 6.20), and the range extent (white line in figure 6.20) caused by the swing of limbs are calculated from the range-time profile along each time bin. The formulas for those two range features are shown below in equation 6.1 and 6.2:

$$D_a(i) = \frac{\sum D(j)R(j,i)}{\sum R(j,i)} \tag{6.1}$$

$$E_D(i) = \sqrt{\frac{\sum_j (D(j) - D_a(i))^2 R(j,i)}{\sum_j R(j,i)}} \tag{6.2}$$

Where $D_a(i)$ and $E_D(i)$ denote the average distance and radar extent at ith time bin, respectively. $R(j, i)$ denotes the component (row j, column i) of the range-time matrix, whereas the $D(j)$ represents the distance value of j^{th} range bin.

The range profiles for six individual gaits are shown in figure 6.20, where the difference across the gait patterns is not as visible as Doppler spectrogram. However, an additional mathematical transform (e.g. STFT, Wavelet transform) on top of the range-time matrix is essential to generate the Doppler spectrogram. Those transforms are extremely computational intensive especially on a portable device like a smart watch or mobile phone. Therefore, it is interesting to test whether using the range information to feed the proposed network can reach similar performance as micro-Doppler Bi-LSTM.

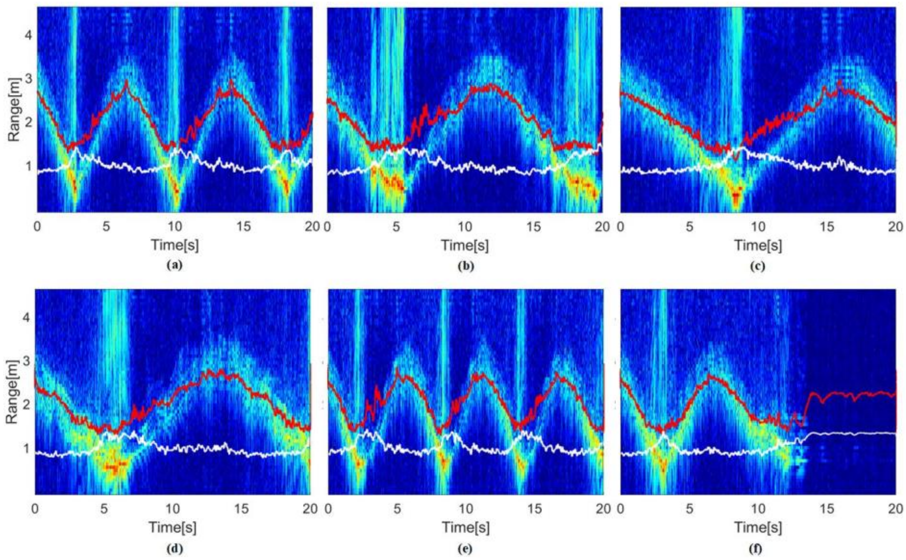


Figure 6.20 Range-time maps for several gaits: (a) walking; (b) dragging foot; (c) small steps; (d) walking with aid; (e) bunny jump; (f) walking and controlled fall. Red line: average radar-subject distance. White line: range extent around average value

- For the range trilateration, the range information from Xethru radars (Xethru P1, P2 and P3) at three different positions can be fused by trilateration algorithm to obtain the 2-D coordinates of the central torso of testing subject. In our case, due to that the Xethru P2 is installed on the ceiling, the R_2 in equation 4.27 is not equal to the Euclidean distance from the corresponded radar to the subject. Instead, the projection of the average distance on a 2-D horizontal plane is considered as the actual length of R_2, as shown below in equation 6.3:

$$R_2^{'} = \sqrt{R_2^{\ 2} - h^2} \tag{6.3}$$

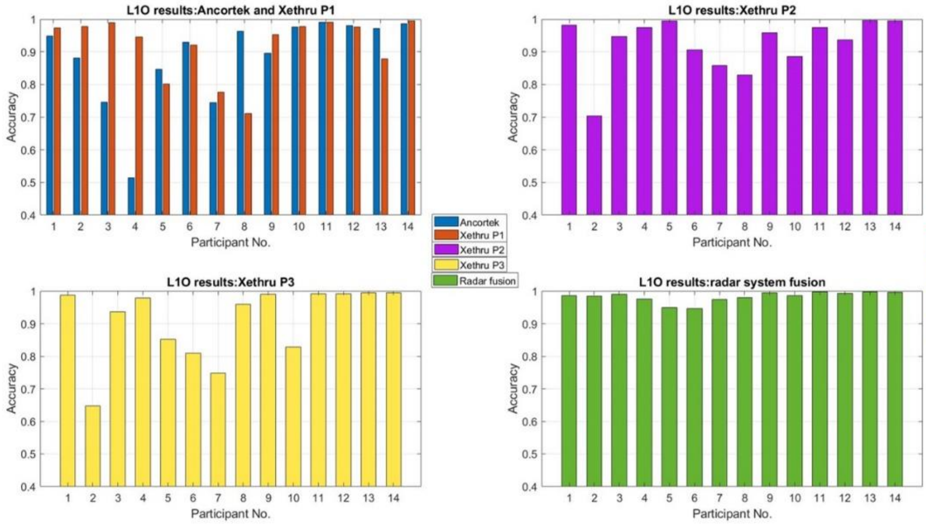
  Where h denotes the height of Xethru P2 with respect to the desk, in our case, h is equal to 1.23m. Because of the limitation of range solution, the average distance of single radar sensor will have some measurement errors (see details in Chapter 4.3.1), trilateration-based signal level fusion algorithm significantly increases the precision of localization, which is beneficial to the following training and testing of the proposed Bi-LSTM network. Beyond that, the range trilateration saves about 90% of processing time than micro-Doppler Bi-LSTM in the MATLAB implementation.

Hyper-parameters	Micro-Doppler as input	Range-Time as input	Range tri-lateration as input
Training, validation and testing ratio	15:0:1	15:0:1	15:0:1
Number of inputs	4	2	2
Number of classes	12		
Training and testing scheme	Leaving one participant out (L1O)		
Mini-Batch size	8		
Max epochs	200	400	200
Initial learning rate	1e-3		
Learning rate drop period	50	100	50
Learning rate drop fact	0.1		
Optimization function	Adaptive Momentum Estimation		

Table 6.4 The hyper-parameters for the proposed Bi-LSTM network

The hyper-parameters for the training of the proposed Bi-LSTM network are summarized in table 6.4, where the initial learning rate is fixed as 0.001, it will drop 90% after the learning rate drop period, which is varied by the type of input. For the range-time matrix, as the training accuracy reaches the plateau much slower than the micro-Doppler and range trilateration, the max epochs is set double as two other types of inputs for the fully convergence of the proposed network. The optimization function is chosen as 'Adam' for leveraging the advantages of RMS propagation and Momentum.

The classification performance of micro-Doppler Bi-LSTM for each testing subject in the 'L1O' training and testing scheme are shown in figure 6.21, with respect to different radar sensors. The average accuracy across all the participants are similar for each cases, where Xethru P2 (radar on the ceiling) outperforms other radar sensors with around 92.4% mean accuracy due to its position and looking angle. However, in terms of the minimum accuracy (the 'worst' participant who yielded the lowest average accuracy in the 'L1O' test) and accuracy variance (standard deviation across all the participants), which are highly related to the robustness of the radar, Xethru P1 and P2 are much better than Ancortek and Xethru P3. It is interesting to note that significant variability of classification accuracy for some participants when tested with same type of radar sensors at different positions (e.g. participant No.2 and No.10 for all Xethru radars), or radar operating at different central frequency and bandwidth but placed at the same positions (e.g. participant No.2, No.3, No.4 and No.8 for Ancortek and Xethru P1). For the fusion, a NBC is used to boost the accuracy for each participant (especially the one with low classification performance) by leveraging the advantages from all the radar sensors. Not only mean accuracy, but also the lowest boundary and standard deviation are improved, the gain for the aforementioned metrics are 8.1%, 23.5% and 6.09% with respect to the best case of single radar.



	Mean Acc(%):	Min Acc(%):	STD(%):
An	88.4	51.4	13.49
X1	91.9	71.1	9.19
X2	92.4	70.4	8.27
X3	90.8	64.8	11.2
Fus	98.2	94.6	2.18

Figure 6.21 The classification performance of the Doppler Bi-LSTM with single radar and fusion for *individual gait data*. Different colour bars indicate results from the different radar sensors used in isolation or with fusion (green)

.

Figure 6.22 shows the dependence of micro-Doppler Bi-LSTM classification results with the radar aspect angle. In this test, the prediction results of the 12 individual gaits from one random participant are concatenated one by one to construct a 240s (12x20) long data sequence (without transitions between different gait patterns). It is observed that radar aspect angle changes in a periodic manner, and this is caused by that the testing subjects move in an elliptical route. Moreover, for radar sensors at different positions, the value of aspect angles varies too. Theoretically, the strength of the Doppler signature will be significantly attenuated with a high aspect angle (because Doppler signature is a function of cosine aspect angle), and this potentially compromises part of radar classification results, in particular, Ancortek (e.g. 122-130s) and Xethru P3 (e.g. 225-230s). Many misclassifications (red line in figure 6.22 on ‘False’) happen at different moments when using different radar sensor separately. For the Ancortek radar, the main errors are in the period 140-160s and 180-240s for the last six individual gaits including falling, whereas the Xethru P2 misclassified the events between 120s and 150s. Apart from that, some minor classification errors also occur in Xethru P1 and P3 (mainly between 200s and 240s). However, with the help of fusion, most of the misclassified events are corrected except some very minor errors around 210s and 230s. This suggests that at least one of the radar sensors in the network is able to provide ‘good’ view of the testing subjects in terms of aspect angles, yielding strong Doppler signature that classifier prefers. Also it can’t be ignored that the misclassification of Xethru P1 at 100s and 230s is less than fusion, this proves that fusion not only gain the strength of participated sensors but also ‘inherits’ their shortcomings.

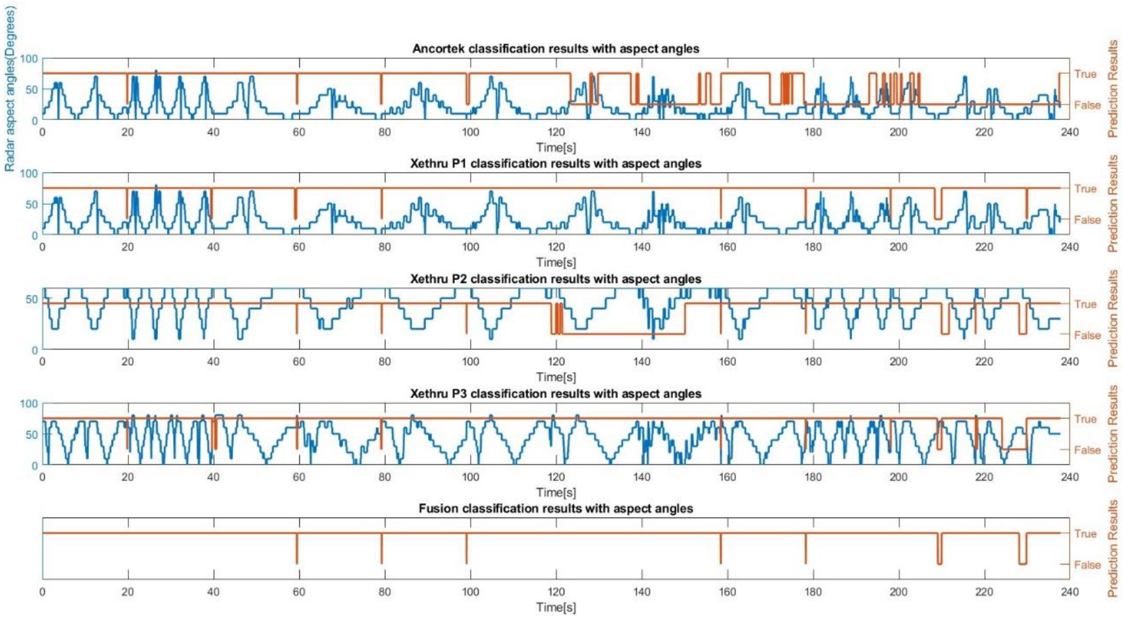


Figure 6.22 The prediction results of the Doppler Bi-LSTM network with respect to the aspect angle for *individual gait data*. From top to bottom: Ancortek, Xethru P1, P2, P3, and radar fusion. Aspect angle values reported in blue; network prediction results as binary true-false values in orange

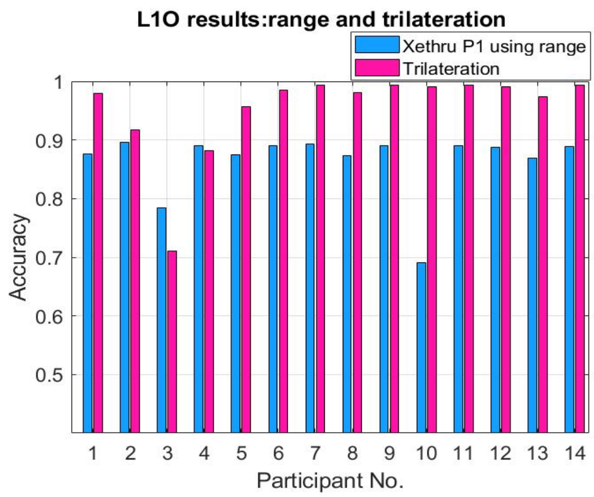


Figure 6.23 The classification accuracy for range-only information from Xethru P1 and trilateration for *individual gait data*

The comparison between ‘L1O’ test results using only Xethru P1 range information and range trilateration is shown in figure 6.23, where the average accuracy across the 14 participants are 84.4% and 95.3% for this two cases respectively. Compared to the Doppler classification results in figure 6.21, proposed network trained with range data from single radar is not as powerful as using the micro-Doppler information, the accuracy difference is approximately 7.5% for the case of Xethru P1. However, most of the participants receive an accuracy gain of roughly 2-30% through using trilateration between multiple Xethru radar sensors. It is interesting that an accuracy drop occurs on participant No. 3, and this suggests that other two radar sensors experiences higher measurement errors or noises during the data collection of this subject, as the precision of localization via trilateration depends on three radar sensors.

