

## DOCTOR OF PHILOSOPHY

### Machine Learning for Human Activity Recognition Using Non-Intrusive Sensors

Karayaneva, Yordanka Lazarova

*Award date:*  
2022

*Awarding institution:*  
Coventry University

[Link to publication](#)

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of this thesis for personal non-commercial research or study
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission from the copyright holder(s)
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Coventry University  
Centre for Computational Science and Mathematical Modelling

**Machine Learning for Human Activity  
Recognition Using Non-Intrusive Sensors**



**Yordanka Lazarova Karayaneva**

A thesis submitted for the degree of  
*Doctor of Philosophy*

September 2021

## **Declaration**

All sentences or passages quoted in this document from other people's work have been specifically acknowledged by clear cross-referencing to author, work and page(s). Any illustrations that are not the work of the author of this report have been used with the explicit permission of the originator and are specifically acknowledged. I understand that failure to do this amounts to plagiarism and will be considered grounds for failure.

Name: Yordanka Karayaneva

Content removed on data protection grounds



## **Certificate of Ethical Approval**

Applicant:

Yordanka Karayaneva

Project Title:

Developing a holistic methodology for determining applications for robotics in elderly care homes

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Medium Risk

Date of approval:

23 July 2018

Project Reference Number:

P65532

Content removed on data protection grounds



Dedicated to my family.

## Acknowledgement

The past four years of my PhD journey have been very exciting and challenging, which helped me grow as a researcher and a person. I am extremely grateful for having being awarded the Data Driven Research and Innovation (DDRI) PhD grant at Coventry University, UK. I am pleased with the progress that I have made, which would not have been possible without the help of my family, supervisory team, academics at Coventry University, and friends.

First and foremost, I would like to thank my family for their outstanding support throughout my PhD, especially in times of struggle and despair. My mother Elena and my father Lazar have always valued education and supported me even when I was feeling lost. My sister Magdalena also provided support in both good and difficult times.

I would like to express the deepest gratitude to my Director of Studies, Dr. Sara Sharifzadeh, whose help and guidance were exceptional. She has helped me understand everything better with her amazing approach towards students. Additionally, she always responded to my questions even outside of working days and has been the kindest person to me. I would never be able to stress enough that I owe my success and publications mostly to her. She and Dr. Bo Tan were working with me during the New Year night 2020/2021 time in order to submit a publication to a reputed journal to meet the deadline.

I would like to extend my gratitude to Prof. Vasile Palade, Prof. Ala Szczepura, Prof. Yanguo Jing, Dr. Bo Tan, Dr. Diana Hintea, and Dr. Mark Elshaw for their support throughout my PhD. I greatly appreciate their guidance and professionalism. I would like to thank specifically Dr. Bo Tan, who continued his supervisory support even after leaving the university.

I would also like to thank my friend Uche Onyekpe, who supported me as a person and student for finalising the documentation for the thesis. I extend my gratitude to my friend Dimitar Popov, who supported me personally.

Finally, I would like to thank Coventry University for supporting this research financially, by allowing me to attend international conferences. I also thank Prof. Damien Foster for providing financial support for attending a conference outside of my PhD allowance.

## Abstract

Human activity recognition with non-intrusive sensors is an emerging topic in the field of computer vision, which has led to applications for supporting the older population. The current studies lack a holistic evaluation of data derived from infrared (IR) sensors including multiple layouts, sensors positions, noise analysis, multi-subject activities, and model generalisation. Micro-Doppler radars are also used extensively for human activity recognition, but the majority of studies fall in the supervised learning category. The very few studies associated with unsupervised human activity recognition with micro-Doppler radars suffer from a number of limitations such as intermediate accuracy and an exploration of a few techniques for feature extraction.

This thesis explores the use of IR sensors and micro-Doppler radars for human activity recognition for healthcare and eldercare applications. An investigation of such data with a variety of feature extraction and classification methods is achieved by using a number of datasets, comprising of multiple scenarios. Hence, these results address the shortcomings of the previous literature by providing a holistic understanding and evaluation of such data for healthcare purposes. The achieved results for the IR sensors data demonstrate the optimum model for feature extraction and classification, optimum sensor position and layout as well as a novel periodic noise reduction technique. The outcomes of this work bring us a step closer to the potential application of such systems in elderly care homes.

In terms of micro-Doppler radar data, unsupervised learning is studied considering its importance for unlabelled and poorly labelled projects. Two unsupervised feature extraction techniques are proposed, which are comparable with the existing Convolutional Variational Autoencoder (CVAE) architecture in terms of classification accuracy. The proposed methods provide a reasonable trade-off between computational time and accuracy, which makes them attractive for unsupervised applications.

Finally, a user acceptance questionnaire is administered in a care home to understand the views and needs of older adults towards sensing technology and physical robots. An extension to the Technology Acceptance Model (TAM) is achieved to cover more characteristics also seen from the perspective of health improvement. Trust and control, as well as physical appearance and functionalities of robots are studied. Two age-based groups are distinguished for statistical analysis with interesting findings.