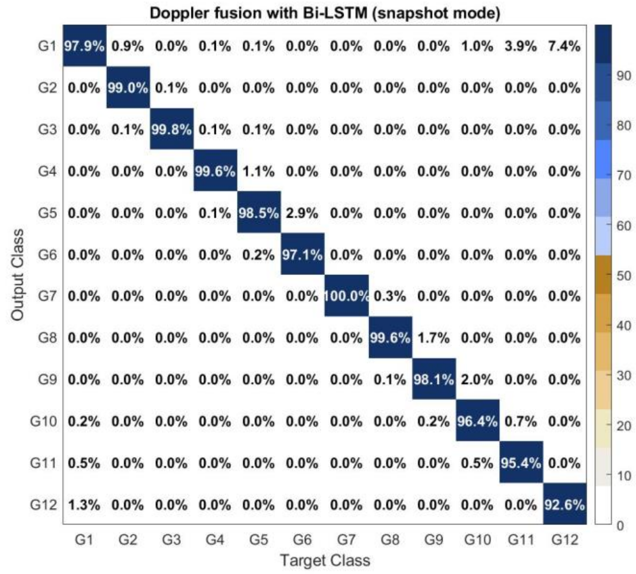


Figure 6.24 Confusion matrix of micro-Doppler Bi-LSTM fusion using a NBC for *individual gait data*

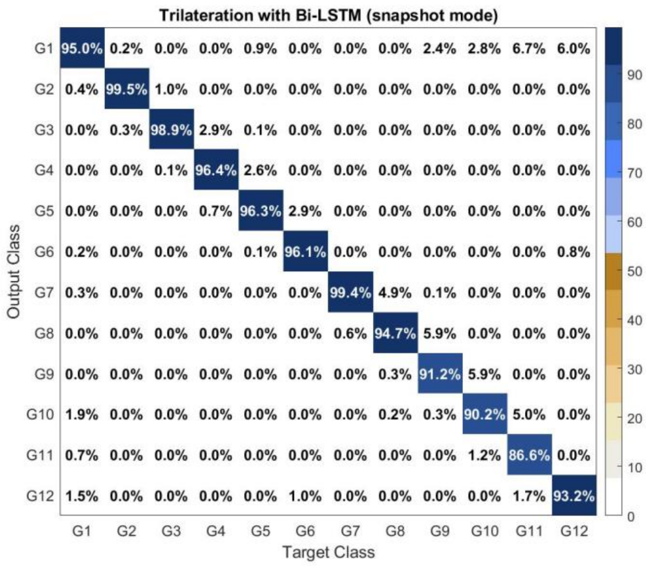
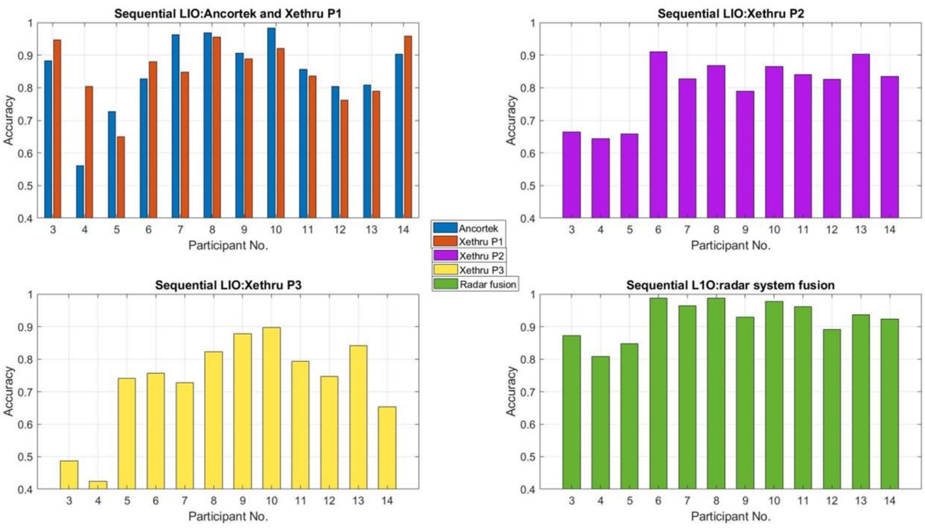


Figure 6.25 Confusion matrix of range information fusion using trilateration for *individual gait data*

Figure 6.24 and 6.25 show the classification results of radar fusion using micro-Doppler information and range trilateration, respectively. For the micro-Doppler information, most of the classes yield a correctly classified rate over 95%, whereas the fall recognition rate is a bit lower than the average. It is reported that some minor misclassifications take place between certain pairs of individual gait (e.g. G1 'walking with normal speed' and G2 'walking with fast speed', G5 'limping with a cart' and G6 'walking with small steps'). Those pairs of gaits are designed to be similar for creating more 'artificial' classification challenges. The sensitivity for G9, G10 and G11 are reduced approximately 6-9% in the case of range trilateration, whereas the range trilateration surpasses the Doppler fusion about 0.6% in terms of the most crucial fall identification. The false alarms of falls are only distributed between G1 and G11, both containing a long-term walking like G12. Besides that, the testing subject will be usually hesitated after hearing the order to fall, leading to an offset between the actual falling moment and the ground truth (at about 12s the falling order was given and the subject performs at 13s).

6.2.4 Classification Results of Sequential Gaits

In this sub-section, the real sequential gaits involving the natural and seamless transitions between different gait styles are processed with the proposed Bi-LSTM network. Those transition parts need to be taken into account carefully, including the processes of gradually speed change (e.g. from walking normally to walking slowly in sequence A and opposite transition in sequence E), starting some abnormal gaits due to the deterioration of personal health conditions (e.g. from normally walking to dragging one foot in sequence A), triggering a fall event during walking with difficulties (e.g. from walking slowly to fall in sequence A and from walking with a cane to fall in sequence C) and the slowly recovered walking ability after a fall accident (e.g. from walking with a walker to walking slowly in sequence E). Beyond that, for sequential classification, same information for the analysis of individual gaits, namely, micro-Doppler, range information from Xethru P1 and range trilateration, are used as inputs of the proposed network respectively.



	Mean Acc(%):	Min Acc(%):	STD(%):
An	85.4	56.1	11.85
X1	84.8	64.9	9.24
X2	82.2	64.3	9.49
X3	75	42.4	14.59
Fus	92.7	80.8	5.82

Figure 6.26 The classification performance of the Doppler Bi-LSTM with single radar and fusion for *sequential gait data*. Different colour bars indicate results from the different radar sensors used in isolation or with fusion (green)

The ‘L1O’ testing results of micro-Doppler Bi-LSTM for the sequential gaits are summarized in figure 6.26 in terms of single radar and fusion. Different from the performance of individual gaits in figure 6.21, Ancortek outperforms other radars with an accuracy equal to 85.4%, followed by the Xethru P1 with 0.6% lower. There is a significant accuracy drop (approximately 10% in mean accuracy compared to the best Xethru) in Xethru P3, especially on the participant No. 3 and 4 with a classification accuracy lower than 50%. The radar fusion through NBC improves the correctly classified rate of about 1.5%-25% for most of the participant, beyond that, with respect to the best case in single radar, other metrics including the worst case scenario of the classification accuracy (about +15.9%) and the standard deviation across the participants (about -3.4%), share the gain as well.

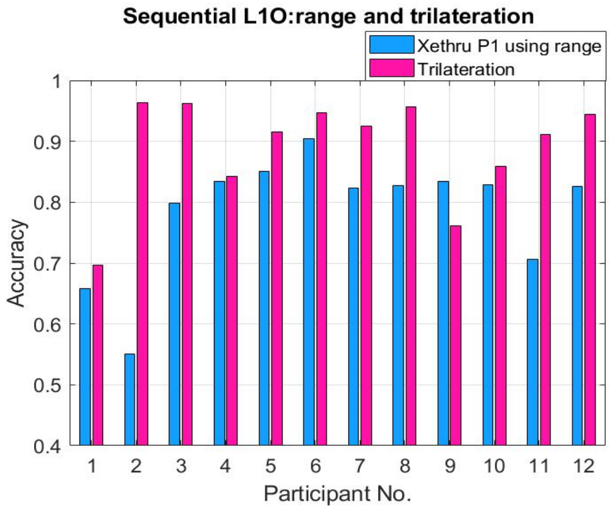


Figure 6.27 Classification accuracy for range only information from Xethru P1 and trilateration for *sequential gait data*

Figure 6.27 compares the classification results of using the range information from only Xethru P1 and the range trilateration by fusing range data from all three Xethru radars (Xethru P1- P3). The average accuracy for using the range data of Xethru P1 is approximately 83.7%, whereas the worst case scenario (participant No.2) across all the participants yields an accuracy of about 55%. Through the range trilateration, the average accuracy is increased by nearly 7.3%, with the minimum accuracy of the previous case rising to 96%. Most of the participants receive an accuracy gain of 2-41%, whereas an accuracy drop of roughly 7.5% is reported by participant No. 9. This could be solved by further fusion with Doppler information since the performance of this subject in figure 6.26 is much higher than range trilateration.

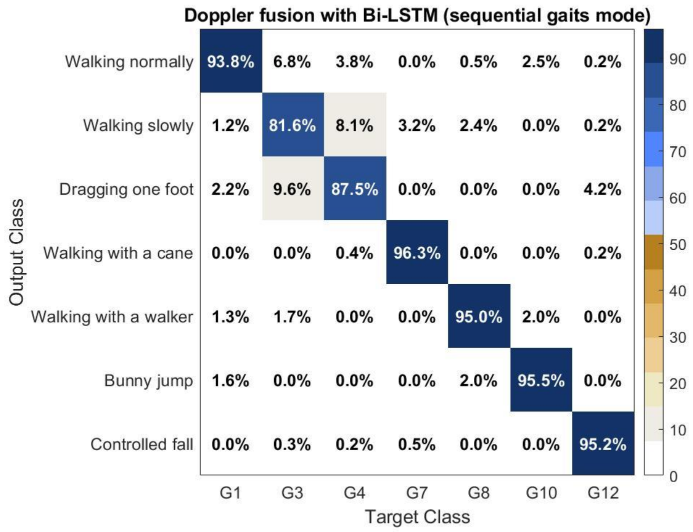


Figure 6.28 Confusion matrix of micro-Doppler Bi-LSTM fusion using a NBC for *sequential gait data*