To conclude, the techniques presented in this thesis significantly increase the usability of IR sensors and micro-Doppler radars for human activity recognition applications. This brings us a step closer to the application of these systems in care home, where their real world significance will be evaluated.

# Contents

<b>List of Figures</b>	<b>xv</b>
<b>List of Tables</b>	<b>xvii</b>
<b>List of Abbreviations</b>	<b>xix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation for Research . . . . .	2
1.2 Research Questions . . . . .	5
1.3 Thesis Novelty . . . . .	6
1.4 List of Publications . . . . .	6
1.5 Thesis Structure . . . . .	7
<b>2 Literature Survey</b>	<b>9</b>
2.1 Introduction . . . . .	9
2.2 Sensing Technologies . . . . .	10
2.2.1 Intrusive Technologies . . . . .	10
2.2.2 Non-Intrusive Technologies . . . . .	11
2.3 Analysis Techniques for Non-Intrusive Sensing Technologies . . . . .	15
2.3.1 Hand-Crafted Techniques . . . . .	15
2.3.2 Deep NN . . . . .	16
2.3.3 Supervised Learning . . . . .	18
2.3.4 Unsupervised Learning . . . . .	19
2.4 Evaluation Metrics for Sensing Technologies . . . . .	22
2.4.1 Supervised Metrics . . . . .	22
2.4.2 Unsupervised Metrics . . . . .	26
2.5 Evaluation Criteria for Sensing Technologies . . . . .	27
2.6 User Acceptance of Sensors and Robots . . . . .	29
2.6.1 User Acceptance of Sensors . . . . .	29
2.6.2 User Acceptance of Robots . . . . .	30
2.7 Chapter Conclusion . . . . .	33

<b>3</b>	<b>IR Sensing Array Based Human Activity Recognition</b>	<b>35</b>
3.1	Introduction . . . . .	35
3.2	Data Description . . . . .	36
3.2.1	Infrared Human Activity Recognition Dataset <i>Coventry-2018</i> . . . . .	36
3.2.2	Infrared Human Activity Recognition Dataset <i>Infra-ADL2018</i> . . . . .	38
3.3	Data Analysis Based on Sparse Techniques . . . . .	39
3.3.1	Methodology . . . . .	39
3.3.2	Data Analysis . . . . .	41
3.3.3	Experimental Results . . . . .	45
3.4	Noise Analysis and Classification Modeling for Holistic Understanding of LRIR Data . . . . .	47
3.4.1	Data Analysis Framework . . . . .	47
3.4.2	Evaluation and Results . . . . .	60
3.5	Chapter Conclusion . . . . .	71
<b>4</b>	<b>Unsupervised Doppler Radar Based Activity Recognition</b>	<b>73</b>
4.1	Introduction . . . . .	73
4.2	Methodology . . . . .	75
4.2.1	Dataset Description . . . . .	75
4.2.2	Number of Classes Estimation . . . . .	77
4.2.3	Feature Extraction Methods . . . . .	80
4.2.4	Clustering Methods . . . . .	85
4.2.5	Visualisation Techniques for High-Dimensional Data . . . . .	86
4.3	Evaluation and Results . . . . .	88
4.3.1	Local DCT-based Analysis Results . . . . .	89
4.3.2	Local Entropy Analysis Results . . . . .	90
4.3.3	CAE and CVAE Robustness Evaluation . . . . .	91
4.3.4	Comparison of the Average Training and Testing Accuracies for K-Means and K-Medoids Using all Feature Extraction Techniques . . . . .	91
4.3.5	Visualisation Results . . . . .	93
4.4	Chapter Conclusion . . . . .	95
<b>5</b>	<b>User Acceptance of Sensors and Robots</b>	<b>96</b>
5.1	Introduction . . . . .	96
5.2	Methodology . . . . .	97
5.2.1	Study Aims . . . . .	97
5.2.2	Survey Design . . . . .	97
5.2.3	Participants and Data Collection . . . . .	98
5.2.4	Analysis of Responses . . . . .	98
5.3	Study Findings . . . . .	99
5.3.1	User Acceptance of Sensors . . . . .	100
5.3.2	Factors Affecting the User Acceptance of Robots . . . . .	100
5.3.3	User Acceptance of Robots Based on Appearance . . . . .	102

5.3.4	User Acceptance of Robots Based on Role and Functions . . . . .	102
5.3.5	User Acceptance of Robots Based on Cost . . . . .	104
5.3.6	Open-Ended Question . . . . .	104
5.3.7	Comparison Between User Acceptance of Sensors and Robots . . .	104
5.3.8	Participants' Responses Based on Age . . . . .	105
5.4	Chapter Conclusion . . . . .	106
<b>6</b>	<b>Discussion</b>	<b>108</b>
6.1	Introduction . . . . .	108
6.2	Discussions on the Analysis of IR Data . . . . .	108
6.2.1	Limitations of the Extrapolation Method . . . . .	108
6.2.2	Optimum Model for Activity Recognition . . . . .	109
6.2.3	The Effect of Changes in Sensing Layout and Model Generalization	109
6.3	Discussions on the Analysis of Doppler Radar Data . . . . .	109
6.3.1	Analysis of Local Patching Strategy . . . . .	109
6.3.2	The Effect of the Clustering Strategy . . . . .	110
6.3.3	Optimum Models for Doppler Radar Data . . . . .	110
6.3.4	Fusion of Different Methods . . . . .	111
6.3.5	Analysis of Higher Standard Deviation . . . . .	111
6.3.6	High-Dimensional Data Projection and Visualisation . . . . .	111
6.3.7	Results Comparison with Supervised Strategies . . . . .	111
6.4	User Acceptance Analysis . . . . .	112
6.4.1	User Acceptance of Non-Intrusive Sensing Technology . . . . .	112
6.4.2	Older People's Acceptance of Robots . . . . .	112
6.4.3	Acceptance and Trust of Robot in Own Home . . . . .	113
6.4.4	Comparison Between User Acceptance of Sensors and Robots . . .	113
6.4.5	Comparison Between Age Groups . . . . .	114
6.4.6	Strengths and Limitations of Study . . . . .	114
6.5	Overall Discussions on the Use of Non-Intrusive Sensing Technologies for Health Monitoring of Aged Population . . . . .	114
6.6	Chapter Conclusion . . . . .	115
<b>7</b>	<b>Conclusion</b>	<b>117</b>
7.1	Research Questions and Thesis Contributions . . . . .	118
7.2	Human Activity Recognition Using IR Sensors . . . . .	119
7.3	Human Activity Recognition Using Doppler Radars . . . . .	120
7.4	User Acceptance of Sensors and Robots By Older Adults . . . . .	121
7.5	Research Limitations . . . . .	122
7.6	Future Directions . . . . .	122
7.7	Chapter Conclusion . . . . .	123
	<b>Bibliography</b>	<b>125</b>

<b>Appendices</b>	<b>144</b>
<b>A User Acceptance Questionnaire</b>	<b>145</b>

# List of Figures

1.1	Signs and symptoms of epileptic seizures [10]. . . . .	3
1.2	The projected stages of Parkinson’s disease [14]. . . . .	4
2.1	Bias-variance trade-off graph based on error [101]. . . . .	23
3.1	Layouts and experiment scenario. (a), Small and large experiment layouts; (b), The real experiment scenario. . . . .	37
3.2	Layout and experiment scenario. (a), Experiment layout; (b), The real experiment scenario. . . . .	38
3.3	Methodology steps for data analysis based on sparse methods. . . . .	39
3.4	Example of (a) a single frame of an IR image and (b) a stream of a number of frames. . . . .	40
3.5	Layout and experiment scenario for Sensor-1. (a), Experiment layout; (b), The real experiment scenario. . . . .	40
3.6	Accuracy of SVM, RF, KNN and LR on both training and testing sets for SPCA and SDA. . . . .	44
3.7	Visualisation of the original data matrix and the transformed samples using SPCA and SDA. . . . .	44
3.8	Accuracy of SVM, RF, KNN and LR on both training and testing sets for SPCA and SDA. . . . .	45
3.9	Comparison of the clean and noisy test data classification performances for RRLDA and SDA as well as PCA and SPCA. . . . .	46
3.10	Visualisation of the first sparse component of SDA $\beta_1$ and first sparse eigenvector of SPCA $V_1$ . . . . .	46
3.11	(a) Examples of $8 \times 8$ pixel frames for the activity sitting, and (b) the vectorised spatio-temporal 2D array over 40 frames and 64 pixels. . . . .	47
3.12	The data processing flow in this work, which shows the two undertaken strategies. . . . .	48
3.13	Data vectorisation and dimension description. One sample 2D spatio-temporal array is visualised in blue color in the middle. . . . .	50
3.14	The (a) one-subject and (b) two-subject vectorised $8 \times 8$ frames of activities (64) over 40 equalised frames. . . . .	51
3.15	The power spectrum with prominent peaks rounded in red. . . . .	52