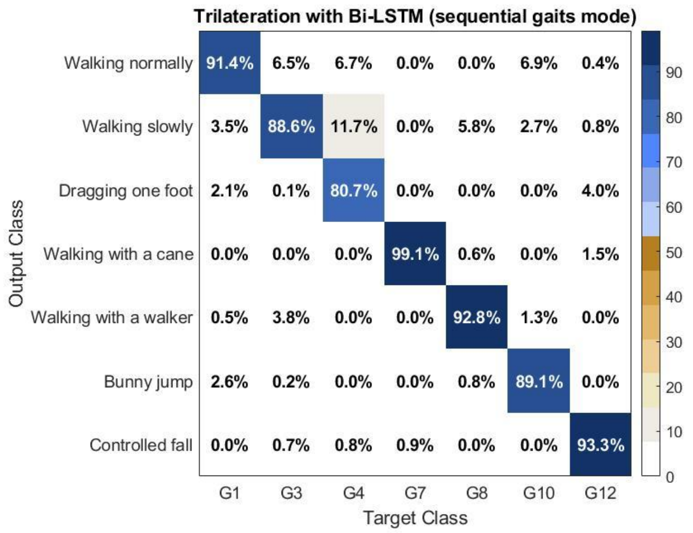


Figure 6.29 Confusion matrix of range information fusion using trilateration for *sequential gait data*

Figure 6.28 and 6.29 show the confusion matrices for the classification results of sequential gaits under 'L1O' training and testing scheme, with respect to Doppler fusion and range trilateration. In figure 6.28, the misclassified events for Doppler fusion are mainly across 'walking normally', 'walking slowly' and 'dragging one foot', which are quite similar in terms of the moving patterns. Apart from that, the sensitivity for the rest classes are over 93%, for the detection of the controlled falls (falling slowly), few fall events are misclassified to 'dragging one foot', 'walking with a cane' and 'walking slowly' because those gaits are neighbours in sequence A, B and C. Compared to the Doppler fusion, range trilateration in figure 6.29 increases the correctly classified rate of 'walking slowly', whereas more than 10% of the 'dragging one foot' are still misclassified, leading to a decrease in the sensitivity of this class. For the G7 'walking with a cane', the classification accuracy improves to nearly 100% with minor misclassification to the controlled fall. It is also reported that the performance of 'bunny jump' is reduced by about 6.4%, this can happen because the difference between the two gait styles in terms of subject positions is not as significant as using micro-Doppler. Additionally, the fall detection is less powerful than Doppler fusion especially in the number of false alarms. Conclusively, it is potential to combine the Doppler fusion and range trilateration results for seeking subsequent improvement, with the price of tripling the computational loads.

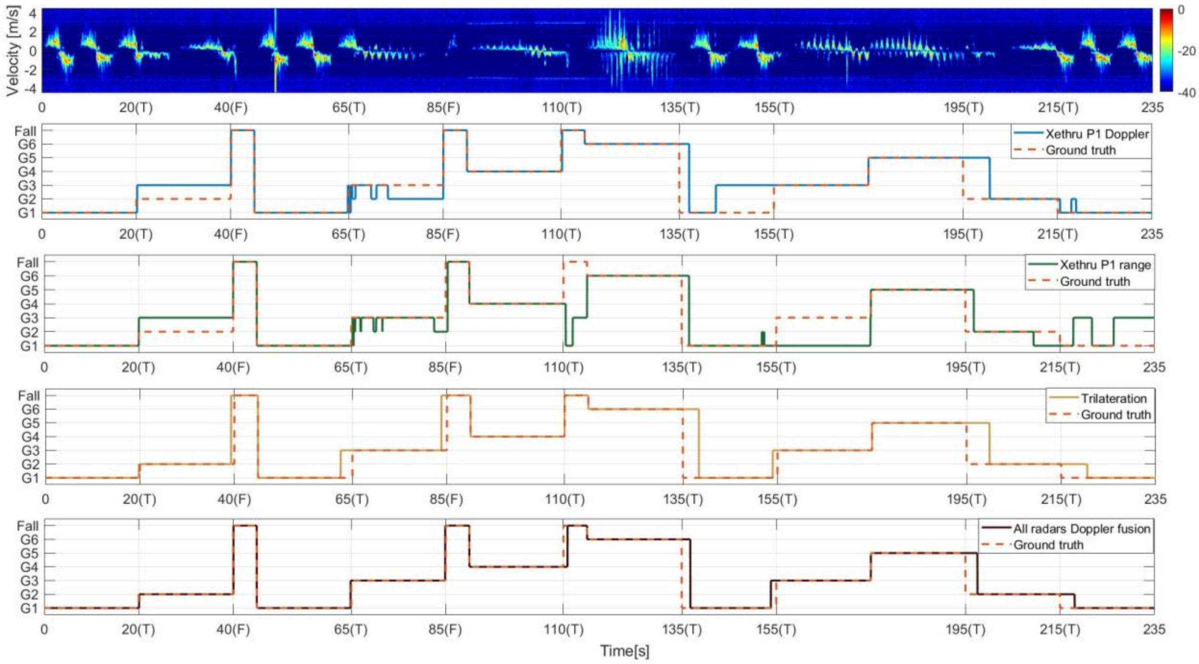


Figure 6.30 Predictions vs ground truth for sequential gaits performed by a participant. From top to bottom: Doppler spectrogram for all sequential gaits recorded with Xethru P1; Xethru P1 results using Doppler; Xethru P1 results using range; signal level range fusion using trilateration; decision level Doppler fusion of all radar sensors. G1= normal walk; G2= slow walk; G3=dragging foot; G4= walk with cane; G5= walk with walker; G6= bunny jump.

Figure 6.30 compares the prediction results of sequential gaits in terms of different types of inputs and fusion methods for the proposed Bi-LSTM network, in particular, Doppler and range data from Xethru P1-only, and information fusion between multi-radar sensors through NBC and trilateration. The Doppler spectrogram of Xethru P1 composed by cascading the 5 sequential gaits (A-E in table 6.3) of one random participant is shown at the top of figure, with respect to the ground truth (red dash line), where markers 'T' and 'F' denote the moments of gait transition and fall event, respectively. There are three different types of classification errors in the prediction results, the first type is single spike on the classifier output (e.g. 220-223s for using Doppler from Xethru P1-only and 150-152s for using range information from same radar sensor), the second type is the rapid oscillation (multi-spikes in a short period) on the prediction results (e.g. 65-85s for both single radar cases), and the third type is long-term classification errors (e.g. 20-40s for both single radar cases and 142-155s for using Doppler from Xethru P1 independently). Fusion of Doppler information by a NBC and combination of range information through trilateration, both correct most of the classification errors, especially the second and third types in the single radar case. It is nice to see that the prediction line fits the ground truth well after applying signal and decision level fusion, only left some offset of the transition parts (135s, 155s, 195s and 215s for the best case Doppler fusion).

6.3 Summary of the Chapter

This chapter extends the analysis of snapshot activities discussed in chapter 5 to continuous activities and gait streams.

In the continuous activity part (chapter 6.1), a novel dual Bi-LSTM layers network is proposed to identify the transitions of activities and its classification results are compared with conventional sliding widow-based method. Meanwhile, a hybrid fusion framework comprised of soft fusion with different sensor weights and a hard combiner is used to fuse the wearable IMU sensor and FMCW radar, and the results of proposed hybrid fusion outperforms using soft or hard fusion individually.

In the continuous gait part, micro-Doppler information of distributed radar sensors (three UWB impulse radars and one FMCW radar at different spatial locations) are combined with the help of NBC (best hard combiner proved in chapter 6.1) and the Doppler fusion results show significant improvement compared to best solo case, whereas a trilateration-based signal level algorithm is proposed to combine the range data of three UWB impulse radars and the range fusion results show comparable performance as Doppler fusion with a reduction of computational budgets.

7 Dynamic and Static Gesture Classification

This chapter presents a sensor fusion framework for human micro-gesture classification through combining wearable pressure sensor array and an UWB Impulse Doppler radar. The wrist-worn pressure sensor array and UWB Doppler radar are used to recognize static and dynamic gestures by a Quadratic-kernel SVM separately. Prior to the fusion, Sequential Forward Selection (SFS) is used to search the optimal feature combination from the original feature set for improving the individual sensor performance. Subsequently, two fusion methods where one sensor is acted as ‘enhancer’ of the other are tested. For the first method, confidence level of the classifier trained by Doppler radar data is combined with the same type of information from pressure sensor for maximizing the static hand gestures classification performance. For the second method, the PSA acts as an ‘Enhancer’ for radar to improve the dynamic gesture classification performance, where different weights of the ‘Enhancer’ sensor in the soft fusion process have been evaluated and compared in terms of classification accuracy. Moreover, a hard fusion approach, in particular, Naïve Bayes Combiner (NBC) is used to fuse the prediction results of both sensors based on their previous classification performance. For moving our simulation closer to the actual product test, ‘L1O’ cross-validation method is chosen to test one unknown participant with the model trained by data from others, demonstrating that this fusion framework for static and dynamic gestures yields approximately 15% improvement in classification accuracy in the best cases.

7.1.1 Experimental Setup

The data collection was taken place in meLAB laboratory at University of Glasgow with one UWB Impulse Doppler radar (Xethru X4M300) and wrist-worn Pressure Sensor Array (PSA). The UWB Doppler radar has an operating frequency equal to 7.3 GHz with more than 1.5 GHz instantaneous bandwidth at -10dB, leading to a range solution about 0.1m. The PRI (Pulse Repetition Interval) is set to 0.005s. The Xethru radar was placed on a plastic desk in figure 7.1 with around 1.2m to the ground, pointing the centre of the subject hand at a distance of nearly 0.4m. Additionally, the radar was connected to the laptop for data transfer and powering, whereas the received radar data is in a format of complex numbers for signal amplitude and phase. The PSA bracelet in figure 7.2 is comprised of five Force Sensitive Resistor (FSR402) nodes and an Arduino DUE board are connected with those nodes as an ADC and temporal buffer to convert the pressure level to the corresponding voltage. Those voltage data were then transferred to the laptop with a sampling frequency of approximately 50 Hz through an interface in LABVIEW. The data processing of both radar and PSA were finished in MATLAB.

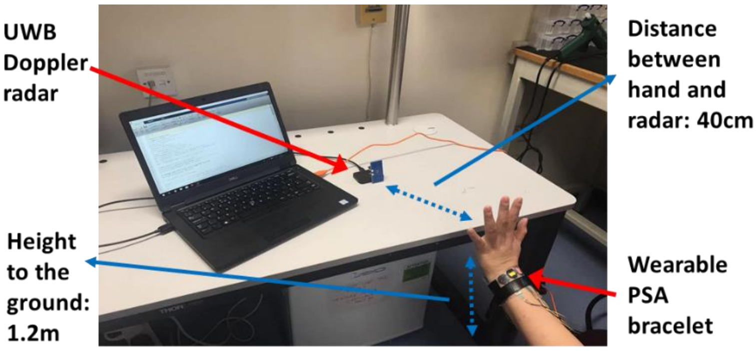


Figure 7.1 Experimental setup (on the participant wrist: wearable PSA bracelet, blue chip on the box: UWB impulse-Doppler radar)

The dataset includes ten male subjects, aged from 21 to 36. During the experiment, the participants were requested to perform four different kinds of static gestures with their left hand, namely, number 0 (fist), number 1, number 2 and number 5 (open hand), and some transitions between pairs of these static gestures. Each static gesture lasts about 4s and then transfers to the following static gestures, the whole gesture sequence contains seven static gesture samples along with six gesture transitions (0-5, 5-2, 2-0, 0-2, 2-1, 1-5). Hence, the length of the gesture sequence for one subject is roughly 28s. The total number of the static and dynamic gesture samples are 70 (10x7) and 60 (10x6) respectively.

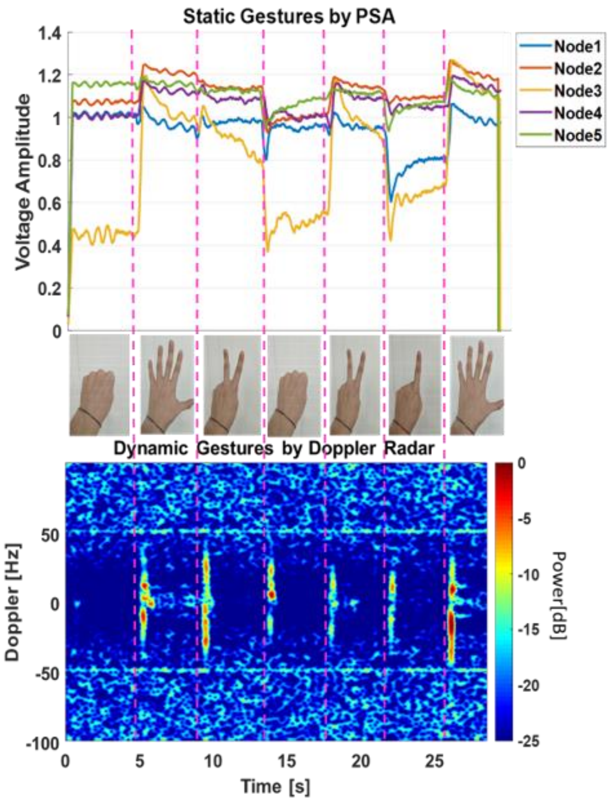


Figure 7.2 Pressure level from PSA (top), radar Doppler signature (bottom)

The readings of the five pressure sensor nodes on the PSA bracelet are shown in the top of figure 7.2 as voltage amplitude, whereas the corresponded radar Doppler signature are presented in the bottom of figure 7.2. When the hand is static, each pressure node generates a different flat response in terms of amplitude, since they are placed at different portions of the wrist tendon to percept the pressure change. The radar information is based on Doppler-effect, in particular micro-Doppler, therefore, it is more sensitive to the moving targets, for our case, the transition part between two static gestures.