3.16	Illustration of the 3D cube of validation accuracy and the 2D average validation array obtained by averaging over K folds of the 3D cube for selection of $T_h^*$ and $Num_h^*$ parameters. . . . .	54
3.17	Heatmaps of average CV validation accuracy for selection of optimum $T$ and $Num$ parameters for (a) horizontal (b) vertical stripes. . . . .	55
3.18	(a) Original image and (b) noise reduced image. . . . .	56
3.19	(a) Forming GLCM matrices in four directions i.e., $0^\circ$ , $45^\circ$ , $90^\circ$ , $135^\circ$ ; (b) Computation of GLCM matrix based on horizontal occurrences at $d = 1$ for an image. . . . .	59
3.20	Illustration of the CNN-LSTM architecture. . . . .	61
3.21	Comparison of the classification performances of random methods and proposed methods for interpolation and extrapolation before denoising (BD) and feature extraction using all four classification methods. The Coventry-2018 large layout data (15-class problem) of Sensor-1 was used. . . . .	62
3.22	Comparison of the classification performances before denoising (BD) and after applying the DFT-based periodic noise removal algorithm (AD) using different feature extraction methods and LR classification. The Coventry-2018 large layout data (15-class problem) of Sensor-1 was used. . . . .	63
3.23	CNN-LSTM model evaluation using Coventry-2018 large layout data (15-class problem) of Sensor-1. . . . .	64
3.24	Comparison of the effect of small layout and large layout data in classification performance. The LR classifier with the four feature extraction methods are applied on the single-subject activities for small layout and the single-subject activities for the large layout as well as CNN-LSTM using Coventry-2018 Sensor-1 large and small layout data sets. . . . .	66
3.25	(a) Single-subject activities and (b) double-subject activities. The vertical axis show the unseen test subject and the numbers show the average test accuracy over three repetition of each activity. . . . .	68
3.26	Patterns for activity 9 using all single-subjects. . . . .	69
3.27	Patterns for activity 2 using all double subjects. . . . .	69
3.28	Classification accuracy for individual sensors and combinations of them for Coventry-2018 dataset using SVD+LR on large layout for the 15-class problem including single and double subjects. . . . .	70
3.29	Comparison between the confusion matrices of the two side sensors, Sensor-1 (a) and Sensor-3 (b), of Coventry-2018 dataset. The large layout with single and double subject data was used for training a model based on SVD+LR. . . . .	71
3.30	Comparison between Coventry-2018 Sensor-1 (a) and Sensor-3 (b) for “Standing + Small Movements” activity. . . . .	71
4.1	Example of an $80 \times 80$ image of the activity walking (1). . . . .	74
4.2	Flowchart, describing the overall analysis framework of the study. . . . .	75
4.3	Experimental layout for conducting the activities in the used Doppler radar dataset. . . . .	76

4.4	Example of an $80 \times 80 = 6400$ image for each activity: (1) walking, (2) running, (3) jumping, (4) turning, and (5) standing. . . . .	76
4.5	An Elbow test used to determine the number of clusters $K$ . . . . .	78
4.6	Silhouette score, Davies score, Dunn’s index and Calinski-Harabasz index used to determine the number of clusters $K$ . . . . .	80
4.7	Flowchart of the proposed local DCT-based method. . . . .	81
4.8	The three patching strategies for entropy analysis. . . . .	82
4.9	Flowchart of the proposed local entropy-based method. . . . .	82
4.10	The CAE architecture. . . . .	83
4.11	The CVAE architecture. . . . .	84
4.12	(a) Illustration of the $40 \times 40$ local patches in the Doppler spectrogram. The optimum patch for feature extraction is ticked in green color. (b) The corresponding Dunn’s index for the four $40 \times 40$ local patches. . . . .	89
4.13	Illustration of the Dunn’s index for the three patching strategies used for feature extraction based on local entropy analysis. . . . .	90
4.14	(a) CAE and (b) CVAE robustness evaluation using the Dunn’s index based on average testing results. . . . .	91
4.15	Confusion matrices for DCT features using (a) K-Means, and (b) K-Medoids for different subjects. . . . .	92
4.16	Confusion matrices for CVAE encoded features using (a) K-Means and (b) K-Medoids for different subjects. . . . .	93
4.17	Visualisation of the first two components of (a) the raw data, and (b) the first components of the CVAE encoded data. . . . .	93
4.18	Visualisation of the transformed raw data $D_{n \times 6400}$ using t-SNE, MDS, and LLE. . . . .	94
4.19	Visualisation of the transformed encoded data $D_{n \times 50}$ using t-SNE, MDS, and LLE. . . . .	94
5.1	Health and well-being improvement based on sensing technology. . . . .	101
5.2	Relative importance of five robot characteristics . . . . .	101
5.3	User acceptance of a robot at home and trusting the robot for making decisions	102
5.4	User acceptance of robots based on physical appearance . . . . .	103
5.5	User acceptance of robots based on functions . . . . .	103
5.6	Health and well-being improvement based on robot’s activities . . . . .	104