7.1.2 Feature Extraction and Selection

Each participant has different wrist size and strength of muscle, apart from that, positions of sensor nodes are also varied by person. Hence, there are significant difference between the amplitude of PSA measurements from one participant to another. Prior to the feature extraction, the raw data are normalized to the same scale by subtracting the mean value and dividing the standard deviation.

PSA Features	No.
2-d Mean of the voltage amplitude	1
Max of the voltage amplitude	1
Min of the voltage amplitude	1
Range of the voltage amplitude	1
Mean of the cross-correlation between data from Node x and Node y	10
Standard deviation of the cross-correlation between data from Node x and Node y	10
The difference between the mean of first 50 data points and the mean of last 50 data points (only for dynamic gestures)	5

Table 7.1 List of the PSA features

For the PSA bracelet, 24 numerical features inspired by our work in wearable sensors are extracted from the standardized dataset to characterize both static and dynamic gestures, listed in table 7.1. These include: the two dimensional mean value, maximum, minimum and range of all sensor nodes data for each static gestures (4 features in total), as well as the mean and standard deviation of the correlation function to represent the relationship between pairs of sensor nodes (20 features in total as there are 5 sensor nodes, 10 different combos). For increasing the robustness of the dynamic gesture classification, 5 more features (29 in total), 1 for each resistor of the PSA, have been utilized. These are the difference between the mean of first 50 data points and the mean of last 50 data points in order to estimate the pressure difference between the previous and next static gestures.

For the Xethru radar, 12 physical features involving Doppler centroid and bandwidth (ID.1-12 in table 4.1) along with 8 SVD-based features (ID. 24-27 and 36-39 in table 4.1) are extracted to train the classifier. The Doppler centroid and bandwidth represent the central mass of the palm movement and the power surrounding the central mass, respectively. In our case, the bandwidth is more relevant to the fingers motions.

The time slots in the data sequence for feature extraction of the static gestures are 2-4s, 6-8s, 10-12s, 14-16s, 18-20s, 22-24s, and 26-28s, which can avoid the sudden change of the voltage amplitude. The data between time slots of two static gestures (4-7s, 8-11s, 12-15s, 16-19s, 20-23s, 24-27s) is used to generate the features for dynamic gestures, as the dynamic gestures are more correlated with significant increase or decrease on the pressure level.

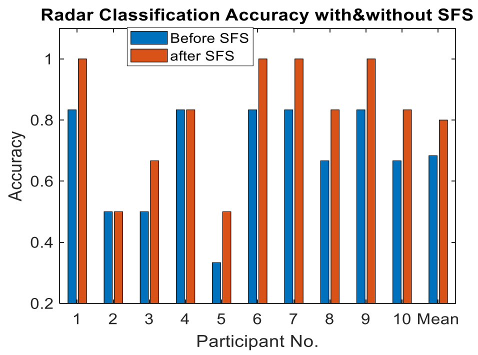


Figure 7.3 Radar SFS results with dynamic gesture

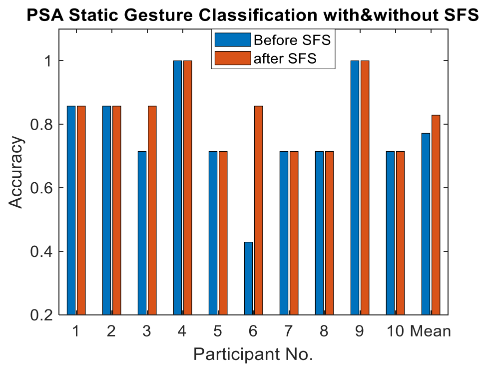


Figure 7.4 PSA SFS results with static gesture

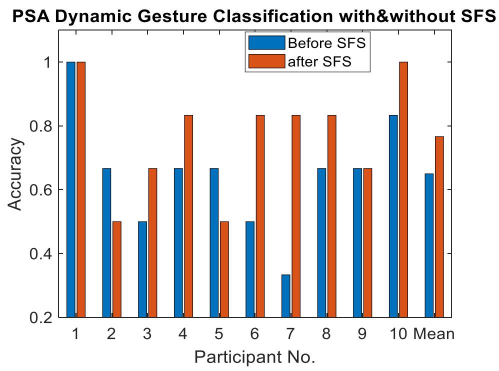


Figure 7.5 PSA SFS results with dynamic gesture

SFS (Sequential Forward Selection) is selected to search the optimal feature combinations for reducing the dimension of the original feature set and gaining some improvement in the classification performance. This wrapper-based feature selection method uses the performance of a Quadratic-kernel SVM classifier (same with the classification) as the metric to evaluate the priority of each possible feature combination. The classification accuracy with and without SFS are compared in Figure 7.3-7.5 for the cases that static gestures using PSA and dynamic gestures using both sensors, where the accuracy bar is plotted for each participant along with the average. For the Doppler radar, the average classification accuracy of the dynamic gestures rises approximately 12% except the first subject, which reports an accuracy drop of around 17%. This is due to the limitation of our wrapper method implementation, which looks for the generalized optimal feature combination across all the pool of participants instead of fitting the process to each of them individually. In the PSA case, the average performance of static gesture recognition increases about 5%, whereas the accuracy gain is more substantial for each participant with respect to dynamic gestures (the mean value is nearly 13%).

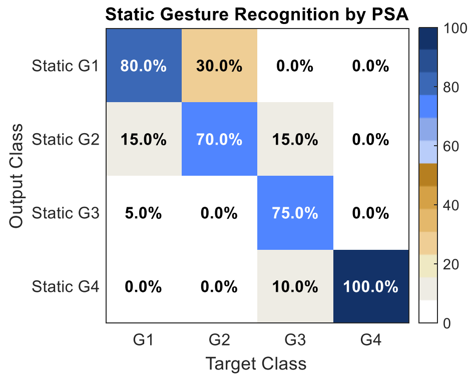


Figure 7.6 Confusion matrix of PSA-only for static gesture recognition

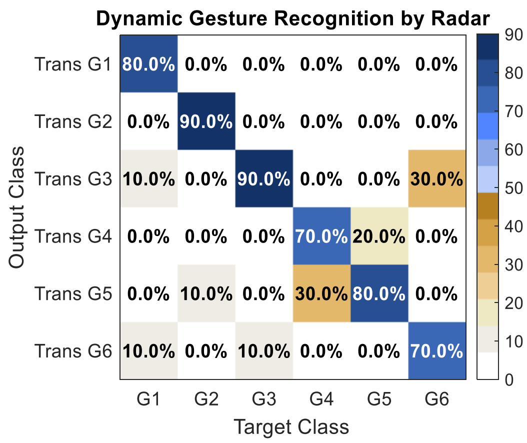


Figure 7.7 Confusion matrix of radar-only for gesture transition recognition

The SVM classifier with a Quadratic kernel (same with the SFS) is trained with the best feature combinations produced by SFS and then used to recognize the gestures from one unknown participant. This training and testing procedure will not stop until all the ten participants have been selected as the testing subject and the final classification accuracy denotes the average of ten testing iterations. Figure 7.6-7.7 summarizes the classification results of static gestures with PSA and dynamic gestures with radar, respectively. The average classification accuracy across all the participants for static gestures (PSA) is approximately 82.9%. Main classification errors occur between one pairs of gestures, namely, G1 and G2, this is caused by the difference in the tendon's pressure is not very significant when performing G1 (number 0) and G2 (number 1). Apart from that, some minor misclassifications (≤15%)

interleaves between G1 and G3, G2 and G3, G3 and G4. For the gesture transitions by Doppler radar, the average classification accuracy drops about 2.9% with respect to the PSA on static gestures. The main misclassifications exists between Trans 4 ‘0-2’ and Trans 5 ‘2-1’, along with about 30% Tran 6 ‘1-5’ being misclassified to Trans 3 ‘2-0’. In the data collection, some of the participants attempt to leave their palm flat (parallel to the ground), leading to a strong Doppler signature in the favour of classifier, whereas the others intends to create an angle between the palm and the radar line-of-sight, and as a consequence of that, the signature is attenuated. Those different performing styles also challenge the classifier as those ‘good’ and ‘bad’ signatures are included in the training and testing set with a complete random manner.

7.1.3 Radar as ‘Enhancer’

The breakdown of the fusion process is illustrated in figure 7.8 when radar is used as an ‘enhancer’ sensor. The classification has been divided into two stages, where the SVM classifier uses the optimal feature set generated by SFS to generate prediction results and related confidence level matrix in the first stage. Due to the difference in number of classes between static and dynamic gestures (4 static gestures for PSA and 6 dynamic gestures for Doppler radar), the PSA and radar yields 6 and 4 columns in the confidence level matrix with respect to the number of classes to identify.

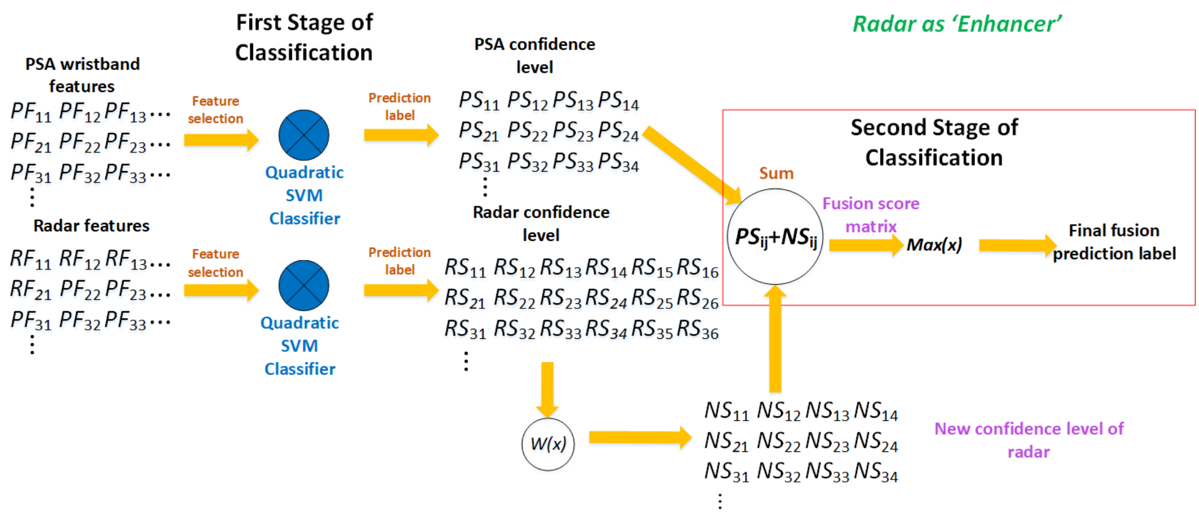


Figure 7.8 Two stages classification model by using confidence level (radar as ‘Enhancer’)

In this case, the confidence level matrix of the two sensors can’t be simply added together. To tackle this problem, it is assumed that in this regard the static gesture contains more significant information and not the gesture transition. Hence, the 6-classes (dynamic gesture) matrix from the Doppler radar can be converted into a 4-class (static gesture) matrix considering the final gesture after each transition. Beyond that, as the Doppler radar in this approach acts as an “enhancer” of the PSA to classify static gestures, a weighted function has been utilized on the new 4-class confidence level matrix to adjust the radar impact in the fusion process, if the i^{th} sample of Doppler radar has a prediction label 'G1', then NS_{ij} (new confidence level matrix of radar) could be derived by equation 7.1:

$$NS_{ij} = \begin{cases} W_1 (j=1) \\ W_2 - \max\{RS_{i1}, RS_{i2}, \cdots, RS_{i6}\} (j=2,3,4) \end{cases} \tag{7.1}$$

Where RS_{i1} to RS_{i6} denote the elements in original confidence level matrix of Doppler radar, with respect to class 1 to 6. W_1 and W_2 are two weight coefficients, which are introduced to control the radar influence in the fusion confidence level matrix. Decreasing the value of W_2 or increasing the value of W_1 will weaken the impact of radar, and vice versa. This equation translates a 6-classes problem to a 4-classes problem with keeping the useful information from Doppler radar simultaneously. Furthermore, figure 7.9 compares the average classification accuracy regarding to different values of the two weight coefficients. The trend suggests that higher classification accuracy is associated with choosing a medium W_1 and a high W_2. It is reported that the optimal classification accuracy (around 91.4%) occurs when the W_1 and W_2 equates to 1 and 0.76 respectively.