# List of Tables

3.1	Two categories of activities in <i>Coventry-2018</i> . . . . .	37
3.2	Three categories of activities of <i>Infra-ADL2018</i> . . . . .	38
3.3	Investigated Activities . . . . .	41
3.4	The f-score and p-value results of the ANOVA analysis for the average horizontal and vertical stripes (excluding the DC components) of the 2D power spectrum for all three sensors (S1, S2, S3), and a combination of them (S1+S2+S3) using large layout and 15 classes data from Coventry-2018 dataset. . . . .	62
3.5	Testing performances for all combinations of the feature extraction and classification methods as well as CNN-LSTM using the Sensor-1 large layout data set from Coventry-2018 for 15 classes of single subject and double subject activities. . . . .	65
3.6	Testing performance of the classification models using SVD+LR for single-subject, double-subject, and all activities of <i>Infra-ADL2018</i> . . . . .	65
3.7	Comparison between single-subject activities and double-subject activities using Coventry-2018 large layout data of Sensor-1 and Infra-ADL2018 Sensor-2 data. LR was used together with all feature extraction methods as well as CNN-LSTM. . . . .	68
3.8	Testing performance of leave-one-subject-out model generalisation for single-subject activities and double-subject activities of <i>Infra-ADL2018</i> . . . . .	68
4.1	Average and standard deviations of testing accuracies based on K-Means and K-Medoids using DCT features from the raw data $80 \times 80$ single patch and those from the selected $40 \times 40$ local patch features over 4 rounds of LOOCV. . . . .	90
4.2	Average training accuracies with standard deviations of K-Means and K-Medoids using the five feature extraction methods over 4-subjects LOOCV. . . . .	91
4.3	Average testing accuracies with standard deviations of K-Means and K-Medoids using the five feature extraction methods over 4-subjects LOOCV. . . . .	92
5.1	All participants including their gender and age. . . . .	100
5.2	One-way ANOVA statistics for the two age-based groups described in Table 5.1 . . . . .	105
5.3	Spearman’s rank correlation coefficient and Kendall’s rank correlation coefficient for the two age-based groups . . . . .	106

6.1	Total computational time over 600 samples of the four folds for the two proposed feature extraction methods as well as CVAE. . . . .	110
-----	--	-----

# List of Abbreviations

**2DPCA** 2-Dimensional Principal Component Analysis

**AE** Autoencoder

**AIC** Akaike Information Criterion

**BIC** Bayesian Information Criterion

**BW** Bandwidth

**CAE** Convolutional Autoencoder

**CNN** Convolutional Neural Network

**CV** Cross Validation

**CVAE** Convolutional Variational Autoencoder

**DCT** Discrete Cosine Transform

**EM** Expectation-Maximisation

**FP** False Positive

**FN** False Negative

**GLCM** Grey Level Co-occurrence Matrix

**GMM** Gaussian Mixture Model

**HMM** Hidden Markov Models

**IR** Infrared

**k-NN** k-Nearest Neighbours

**LLE** Locally Linear Embedding

**LR** Logistic Regression

**LRIR** Low-Resolution Infrared

**LSTM** Long Short-Term Memory

**MDS** Multidimensional Scaling

**MLP** Multi-Layer Perceptron  
**MSE** Mean Square Error  
**NN** Neural Networks  
**PCA** Principal Component Analysis  
**PIR** Pyroelectric Infrared  
**PRC** Precision Recall Curve  
**PRIM** Patient Rule Induction Method  
**RF** Radio Frequency  
**RF** Random Forest  
**RNN** Recurrent Neural Network  
**ROC** Receiver Operating Characteristic  
**RRLDA** Reduced Rank Linear Discriminant Analysis  
**SDA** Sparse Discriminant Analysis  
**SOM** Self-Organising Map  
**SPCA** Sparse Principal Component Analysis  
**SSE** Sum of Squared Errors  
**SVD** Singular Value Decomposition  
**SVM** Support Vector Machine  
**TAM** Technology Acceptance Model  
**TP** True Positive  
**TN** True Negative  
**t-SNE** t-Distributed Stochastic Neighbour Embedding  
**VAE** Variational Autoencoder

# Chapter 1

## Introduction

Currently, an ageing population represents a global challenge. The problem is further exacerbated by the gradual fall of the working population due to lower birth rates. In addition, elderly care homes are faced with staff shortage due to unfavourable working conditions. The complicated problem with an expected ageing population has led to the pursuit of cost-effective solutions for human activity monitoring and recognition to supplement the envisaged considerably reduced carer to cared ratio. The detection and recognition of human activity is crucial in the field of early disease diagnosis, such as dementia, as well as fall detection. A diverse range of devices is proposed for this purpose such as traditional cameras, wearable technologies, acoustic sensors, infrared (IR) sensors, and radar sensors. When considering common constraints and limitations associated with cameras, wearables, and acoustic sensors, present healthcare provision has mostly adopted IR sensors and radar sensors. The selection of these two types of devices for human activity monitoring and recognition is attributed to their non-obtrusiveness, relatively small size, cost-effectiveness, and reliable recognition accuracies.

Human activity recognition is an emerging field of computer vision. The goal of human activity recognition is to detect and characterise a human activity, and more importantly, a deviation from normal activity, from sensor data in real world settings. Most commonly, human activity recognition is aimed at assistive applications for healthcare and eldercare. The challenging problem of detecting and recognising activities is exacerbated by the individual constraints associated with each sensor device for this purpose. While vision-based human detection is the most common area of research, the modern healthcare and eldercare block its usage due to privacy issues. Thermal-based and radio frequency-based human detection are emerging fields in the literature. Still, a number of limitations and shortcomings have been discovered, which block the continuous application of these approaches in real world scenarios.

This thesis explores the use of IR sensors and Doppler radar sensors for the detection and recognition of common human activities for the purpose of modern e-healthcare and eldercare. Considering the limitations of current research on both types of devices, improvements are proposed accordingly. In terms of IR sensors, a novel dataset is proposed to cover a

broad range of scenarios such as number of sensors, sensor positioning, layouts, and multi-subject activities. Additionally, a novel method for noise alleviation is developed as well as a holistic consideration of machine learning approaches for feature extraction and classification. Radar sensors are considered in terms of unsupervised learning, which is a far less researched area for human activity recognition. Based on this study, two novel methods for unsupervised feature extraction are proposed, which show a recognition improvement over the state-of-the-art approaches. Finally, user acceptance of non-obtrusive technologies and robots with a physical presence are evaluated among older adults in a residential care home with interesting findings.

## 1.1 Motivation for Research

The need to support the older adults is continuously growing with the increasing rates of ageing population. An ageing population is a global challenge, which is characterised by the increase of older people compared to the decrease of younger generations due to a decline in fertility and increased life expectancy. The most recent report by the United Nations (UN) identified 727 million people aged 65+ globally in 2020 [1]. This number is projected to increase by more than double to 1.5 billion by 2050. A previous report by the UN identified that two thirds of older people are reported to reside in developing countries, which disproves the common assumption of an ageing population being restricted to developed countries [2]. While the percentage of older populations in Europe and Northern America is higher, the same processes extending lifespan are gaining control of the population in the remaining world. The pessimistic prognosis is further complicated by the lack of trained staff in elder-care. The shortage of staff has been exacerbated by lack of job satisfaction, stress at work, and no clear career progression [3]. Data from various countries report disproportional results for older generations living independently or with families. Despite these results having been generated mostly by cultural differences, recent results show that older people are becoming more likely to live independently. Home care is also seen to be a far cheaper option than care homes. The ageing population issue is inevitably evident in the United Kingdom (UK) as well. The Office for National Statistics (ONS) reported 11.8 million UK residents aged 65+ in 2016 [4]. As with the global trends, this number is expected to grow to 20.4 million by 2066.

Major concerns surrounding ageing populations are neurological health problems, most commonly dementia leading to a cognitive decline. Other neurological diseases affecting people's movement include epilepsy, Parkinson's disease, and Motor Neurone Disease (MND). Dementia is an umbrella term for a number of neurocognitive diseases including vascular dementia, dementia from Parkinson's disease, dementia with Lewy bodies, frontotemporal dementia, Creutzfeldt-Jakob disease, mixed dementia, Huntington's disease, Traumatic brain injury, and most commonly Alzheimer's disease [5]. Currently, there are approximately one million people suffering from dementia in the UK. This number is projected to double to over two million in 2051. Needless to say, the prevalence of the disease

leads to huge economic costs. The total cost for dementia in the UK is estimated at £26.3 billion and is one of the largest impediments in the NHS [6]. Early motor-related signs and symptoms of dementia include coordination difficulties and disorientation [7]. In addition, changes in behaviour are observed, which can lead to disrupting daily schedule routines such as repeating or missing activities. Therefore, human activity recognition can be useful for detecting these early symptoms of dementia.

Regarding epilepsy, it is a chronic disease known to affect one in 100 people in the UK, while one in every new four diagnosed is over 65 [8]. Epileptic seizures can be life-threatening, however, many of them can be prevented with the accurate treatment. In addition, seizures are often accompanied with motor-related symptoms [9]. The overall signs and symptoms of epileptic seizures are illustrated in Figure 1.1. As observed, motor-related symptoms include uncontrollable jerking movements as well as sudden falls. Hence, activity recognition models can be useful for detecting the onsets of some epileptic seizures.

Some materials have been removed from this thesis due to Third Party Copyright. Pages where material has been removed are clearly marked in the electronic version. The unabridged version of the thesis can be viewed at the Lanchester Library, Coventry University

---

**Figure 1.1:** *Signs and symptoms of epileptic seizures [10].*

Another similar disease to dementia that affects older adults and specifically their movement is Parkinson's disease. Adults over 50 are usually affected by the disease, while one in 20 adults with the condition developed Parkinson under 40 [11]. The symptoms of Parkinson include involuntary shaking of different parts of the body (also known as tremor), stiff and inflexible muscles, slow movement, loss of automatic movements, and impaired posture and imbalance potentially leading to falls. [12]. With the increasing rates of ageing population, the prevalence of Parkinson's disease is estimated to increase by 18% between 2018 and the quarter of the century in the UK [13]. The projected stages of Parkinson's development shown in Figure 1.2 reveal the need for early intervention.