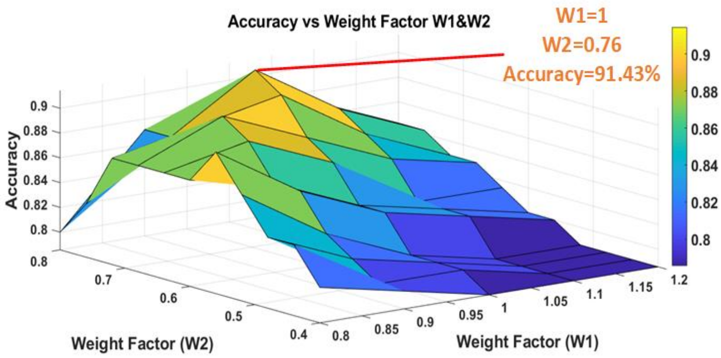


Figure 7.9 Classification accuracy with different weights (radar as 'Enhancer')

Figure 7.10 presents the classification results when soft fusion is applied between Doppler radar and PSA to boost the performance of static gesture recognition. Compared to the PSA-only in figure 7.6, the average classification accuracy across all the ten participants raises by about 8.5% to 91.4%, whereas the misclassifications between G1 'number 0' and G2 'number 1' has been decreased to a low level (from 30% to 10%). Moreover, Class G2 yields the lowest correctly classified rate (sensitivity) in the circumstance that using PSA independently, as shown in figure 7.6. However, this number is boosted to 90% with approximately 20% accuracy gain thanks to the information fusion with radar.

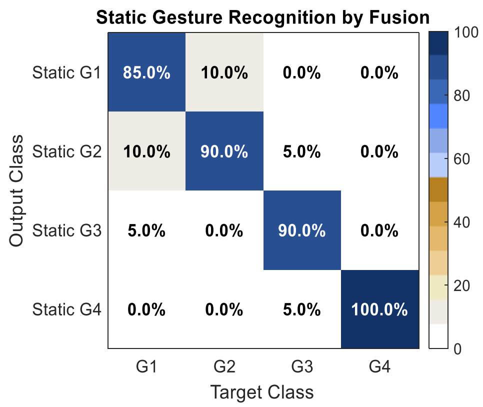


Figure 7.10 Confusion matrix of static gesture recognition using soft fusion

7.1.4 PSA as 'Enhancer'

Similarly, PSA could also be used as 'enhancer' for the Doppler radar to recognize the gesture transitions (dynamic gestures), where there is no class imbalance in the second classification stage (PSA data can be directly used for dynamic gestures). The confidence level matrix of radar and PSA are summed to build a fusion confidence level matrix, whereas the weight function is used to control the PSA impact as we did in the case that radar is 'enhancer'. Figure 7.11 indicates the dependence of the accuracy and the weight coefficient between Doppler radar and pressure sensor, the fusion accuracy for dynamic gestures reaches the maximum point when the ratio of PSA information in the fusion confidence level is between 50% and 60%. The confusion matrix corresponded to the best soft fusion cases for the dynamic gestures is illustrated in figure 7.12. Most of the misclassifications in figure 7.7 are corrected, however still some minor errors (about 10%, 6 samples out of 60 are misclassified) between dynamic gestures Trans 2, Trans 3, Trans 4 and Trans 5.

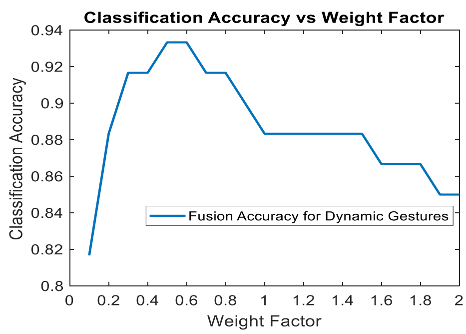


Figure 7.11 The classification accuracy with different weight factors (PSA as 'Enhancer')

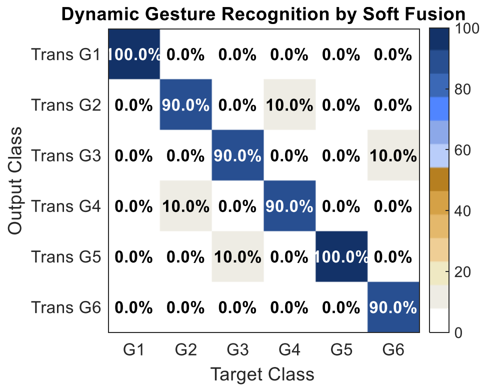


Figure 7.12 Confusion matrix of dynamic gesture recognition using soft fusion

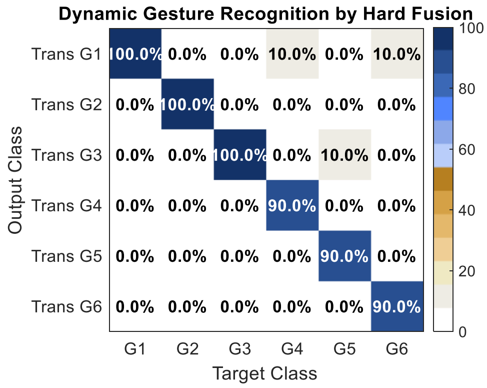


Figure 7.13 Confusion matrix of dynamic gesture recognition with the help of NBC

Hard fusion, in particular, a Naïve Bayes Combiner (NBC) is used to seek subsequent improvement on top of the soft fusion results. For the dynamic gestures, the average classification accuracy of hard fusion is approximately 95%, whereas there is no significant improvement on the static gestures using the same approach. The classification results of dynamic gestures with the help of NBC are shown in figure 7.13 through a confusion matrix, where the hard fusion results is raised about 1.7% in terms of average classification accuracy. However, it is interesting to notice that the misclassified events are varied by soft and hard fusion between PSA and radar, suggesting a potential hybrid fusion framework to combine the advantages of both.

7.1.5 Classification Results of 'L1O' Cross-Validation

Figure 7.14 compares the 'L1O' classification results for both static and dynamic gesture, with respect to using Doppler radar individually, and the soft fusion and hard fusion frameworks. In the case of static gestures, it is reported that an accuracy improvement of 14-28% is achieved for most of the participants through soft fusion with the exception of No. 2 and 3. Instead, hard fusion improves the performance of these two subjects on top of soft fusion results, but an accuracy drop occurs at participant No. 1 and 10. For the dynamic gesture recognition, the gain of average accuracy is approximately 13.3% through soft fusion while the hard fusion can earn more especially on participant No. 5. Furthermore, the variance level across all the ten participants is shown in figure 7.15, compared to the radar-only case, soft fusion slightly decreases the accuracy variance for static gestures, whereas significant reduction of variance level is provided by hard fusion (half for static gestures and about 30% for dynamic gestures). This indicates that hard fusion framework is more robust and stable than soft fusion.

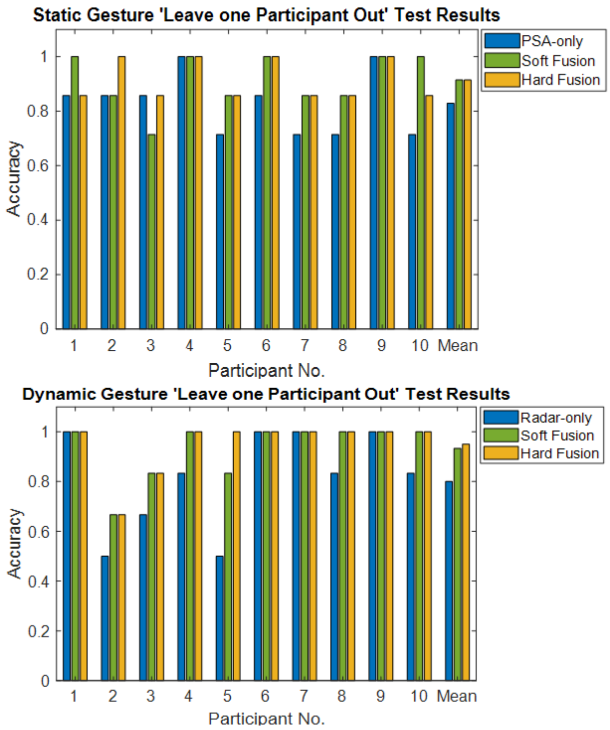


Figure 7.14 Classification results of each participant for radar-only and different fusion methods (top: static gesture, bottom: dynamic gesture)

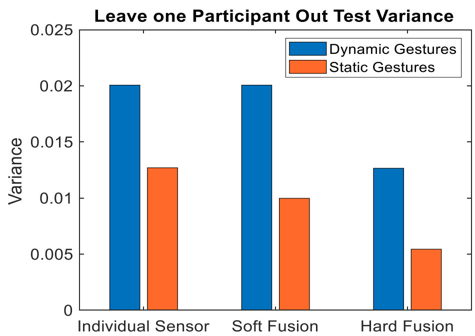


Figure 7.15 Accuracy variance in the 'L1O' test for individual sensor and different fusion cases

7.1.6 Summary of the chapter

In this chapter, a hierarchical sensor fusion framework that combines the strength of UWB impulse radar and PSA wristband is proposed to improve the recognition rate of static and dynamic hand gestures. For the static gestures, the fusion framework needs to be fine-tuned to overcome the limitation of Doppler Effect (only moving targets could be captured). The ‘L1O’ results of soft and hard fusion show better average performance (accuracy) and system stability (variance) than using these two sensing approaches individually.

8 Conclusions and Future Work

8.1 Conclusions

The major objective of this thesis is to explore possible fusion between additional radar nodes as well as between radar and additional heterogeneous sensors for classifying human behaviours and detecting crucial fall events. The proposed fusion frameworks have been successfully verified through the classification results of different datasets collected at University of Glasgow CSI Lab and University of Alabama-Tuscaloosa CI4R Lab. In addition to novel approaches to combine information of radar and other sensors in a multimodal framework, this thesis has explored classification of continuous sequences of human activities and gait, a step forward compared to the analysis of time-limited, individual activities recorded in separated datasets.

Given some experimental data, a machine learning-based classifier can be trained to learn the common properties within each activity (e.g. fall events always yield a strong acceleration towards ground). Hence, different classification methods have been studied in the literature, and this thesis has provided a comprehensive review to compare their robustness and computational cost for classifying human activities and movements. Compared to computer vision-based sensing approaches, such as image and video cameras, radar tracks the moving trajectory of the target without generating plain images or videos, therefore with less privacy issues induced. Even radar may have a blind zone due to its transceiver beam width, with the help of multistatic configuration, it is still able to provide comparable performance as cameras. This thesis also reviews different radar technologies (CW, FMCW, UWB impulse) in the field of AAL as well as the advantages and disadvantages of other sensing technologies. Due to the small size and light weight, wearable IMU sensor has been considered as the potential helper to the radar in this thesis. Wearable IMU sensor attempts to measure acceleration, angular speed, and magnetic field strength of individual body parts (e.g. wrist, waist, and ankle). The information collected by IMU could be used to complement with radar when the range resolution is not enough to characterize small movements or the strength of micro-Doppler signature is attenuated because of aspect angle problems, and this thesis lists some possible fusion approaches among heterogeneous/non-heterogeneous sensors, which are in relation to the novel hybrid fusion presented in this thesis.

Radar information could be expressed in a 3-D space, containing physical distance to the target, time, and target radial velocity based on Doppler Effect. Micro-Doppler signatures, generated by applying TF analysis on the range-time maps, are able to characterize ‘micro-motion’ of individual body parts (e.g. swinging of limbs and rotation of head) when the human body is performing ‘bulk motion’ (e.g. walking, drinking water and falling). Therefore, they have been widely used in classifying human behaviours, detecting the presence of the target, distinguishing the arm/un-armed personnel as well as separating human from other objects. This thesis also presents the mathematical analysis from transmitted and receive chirp signals to extraction of micro-Doppler information in details, whereas the parameters that have an impact on the radar performance have been discussed too.

Prior to the classification, numerical features need to be extracted from raw data to characterize human activities and movements at a high level as well as to reduce the input dimension of the classifier. This thesis introduces the handcrafted features used in this thesis, involving physical features, transform-based features and range-Doppler features for radar, and time and frequency domain features for wearable IMU sensor. However, not all handcrafted features contribute in the classifier training and testing, thus feature selection techniques are employed to drop the redundant features and subsequently improving the classification performance. Regarding this, different feature selection techniques have been discussed, ranging from filter-based methods (Fisher score, Relief-F) to wrapper-based methods (SFS, SBS). For the classifiers used in this thesis, a detailed description has been provided involving the training and testing processes as well as the metrics to evaluate classification performance. Multi-modal sensing/sensor fusion combines the strength of different sensors/sensing technologies to improve the correctly classified rate of certain human behaviours when single sensor cannot provide desirable results due to its own limitations or external conditions. Additionally, this thesis also introduces some novel fusion methods, including hybrid fusion and trilateration-based signal level fusion.

This thesis firstly shows the classification results of ten snapshot activities (testing subjects only perform one activity in the duration of sensor measurement) including frontal fall on a mattress. These data have been collected through a wearable IMU sensor and a FMCW radar operating at 5.8 GHz. Information fusion is implemented to combine the advantages of two sensors and the gain of fusion at distinct levels are compared, where a voting system comprised of two KNN and SVM classifiers outperforms other fusion schemes using the optimal feature set generated by SFS. However, magnetometer yields the worst performance among all three inertial sensors when used individually, therefore it was picked out to fuse with FMCW radar for validating the effectiveness of sensor fusion. Aside from conventional classifier like SVM, an ANN with one to three hidden layers have been utilized to classify the ten activities under a more challenging and realistic 'L1O' cross validation scheme (simulating the situation that classifier can't see the users' data before the actual test), and the results of feature level fusion yields about 7% improvement of accuracy compared to using radar alone.

This thesis also extends the work to continuous activity streams, compared to snapshot activity, continuous activity stream contains a combination of activities (six in our case) and natural transitions between them. The conventional sliding window-based method has been tested on the continuous data collected by three wearable IMU sensors at different body positions (wrist, waist, and ankle) and a FMCW radar, whereas different sensor combinations have also been evaluated. The results point out that the IMU on the wrist is the best partner of FMCW radar. However, the limitation of sliding window-based is hard to identify the transition of two activities due to that it segments the long data sequence into smaller frames before processing. A dual Bi-LSTM layer network has therefore been proposed to address this problem by considering the continuous activity stream as a united sequence, and this allows that the prediction labels are produced in high time resolution and the activity transitions are localized accurately. Meanwhile, a hybrid fusion framework containing soft fusion with different sensor weights and a hard combiner (e.g. RC, NBC) has been proposed in this thesis, aiming to integrate the advantages of both soft fusion and hard fusion, and it can reach nearly 95.8% accuracy with the help of proposed Bi-LSTM network when classifying the continuous activity data. Gait movements are usually considered as a cyclic motion (repetition of same pattern), thus the proposed Bi-LSTM network can be directly used on either snapshot data or continuous data. The classification results show that information fusion of Doppler information using an NBC surpasses the best single radar sensor about 7.3%, whereas a trilateration-based algorithm has been proposed to combine the range data of three radar sensors at signal level. It is interesting to see that fusion with range-based fusion yield comparable performance as Doppler (Doppler slightly higher) with >90% less in computational loads.

As an additional work, this thesis confirms the validity of information fusion through the classification results of micro-gestures. A hierarchical sensor fusion framework about UWB impulse radar and PSA has been established to better recognize static and dynamic gestures, for the dynamic gestures, it is similar to soft fusion in activity recognition, whereas it needs to be fine-tuned when used on static gestures since radar can only sense the dynamic gestures. Additionally, the robustness of hard combiner has been proved to surpass weighted soft fusion, particularly for dynamic gesture.

8.2 Future Work

Although the proposed information fusion schemes have shown significant gain in classification accuracy compared to using single sensor, there is still possible further work to improve the system performance. The most important future action would be validating the method in a wider cohort of participants and activities, including a larger set of measurement environments (e.g. rooms with different shapes or outdoor), aspect angles with respect to the radar, and span of age and physical conditions of the participants. In terms of the implementation of the neural networks, deeper and more complex architectures can be considered with more data collected, for instance training a hybrid deep model [45], [50] to leverage the strength of 3-D CNN and Bi-LSTM network used in this thesis, as well as customization to the structure and hyper-parameters to avoid overfitting while managing the diversity of data for each participant and scenario, for instance adding a CTC (Connectionist Temporal Classification) layer after softmax function to resolve the misalignment between input and output sequences and using GA (Generic Algorithm) to optimize the initial weights of neural networks.

The format of the input data also has scope for further work, considering, for example, radar data from the range-Doppler domain as complementary or alternative to range-time maps and spectrograms, and other sensing modalities if available. Besides that, testing the classification model with different sensors (e.g. training with radar and three IMUs and evaluate with only radar data or cross frequency testing on different radar dataset) is very worth to explore in terms of evaluating the capability of the classifier under more complex condition and for cross-modality learning.

The current HAR system is not in real-time since the classification results are produced through the program in MATLAB after the data collection. Therefore, future work would also seek to implement the information fusion algorithms on embedded platforms and in real-time, moving towards more realistic deployment conditions, for instance, transplanting the weights of classification models trained on the GPU server to a FPGA developing board and then using this portable device to recognize daily activities.

Furthermore, particularly interest is in the fusion techniques, involving signal level fusion using micro-Doppler spectrograms, fusion level fusion using the strength of neural networks (concatenating the feature vectors extracted from two sensors at the final fully connected layer), decision level fusion using the trainable soft combiners such as Fuzzy Integral (FI) [91], KL weights [103], decision template [153] as well as Dempster-Shafer (DS) methods [91] and the variants of NBC such as Behavior-Knowledge Space (BKS) [154] and WERnecke's (WER) combiners [153]. The hybrid fusion presented in this thesis could also be further explored to integrate the power of signal, feature and decision level fusion, also referred as ‘fusion of fusion’.

Reference

[1] NIHR Dissemination Centre, "HELP AT HOME Use of assistive technology for older people," *Natl. Inst. Heal. Res.*, pp. 3–6, 2018, doi: doi: 10.3310/themedreview-03345.

[2] W. H. O. Ageing and L. C. Unit, "WHO global report on falls prevention in older age," *World Heal. Organ.*, 2008.

[3] M. Mubashir, L. Shao, and L. Seed, "A survey on fall detection: Principles and approaches," *Neurocomputing*, vol. 100, pp. 144–152, 2013.

[4] 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 Sens. J.*, vol. 17, no. 3, pp. 812–822, 2017, doi: 10.1109/JSEN.2016.2628099.

[5] M. Terroso, N. Rosa, A. T. Marques, and R. Simoes, "Physical consequences of falls in the elderly: a literature review from 1995 to 2010," *Eur. Rev. Aging Phys. Act.*, vol. 11, no. 1, p. 51, 2014.

[6] C. S. Florence, G. Bergen, A. Atherly, E. Burns, J. Stevens, and C. Drake, "Medical costs of fatal and nonfatal falls in older adults," *J. Am. Geriatr. Soc.*, 2018.

[7] I. H. López-Nava and A. Muñoz-Meléndez, "Wearable inertial sensors for human motion analysis: A review," *IEEE Sens. J.*, vol. 16, no. 22, pp. 7821–7834, 2016.

[8] S. Z. Gurbuz and M. G. Amin, "Radar-Based Human-Motion Recognition With Deep Learning: Promising applications for indoor monitoring," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 16–28, 2019.

[9] T. R. Bennett, J. Wu, N. Kehtarnavaz, and R. Jafari, "Inertial measurement unit-based wearable computers for assisted living applications: A signal processing perspective," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 28–35, 2016.

[10] P. Pierleoni *et al.*, "A wearable fall detector for elderly people based on AHRS and barometric sensor," *IEEE Sens. J.*, vol. 16, no. 17, pp. 6733–6744, 2016.

[11] P. Pierleoni, A. Belli, L. Palma, M. Pellegrini, L. Pernini, and S. Valenti, "A high reliability wearable device for elderly fall detection," *IEEE Sens. J.*, vol. 15, no. 8, pp. 4544–4553, 2015.

[12] X. Liang, H. Heidari, and R. Dahiya, "Wearable Capacitive-Based Wrist-Worn Gesture Sensing System," in *2017 New Generation of CAS (NGCAS)*, 2017, pp. 181–184, doi: 10.1109/NGCAS.2017.80.

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

[14] P. Jung, G. Lim, S. Kim, and K. Kong, "A Wearable Gesture Recognition Device for Detecting Muscular Activities Based on Air-Pressure Sensors," *IEEE Trans. Ind. Informatics*, vol. 11, no. 2, pp. 485–494, 2015, doi: 10.1109/TII.2015.2405413.

[15] J. Maitre, K. Bouchard, C. Bertuglia, and S. Gaboury, "Recognizing activities of daily living from UWB radars and deep learning," *Expert Syst. Appl.*, vol. 164, p. 113994, 2021, doi: https://doi.org/10.1016/j.eswa.2020.113994.

[16] S. U. Park, J. H. Park, M. A. Al-masni, M. A. Al-antari, M. Z. Uddin, and T.-S. Kim, "A Depth Camera-based Human Activity Recognition via Deep Learning Recurrent Neural Network for Health and Social Care Services," *Procedia Comput. Sci.*, vol. 100, pp. 78–84, 2016, doi: https://doi.org/10.1016/j.procs.2016.09.126.

[17] J. Yun and S.-S. Lee, "Human Movement Detection and Identification Using Pyroelectric Infrared Sensors," *Sensors* , vol. 14, no. 5. 2014, doi: 10.3390/s140508057.

[18] S. Tao, M. Kudo, and H. Nonaka, "Privacy-preserved behavior analysis and fall detection by an infrared ceiling sensor network," *Sensors*, vol. 12, no. 12, pp. 16920–16936, 2012.

[19] B. Erol and M. G. Amin, "Fall motion detection using combined range and Doppler features," in *2016 24th European Signal Processing Conference (EUSIPCO)*, 2016, pp. 2075–2080, doi: 10.1109/EUSIPCO.2016.7760614.

[20] 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 Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1933–1937, 2015.

[21] A. Seifert, M. G. Amin, and A. M. Zoubir, "Toward Unobtrusive In-Home Gait Analysis Based on Radar Micro-Doppler Signatures," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 9, pp. 2629–2640, 2019, doi: 10.1109/TBME.2019.2893528.

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

[23] 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 Trans. Mob. Comput.*, vol. 16, no. 2, pp. 511–526, 2017.

[24] X. Yang *et al.*, "Freezing of Gait Detection Considering Leaky Wave Cable," *IEEE Trans. Antennas Propag.*, vol. 67, no. 1, pp. 554–561, 2019.

[25] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless Sensing for Human Activity: A Survey," *IEEE Commun. Surv. Tutorials*, vol. 22, no. 3, pp. 1629–1645, 2020, doi: 10.1109/COMST.2019.2934489.

[26] X. Li, Y. He, and X. Jing, "A Survey of Deep Learning-Based Human Activity Recognition in Radar," *Remote Sens.*, vol. 11, p. 1068, May 2019, doi: 10.3390/rs11091068.

[27] Y. Kim and B. Toomajian, "Application of Doppler radar for the recognition of hand gestures using optimized deep convolutional neural networks," in *2017 11th European Conference on Antennas and Propagation (EUCAP)*, 2017, pp. 1258–1260, doi: 10.23919/EuCAP.2017.7928465.

[28] S.-M. Lee, S. M. Yoon, and H. Cho, "Human activity recognition from accelerometer data using Convolutional Neural Network," in *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2017, pp. 131–134, doi: 10.1109/BIGCOMP.2017.7881728.

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

[30] M. Wang, Y. D. Zhang, and G. Cui, "Human motion recognition exploiting radar with stacked recurrent neural network," *Digit. Signal Process.*, vol. 87, pp. 125–131, 2019, doi: https://doi.org/10.1016/j.dsp.2019.01.013.

[31] Z. Chen, G. Li, F. Fioranelli, and H. Griffiths, "Personnel Recognition and Gait Classification Based on Multistatic Micro-Doppler Signatures Using Deep Convolutional Neural Networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 669–673, 2018, doi: 10.1109/LGRS.2018.2806940.

[32] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, "A Multisensory Approach for Remote Health Monitoring of Older People," *IEEE J. Electromagn. RF Microwaves Med. Biol.*,

vol. 2, no. 2, pp. 102–108, 2018, doi: 10.1109/JERM.2018.2827099.

[33] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Short-range FMCW monopulse radar for hand-gesture sensing," in *2015 IEEE Radar Conference (RadarCon)*, 2015, pp. 1491–1496, doi: 10.1109/RADAR.2015.7131232.