Some materials have been removed from this thesis due to Third Party Copyright. Pages where material has been removed are clearly marked in the electronic version. The unabridged version of the thesis can be viewed at the Lanchester Library, Coventry University

**Figure 1.2:** *The projected stages of Parkinson's disease [14].*

Furthermore, MND is a uncommon neurological condition, which negatively affects the muscles. In terms of movements, weakness in legs, a weak grip, muscle cramps and twitches are examples of some of the symptoms. [15]. The disease usually affects adults in their 60s or 70s, however, it can affect people in all age groups. In addition, early disease diagnosis can be useful for reducing the influence of MND's symptoms on one's life.

When considering the huge economic costs needed to support ageing populations across the world, governments are investing into cost-effective solutions for both home monitoring and eldercare monitoring solutions. A number of devices are capable of obtaining monitoring data for the purpose of activity recognition, respiration and vital signs monitoring or gait analysis. A straightforward example of activity recognition is a fall detection system, which can notify care staff if an accident occurs. In addition, common daily activities such as sitting, standing, walking, and others can reveal important patterns for the subject. The obtained information can subsequently be used for early disease diagnosis. The devices considered for activity recognition are cameras, wearable technologies, acoustic sensors, IR sensors, and radio frequency (RF) radars.

There is a plethora of current literature covering vision-based approaches for human activity recognition. While these approaches can be useful for healthcare applications for posture correction for instance, the modern eldercare blocks its use due to privacy issues. In addition, IR sensors and radar sensors are currently being examined, but the findings lack some important practical aspects related to the use of such sensing systems. For example, research investigating IR sensors has major shortcomings including a lack of an evaluation of multiple sensor scenarios [16, 17, 18, 19, 20, 21, 22]. The existing studies focus on a

single scenario, which leads to the lack of holistic evaluation. Noise reduction in the sense of periodic noise is not evident in any of the previous studies. In regard to radar sensors for human activity recognition, unsupervised learning, that uses machine learning algorithms to analyze and cluster unlabelled datasets, is a beneficial area despite its unpopularity. Unsupervised learning can be applied to scenarios with poor or no data labelling, especially in cases of large and diverse datasets. Only two studies exist for the purpose revealing mediocre recognition tasks, which is a source of motivation for further research [23, 24]. In addition, these studies also lack an evaluation of a broad number of methods, which can provide means for comparison. Furthermore, user acceptance of non-obtrusive sensing technology and robots with a physical presence have been studied as individual paradigms [25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39]. In addition, perceived health improvement, trust, cost, physical appearance and preferred activities conducted by robots are less researched among Long-Term Care (LTC) residents. This thesis also aims to provide a comparison between user acceptance of the two types of technologies as well as an evaluation on a broad range of factors, which can influence the potential adoption of the two agents.

Taking into consideration the limitations of previous research that poses barriers to the potential application of human activity recognition in healthcare and eldercare, this thesis presents novel machine learning approaches for noise alleviation and feature extraction as well as a holistic understanding and evaluation of such sensing systems. Therefore, the findings and contribution of the current work will advance the field of passive human activity recognition for healthcare, specifically using IR sensors and radar sensors for healthcare. Finally, user acceptance by older adults in terms of sensing technology and robots with a physical presence reveals key findings for a meaningful interaction.

## 1.2 Research Questions

The research questions addressed in this thesis are as follows:

1. *“Is it possible to develop novel approaches for data analysis and machine learning to correctly classify human activities derived from non-intrusive remotely sensed data?”*

By “*data analysis*”, the author refers to a novel method for periodic noise alleviation as well as frames equalisation by the means of interpolation and extrapolation. “*Machine learning*” is used in the sense of approaches for feature extraction such as the two novel methods for unsupervised feature extraction of 2D Doppler radar images. The overall goal is to provide improvements to the recognition performances of the architectures in order to reach a significant level for potential applications in real world scenarios.

2. *“What factors influence the user acceptance of non-intrusive sensing technology and robots with a physical presence by older adults?”*

“*User acceptance*” refers to the potential adoption of non-intrusive sensing technology and robots with a physical presence among older adults. In terms of factors that influence the potential adoption, the Technology Acceptance Model (TAM) framework is used, which stresses on only two factors: (1) perceived usefulness, and (2) perceived ease-of-use. To extend it further, a number of new factors are included in this thesis, such as perceived health improvement, cost, functionalities, and physical appearance.

### 1.3 Thesis Novelty

The novelty of research presented in the thesis is in the form of data analysis and machine learning methods for human activity recognition. The data for this research is acquired from non-obtrusive sensors. In addition, a novel user acceptance framework of non-obtrusive technology and robots with a physical presence is developed and applied in care home settings.