[34] H. Li *et al.*, "Hierarchical Sensor Fusion for Micro-Gesture Recognition With Pressure Sensor Array and Radar," *IEEE J. Electromagn. RF Microwaves Med. Biol.*, vol. 4, no. 3, pp. 225–232, 2020, doi: 10.1109/JERM.2019.2949456.

[35] X. Liang *et al.*, "Fusion of Wearable and Contactless Sensors for Intelligent Gesture Recognition," *Adv. Intell. Syst.*, vol. 1, no. 7, p. 1900088, Nov. 2019, doi: https://doi.org/10.1002/aisy.201900088.

[36] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, 2009.

[37] Y. Kim and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 8–12, 2016.

[38] I. D. Castro, M. Mercuri, T. Torfs, I. Lorato, R. Puers, and C. Van Hoof, "Sensor fusion of capacitively-coupled ECG and continuous-wave Doppler radar for improved unobtrusive heart rate measurements," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. PP, no. 99, p. 1, 2018, doi: 10.1109/JETCAS.2018.2802639.

[39] Y. Kim and B. Toomajian, "Hand gesture recognition using micro-Doppler signatures with convolutional neural network," *IEEE Access*, vol. 4, pp. 7125–7130, 2016.

[40] M. A. Richards, J. Scheer, W. A. Holm, and W. L. Melvin, *Principles of modern radar*. Citeseer, 2010.

[41] A. Seifert, A. M. Zoubir, and M. G. Amin, "Detection of Gait Asymmetry Using Indoor Doppler Radar," in *2019 IEEE Radar Conference (RadarConf)*, 2019, pp. 1–6, doi: 10.1109/RADAR.2019.8835611.

[42] M. Wang, G. Cui, X. Yang, and L. Kong, "Human body and limb motion recognition via stacked gated recurrent units network," *IET Radar, Sonar Navig.*, vol. 12, no. 9, pp. 1046–1051, 2018.

[43] R. P. Trommel, R. I. A. Harmanny, L. Cifola, and J. N. Driessen, "Multi-target human gait classification using deep convolutional neural networks on micro-doppler spectrograms," in *2016 European Radar Conference (EuRAD)*, 2016, pp. 81–84.

[44] D. E. Barrick, "FM/CW radar signals and digital processing," NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION BOULDER CO WAVE PROPAGATION LAB, 1973.

[45] Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic continuous hand gesture recognition using FMCW radar sensor," *IEEE Sens. J.*, vol. 18, no. 8, pp. 3278–3289, 2018.

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

[47] C. Ding *et al.*, "Continuous Human Motion Recognition With a Dynamic Range-Doppler Trajectory Method Based on FMCW Radar," *IEEE Trans. Geosci. Remote Sens.*, 2019.

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

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

[50] S. Wang, J. Song, J. Lien, I. Poupyrev, and O. Hilliges, *Interacting with Soli: Exploring Fine-Grained Dynamic Gesture Recognition in the Radio-Frequency Spectrum*. 2016.

[51] B. Jokanović and M. Amin, "Fall Detection Using Deep Learning in Range-Doppler Radars," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 1, pp. 180–189, 2018, doi: 10.1109/TAES.2017.2740098.

[52] 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 Navig.*, vol. 12, no. 7, pp. 729–734, 2018, doi: 10.1049/iet-rsn.2017.0511.

[53] R. J. Fontana, "Recent system applications of short-pulse ultra-wideband (UWB) technology," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 9, pp. 2087–2104, 2004.

[54] G. Ossberger, T. Buchegger, E. Schimback, A. Stelzer, and R. Weigel, "Non-invasive respiratory movement detection and monitoring of hidden humans using ultra wideband pulse radar," in *2004 International Workshop on Ultra Wideband Systems Joint with Conference on Ultra Wideband Systems and Technologies. Joint UWBST & IWUWBS 2004 (IEEE Cat. No.04EX812)*, 2004, pp. 395–399, doi: 10.1109/UWBST.2004.1321003.

[55] A. Lazaro, D. Girbau, and R. Villarino, "Analysis of vital signs monitoring using an IR-UWB radar," *Prog. Electromagn. Res.*, vol. 100, pp. 265–284, 2010.

[56] Z. Duan and J. Liang, "Non-contact detection of vital signs using a UWB radar sensor," *IEEE Access*, vol. 7, pp. 36888–36895, 2018.

[57] M. Aftanas, "Through wall imaging with UWB radar system," *Dep. Electron. Multimed. Commun. Tech. Univ. Kosice*, 2009.

[58] Q. Huang, L. Qu, B. Wu, and G. Fang, "UWB through-wall imaging based on compressive sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1408–1415, 2009.

[59] H. Sadreazami, M. Bolic, and S. Rajan, "On the Use of Ultra Wideband Radar and Stacked LSTM-RNN for at Home Fall Detection," in *2018 IEEE Life Sciences Conference (LSC)*, 2018, pp. 255–258, doi: 10.1109/LSC.2018.8572048.

[60] C. Ding *et al.*, "Non-contact Human Motion Recognition based on UWB Radar," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. PP, no. 99, p. 1, 2018, doi: 10.1109/JETCAS.2018.2797313.

[61] Y. Lang, C. Hou, Y. Yang, D. Huang, and Y. He, *Convolutional neural network for human micro-Doppler classification*. 2017.

[62] H. Du, Y. He, and T. Jin, "Transfer Learning for Human Activities Classification Using Micro-Doppler Spectrograms," in *2018 IEEE International Conference on Computational Electromagnetics (ICCEM)*, 2018, pp. 1–3, doi: 10.1109/COMPEM.2018.8496654.

[63] H. Du, T. Jin, Y. He, Y. Song, and Y. Dai, "Segmented convolutional gated recurrent neural networks for human activity recognition in ultra-wideband radar," *Neurocomputing*, vol. 396, pp. 451–464, 2020, doi: https://doi.org/10.1016/j.neucom.2018.11.109.

[64] Y. Lang, Q. Wang, Y. Yang, C. Hou, D. Huang, and W. Xiang, "Unsupervised Domain Adaptation for Micro-Doppler Human Motion Classification via Feature Fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 392–396, 2019, doi: 10.1109/LGRS.2018.2873776.

[65] J. Bryan and Y. Kim, "Classification of human activities on UWB radar using a support vector machine," in *2010 IEEE Antennas and Propagation Society International Symposium*, 2010, pp. 1–4, doi: 10.1109/APS.2010.5561935.

[66] H. Li, A. Shrestha, H. Heidari, J. L. Kernec, and F. Fioranelli, “Magnetic and Radar Sensing for Multimodal Remote Health Monitoring,” *IEEE Sens. J.*, p. 1, 2018, doi: 10.1109/JSEN.2018.2872894.

[67] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, “A robust human activity recognition system using smartphone sensors and deep learning,” *Futur. Gener. Comput. Syst.*, vol. 81, pp. 307–313, 2018, doi: https://doi.org/10.1016/j.future.2017.11.029.

[68] H. S. AlZubi, S. Gerrard-Longworth, W. Al-Nuaimy, Y. Goulermas, and S. Preece, “Human activity classification using a single accelerometer,” in *Computational Intelligence (UKCI), 2014 14th UK Workshop on*, 2014, pp. 1–6.

[69] K. E. Bliley *et al.*, “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*, 2005, pp. 92–93, doi: 10.1109/MMB.2005.1548392.

[70] T. Kose, Y. Terzioglu, K. Azgin, and T. Akin, “A single-mass self-resonating closed-loop capacitive MEMS accelerometer,” in *2016 IEEE SENSORS*, 2016, pp. 1–3, doi: 10.1109/ICSENS.2016.7808711.

[71] H. Dong and Y. Gao, “Comparison of Compensation Mechanism Between an NMR Gyroscope and an SERF Gyroscope,” *IEEE Sens. J.*, vol. 17, no. 13, pp. 4052–4055, 2017, doi: 10.1109/JSEN.2017.2703601.

[72] B. Hazarika, N. Afzulpurkar, C. Punyasai, and D. K. 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*, 2012, pp. 1–4, doi: 10.1109/ECTICon.2012.6254223.

[73] V. Nabaei, R. Chandrawati, and H. Heidari, “Magnetic biosensors: Modelling and simulation,” *Biosens. Bioelectron.*, 2017.

[74] S. Zuo, H. Heidari, D. Farina, and K. Nazarpour, “Magnetomyography: Miniaturized Magnetic Sensors for Implantable Magnetomyography (Adv. Mater. Technol. 6/2020),” *Adv. Mater. Technol.*, vol. 5, no. 6, p. 2070033, Jun. 2020, doi: https://doi.org/10.1002/admt.202070033.

[75] H. Heidari, E. Bonizzoni, U. Gatti, F. Maloberti, and R. Dahiya, “CMOS vertical Hall magnetic sensors on flexible substrate,” *IEEE Sens. J.*, vol. 16, no. 24, pp. 8736–8743, 2016.

[76] S. Chernbumroong, A. S. Atkins, and H. Yu, “Activity classification using a single wrist-worn accelerometer,” in *Software, Knowledge Information, Industrial Management and Applications (SKIMA), 2011 5th International Conference on*, 2011, pp. 1–6.

[77] C. Zhu and W. Sheng, “Human daily activity recognition in robot-assisted living using multi-sensor fusion,” in *2009 IEEE International Conference on Robotics and Automation*, 2009, pp. 2154–2159.

[78] C. Zhu and W. Sheng, “Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living,” *IEEE Trans. Syst. Man, Cybern. A Syst. Humans*, vol. 41, no. 3, pp. 569–573, 2011.

[79] F. J. Ordóñez and D. Roggen, “Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition,” *Sensors*, vol. 16, no. 1, p. 115, 2016.

[80] R. M. Gibson, A. Amira, N. Ramzan, P. Casaseca-de-la-Higuera, and Z. Pervez, “Multiple comparator classifier framework for accelerometer-based fall detection and diagnostic,” *Appl. Soft Comput.*, vol. 39, pp. 94–103, 2016.

[81] Y. Li and J. Ibanez-Guzman, *Lidar for Autonomous Driving: The principles, challenges, and*

trends for automotive lidar and perception systems. 2020.

[82] H. Gao, B. Cheng, J. Wang, K. Li, J. Zhao, and D. Li, "Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment," *IEEE Trans. Ind. Informatics*, vol. 14, no. 9, pp. 4224–4231, 2018.

[83] H. Wang, B. Wang, B. Liu, X. Meng, and G. Yang, "Pedestrian recognition and tracking using 3D LiDAR for autonomous vehicle," *Rob. Auton. Syst.*, vol. 88, pp. 71–78, 2017.

[84] Q. Guan, X. Yin, X. Guo, and G. Wang, "A Novel Infrared Motion Sensing System for Compressive Classification of Physical Activity," *IEEE Sens. J.*, vol. 16, no. 8, pp. 2251–2259, 2016, doi: 10.1109/JSEN.2016.2514606.

[85] A. Jalal, S. Kamal, and D. Kim, "A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments," *Sensors*, vol. 14, no. 7, pp. 11735–11759, 2014.

[86] S.-R. Ke, H. L. U. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, and K.-H. Choi, "A review on video-based human activity recognition," *Computers*, vol. 2, no. 2, pp. 88–131, 2013.

[87] N. Robertson and I. Reid, "A general method for human activity recognition in video," *Comput. Vis. Image Underst.*, vol. 104, no. 2–3, pp. 232–248, 2006.

[88] P. C. Ribeiro, J. Santos-Victor, and P. Lisboa, "Human activity recognition from video: modeling, feature selection and classification architecture," in *Proceedings of International Workshop on Human Activity Recognition and Modelling*, 2005, pp. 61–78.

[89] W. Lin, M.-T. Sun, R. Poovandran, and Z. Zhang, "Human activity recognition for video surveillance," in *2008 IEEE International Symposium on Circuits and Systems*, 2008, pp. 2737–2740, doi: 10.1109/ISCAS.2008.4542023.

[90] E. P. Ijjina and K. M. Chalavadi, "Human action recognition using genetic algorithms and convolutional neural networks," *Pattern Recognit.*, vol. 59, pp. 199–212, 2016, doi: https://doi.org/10.1016/j.patcog.2016.01.012.

[91] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, "Multi-classification approaches for classifying mobile app traffic," *J. Netw. Comput. Appl.*, vol. 103, pp. 131–145, 2018.

[92] R. M. Gibson, A. Amira, N. Ramzan, P. Casaseca-de-la-Higuera, and Z. Pervez, "Multiple comparator classifier framework for accelerometer-based fall detection and diagnostic," *Appl. Soft Comput.*, vol. 39, pp. 94–103, 2016, doi: https://doi.org/10.1016/j.asoc.2015.10.062.

[93] B. Erol and M. G. Amin, "Radar Data Cube Processing for Human Activity Recognition Using Multi Subspace Learning," *IEEE Trans. Aerosp. Electron. Syst.*, 2019.

[94] T. D. Bufler and R. M. Narayanan, "Radar classification of indoor targets using support vector machines," *IET Radar, Sonar Navig.*, vol. 10, no. 8, pp. 1468–1476, 2016.

[95] D. Wu *et al.*, "Deep Dynamic Neural Networks for Multimodal Gesture Segmentation and Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1583–1597, 2016, doi: 10.1109/TPAMI.2016.2537340.

[96] Y. LeCun *et al.*, "Handwritten digit recognition with a back-propagation network," in *Advances in neural information processing systems*, 1990, pp. 396–404.

[97] X. Bai, Y. Hui, L. Wang, and F. Zhou, "Radar-Based Human Gait Recognition Using Dual-Channel Deep Convolutional Neural Network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9767–9778, 2019, doi: 10.1109/TGRS.2019.2929096.

[98] N. Dawar and N. Kehtarnavaz, "Action Detection and Recognition in Continuous Action Streams by Deep Learning-Based Sensing Fusion," *IEEE Sens. J.*, vol. 18, no. 23, pp. 9660–

9668, 2018, doi: 10.1109/JSEN.2018.2872862.

[99] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, “Multi-sensor system for driver’s hand-gesture recognition,” in *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, 2015, vol. 1, pp. 1–8.

[100] M. S. Seyfioğlu, A. M. Özbayğglu, and S. Z. Gurbuz, “Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities,” *IEEE Trans. Aerosp. Electron. Syst.*, 2018.

[101] 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,” *Med. Eng. Phys.*, vol. 42, pp. 1–12, 2017, doi: https://doi.org/10.1016/j.medengphy.2016.12.011.

[102] C. Kownacki, “Optimization approach to adapt Kalman filters for the real-time application of accelerometer and gyroscope signals’ filtering,” *Digit. Signal Process.*, vol. 21, no. 1, pp. 131–140, 2011, doi: https://doi.org/10.1016/j.dsp.2010.09.001.

[103] J. A. Benediktsson, J. R. Sveinsson, O. K. Ersoy, and P. H. Swain, “Parallel consensual neural networks,” *IEEE Trans. Neural Networks*, vol. 8, no. 1, pp. 54–64, 1997, doi: 10.1109/72.554191.

[104] N. Twomey *et al.*, *The SPHERE Challenge: Activity Recognition with Multimodal Sensor Data*. 2016.

[105] C. Chen, R. Jafari, and N. Kehtarnavaz, “A real-time human action recognition system using depth and inertial sensor fusion,” *IEEE Sens. J.*, vol. 16, no. 3, pp. 773–781, 2016.

[106] C. Zhu, Q. Cheng, and W. Sheng, “Human activity recognition via motion and vision data fusion,” in *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, 2010, pp. 332–336, doi: 10.1109/ACSSC.2010.5757529.

[107] S. Liu, R. X. Gao, D. John, J. W. Staudenmayer, and P. S. Freedson, “Multisensor Data Fusion for Physical Activity Assessment,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 3, pp. 687–696, 2012.

[108] Y. Kim, J. Park, and T. Moon, *Classification of micro-Doppler signatures of human aquatic activity through simulation and measurement using transferred learning*. 2017.

[109] S. Z. Gürbüz, B. Erol, B. Çağlıyan, and B. Tekeli, “Operational assessment and adaptive selection of micro-Doppler features,” *IET Radar, Sonar Navig.*, vol. 9, no. 9, pp. 1196–1204, 2015.

[110] X. Liang, R. Ghannam, and H. Heidari, “Wrist-Worn Gesture Sensing With Wearable Intelligence,” *IEEE Sens. J.*, vol. 19, no. 3, pp. 1082–1090, 2018.

[111] I. D. Castro, M. Mercuri, T. Torfs, I. Lorato, R. Puers, and C. Van Hoof, “Sensor fusion of capacitively-coupled ECG and continuous-wave Doppler radar for improved unobtrusive heart rate measurements,” *IEEE J. Emerg. Sel. Top. Circuits Syst.*, 2018.

[112] N. Levanon and E. Mozeson, *Radar signals*. John Wiley & Sons, 2004.

[113] M. Ash, M. Ritchie, and K. Chetty, “On the application of digital moving target indication techniques to short-range FMCW radar data,” *IEEE Sens. J.*, vol. 18, no. 10, pp. 4167–4175, 2018.

[114] V. C. Chen, D. Tahmoush, and W. J. Miceli, *Radar Micro-Doppler Signatures*. Institution of Engineering and Technology, 2014.

[115] V. Chen and H. Ling, *Time-Frequency Transforms for Radar Imaging and Signal Analysis*. Artech, 2001.

[116] D. Tahmoush, “Review of micro-Doppler signatures,” *IET Radar, Sonar Navig.*, vol. 9, no. 9, pp. 1140–1146, 2015.

[117] A. W. Miller, C. Clemente, A. Robinson, D. Greig, A. M. Kinghorn, and J. J. Soraghan, “Micro-Doppler based target classification using multi-feature integration,” 2013.

[118] H. Li, A. Shrestha, H. Heidari, J. Le Kernec, and F. Fioranelli, “A Multi-sensory Approach for Remote Health Monitoring of Older People,” *IEEE J. Electromagn. RF Microwaves Med. Biol.*, 2018.

[119] S. Mallat, *A wavelet tour of signal processing*. Elsevier, 1999.

[120] M. Li and J. Lin, “Wavelet-transform-based data-length-variation technique for fast heart rate detection using 5.8-GHz CW Doppler radar,” *IEEE Trans. Microw. Theory Tech.*, vol. 66, no. 1, pp. 568–576, 2017.

[121] B. Y. Su, K. C. Ho, M. J. Rantz, and M. Skubic, “Doppler radar fall activity detection using the wavelet transform,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 3, pp. 865–875, 2014.

[122] C.-P. Lai, Q. Ruan, and R. M. Narayanan, “Hilbert-Huang transform (HHT) processing of through-wall noise radar data for human activity characterization,” in *2007 IEEE Workshop on Signal Processing Applications for Public Security and Forensics*, 2007, pp. 1–6.

[123] C.-P. Lai, R. M. Narayanan, Q. Ruan, and A. Davydov, “Hilbert–Huang transform analysis of human activities using through-wall noise and noise-like radar,” *IET Radar, Sonar Navig.*, vol. 2, no. 4, pp. 244–255, 2008.

[124] S. Rahman and D. A. Robertson, “Radar micro-Doppler signatures of drones and birds at K-band and W-band,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018.

[125] V. C. Chen, W. J. Miceli, and D. Tahmoush, *Radar micro-Doppler signatures: processing and applications*. The Institution of Engineering and Technology, 2014.

[126] W. Zhang and G. Li, “Detection of multiple micro-drones via cadence velocity diagram analysis,” *Electron. Lett.*, vol. 54, no. 7, pp. 441–443, 2018.

[127] B. K. Kim, H.-S. Kang, and S.-O. Park, “Drone classification using convolutional neural networks with merged Doppler images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 38–42, 2016.

[128] R. Ricci and A. Balleri, “Recognition of humans based on radar micro-Doppler shape spectrum features,” *IET Radar, Sonar Navig.*, vol. 9, no. 9, pp. 1216–1223, 2015.

[129] B. Erol, S. Z. Gurbuz, and M. G. Amin, “Frequency-Warped Cepstral Heatmaps for Deep Learning of Human Motion Signatures,” in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, 2018, pp. 1234–1238.

[130] P. Lei, Y. Zhang, J. Wang, and J. Sun, “Estimation of human gait cycle based on cepstrum of radar micro-Doppler signatures,” in *2017 Progress in Electromagnetics Research Symposium - Fall (PIERS - FALL)*, 2017, pp. 2356–2359, doi: 10.1109/PIERS-FALL.2017.8293530.

[131] W. D. Van Eeden, J. P. De Villiers, R. J. Berndt, W. A. J. Nel, and E. Blasch, “Micro-Doppler radar classification of humans and animals in an operational environment,” *Expert Syst. Appl.*, vol. 102, pp. 1–11, 2018.

[132] F. Fioranelli, M. Ritchie, and H. Griffiths, “Centroid features for classification of armed/unarmed multiple personnel using multistatic human micro-Doppler,” *IET Radar, Sonar Navig.*, vol. 10, no. 9, pp. 1702–1710, 2016.

[133] F. Fioranelli, M. Ritchie, S. Z. Gürbüz, and H. Griffiths, “Feature diversity for optimized human micro-Doppler classification using multistatic radar,” *IEEE Trans. Aerosp. Electron. Syst.*, vol.

53, no. 2, pp. 640–654, 2017.

[134] C. Karabacak, S. Z. Gurbuz, A. C. Gurbuz, M. B. Guldogan, G. Hendeby, and F. Gustafsson, "Knowledge exploitation for human micro-Doppler classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 10, pp. 2125–2129, 2015.

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

[136] R. J. Javier and Y. Kim, "Application of Linear Predictive Coding for Human Activity Classification Based on Micro-Doppler Signatures," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1831–1834, 2014, doi: 10.1109/LGRS.2014.2311819.

[137] R. I. A. Harmanny, J. J. M. de Wit, and G. P. Cabic, "Radar micro-Doppler feature extraction using the spectrogram and the cepstrogram," in *2014 11th European Radar Conference*, 2014, pp. 165–168, doi: 10.1109/EuRAD.2014.6991233.

[138] L. Atallah, B. Lo, R. Ali, R. King, and G. Yang, "Real-Time Activity Classification Using Ambient and Wearable Sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 6, pp. 1031–1039, 2009.

[139] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/bf00994018.

[140] 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)*, 2017, pp. 167–170.

[141] 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)*, 2019, pp. 1–2.

[142] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[143] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural networks*, vol. 18, no. 5–6, pp. 602–610, 2005.

[144] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," *arXiv Prepr. arXiv1508.01991*, 2015.

[145] Z. Yu *et al.*, "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)*, 2015, pp. 338–345.

[146] M. Seyfioğlu, B. Erol, S. Gurbuz, and M. Amin, "DNN Transfer Learning From Diversified Micro-Doppler for Motion Classification," *IEEE Trans. Aerosp. Electron. Syst.*, vol. PP, p. 1, Dec. 2018, doi: 10.1109/TAES.2018.2883847.

[147] A. Shrestha *et al.*, "Cross-Frequency Classification of Indoor Activities with DNN Transfer Learning," in *2019 IEEE Radar Conference (RadarConf)*, 2019, pp. 1–6, doi: 10.1109/RADAR.2019.8835844.

[148] H. Li, J. le Kernec, A. Mehul, S. Z. Gurbuz, and F. Fioranelli, "Distributed Radar Information Fusion for Gait Recognition and Fall Detection," in *2020 IEEE Radar Conference (RadarConf20)*, 2020, pp. 1–6, doi: 10.1109/RadarConf2043947.2020.9266319.

[149] 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," *Med.*

Eng. Phys., vol. 42, pp. 1–12, 2017.

[150] R. P. L. Durgabai and Y. Ravi Bhushan, “Feature selection using ReliefF algorithm,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 3, no. 10, pp. 8215–8218, 2014.

[151] G. Roffo, “Feature selection library (MATLAB toolbox),” *arXiv Prepr. arXiv1607.01327*, 2016.

[152] L. Kuncheva and J. Rodríguez, *A weighted voting framework for classifiers ensembles*, vol. 38. 2014.

[153] L. I. Kuncheva, *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2014.

[154] Y. S. Huang and C. Y. Suen, “A method of combining multiple experts for the recognition of unconstrained handwritten numerals,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 1, pp. 90–94, 1995.

Appendix

University of Glasgow | College of Science & Engineering

Dr. Christoph Scheepers
Senior Lecturer
School of Psychology
University of Glasgow
58 Hillhead Street
Glasgow
G12 8QB
Tel.: +44 141 330 3606
Christoph.Scheepers@glasgow.ac.uk

Glasgow, January 29, 2018

Ethical approval for:

Application Number: 300170057

Project Title: Human Activities Classification and Fall Detection with Wearable Sensors and Radar

Lead Researcher: Dr Francesco Fioranelli

This is to confirm that the above application has been reviewed by the College of Science and Engineering Ethics Committee and **approved.** Please refer to the collated reviews on the system for additional comments, if any. Good luck with your research.

Sincerely,

Dr Christoph Scheepers

Ethics Officer

College of Science and Engineering

University of Glasgow