The contributions presented in Chapters 3, 4, 5 can be summarised as follows:

- A periodic noise reduction method aimed for the 2D spatio-temporal maps of the low resolution IR (LRIR) streams, which is based on Discrete Fourier Transform (DFT) with model selection strategies evaluated with K-fold cross validation (CV).
- Two methods for unsupervised feature extraction of 2D Doppler radar images based on local Discrete Fourier Transform (DCT) and local entropy analysis respectively, which are evaluated with Dunn’s index as an unsupervised metric.
- A holistic understanding and interpretation of IR data and Doppler radar data for human activity recognition.
- A human activity dataset *Coventry-2018* acquired using IR sensor arrays with multiple capabilities, which address the shortcomings of previous datasets in the literature.
- A user acceptance framework of non-obtrusive sensors and robots with a physical presence among older adults, which includes comparison between the two agents, trust, cost, preferred appearance and activities, and acceptance evaluated from the potential improvement of physical health and mental health.

### 1.4 List of Publications

Chapters 3, 4, 5 contain findings of the following peer-reviewed and accepted publications:

Y. Karayaneva, S. Sharifzadeh, W. Li, Y. Jing and B. Tan, “Unsupervised Doppler Radar Based Activity Recognition for e-Healthcare,” in *IEEE Access*, vol. 9, pp. 62984-63001, 2021, doi: 10.1109/ACCESS.2021.3074088.

Y. Karayaneva, S. Sharifzadeh, Y. Jing, K. Chetty, B. Tan, “Sparse Feature Extraction for Activity Detection Using Low-Resolution IR Streams”, in IEEE International Conference on Machine Learning and Applications, Florida, US, 2019.

Y. Karayaneva, S. Baker, Y. Jing, B. Tan, “Use of Low-Resolution Infrared Pixel Array for Passive Human Motion Movement and Recognition” British Computer Society (BCS) Human-Computer Interaction Conference, Belfast, UK, pp. 1-5, 2018.

The following publications are currently under review at peer-reviewed journals:

Y. Karayaneva, S. Sharifzadeh, Y. Jing, B. Tan, “Human activity recognition for AI-enabled healthcare using low resolution infrared sensor data”, submitted for IEEE Access, 2022.

Y. Karayaneva, A. Szczepura, Y. Jing, “Conceptual framework for co-design of sensors and robots with older adults living in a UK long-term residential care facility”, submitted for Robotica, 2022.

## 1.5 Thesis Structure

The following Chapter 2 reviews the existing findings in three main domains: 1) human activity recognition using IR sensors; 2) human activity recognition using Doppler radars; 3) user acceptance of human activity recognition sensing technology and robots with a physical presence among older adults. An in-depth study of existing approaches of IR data are presented including their major limitations and constraints. The Doppler radar data for human activity recognition are evaluated from the perspective of unsupervised learning, which is a minimally studied research area. Overall, existing machine learning approaches are investigated based on two categorisations: 1) traditional machine learning and deep neural networks (NN); 2) supervised and unsupervised. Evaluation metrics for activity recognition models as well as evaluation criteria for sensing technologies are described as well. Finally, user acceptance of the aforementioned two types of agents is evaluated from individual perspectives.

Chapter 3 presents both the implementation of novel methods for data analysis and machine learning as well novel applications of existing methods to IR data. Novel methods for interpolation, extrapolation and periodic noise reduction are presented and evaluated. In regard to existing techniques, sparse feature extraction methods are applied for dimensionality reduction and variable selection. Additionally, a wide range of both unsupervised and supervised methods for feature extraction and classification are applied in the pursuit of the most accurate architecture.

Chapter 4 is concerned with unsupervised learning for the Doppler radar data. Two novel feature extraction methods are implemented and compared with state-of-the-art techniques. In addition, a novel application of the existing Convolutional Variational Autoencoder (CVAE) is presented. Moreover, a number of existing methods for finding the number of clusters are presented and compared.

Chapter 5 presents the novel user acceptance questionnaire framework. This includes an extension to the TAM framework as well as additional novel insights, which can influence the potential acceptance of sensing technology and robots with a physical presence in the care home settings.

The findings of Chapters 3, 4, 5 are evaluated in Chapter 6. The chapter includes analysis of the proposed findings as well as limitations of the work. Then, Chapter 7 provides a conclusion to the study with insights for future work. This is followed by a list of references and an appendix. Finally, Appendix A includes a copy of the user acceptance questionnaire used in Chapter 5.

# Chapter 2

## Literature Survey

### 2.1 Introduction

This chapter discusses the existing sensing technologies and their relevant AI methods for human activity recognition. The sensing technologies in this case are categorized into intrusive technologies and non-intrusive technologies. The available machine learning techniques for human activity recognition can be grouped into two categorisations: handcrafted (manual) techniques and deep learning techniques. In addition, evaluation metrics for both supervised and unsupervised activity recognition models will be distinguished and reviewed. Then, both intrusive and non-intrusive sensing devices will be evaluated considering key metrics such as ease-of-use, cost-effectiveness, comfort, energy consumption. Finally, user acceptance of sensors and robots by older adults will be included.

The modern healthcare is concerned with the privacy of patients. Intrusive technologies such as cameras and wearables are therefore not recommended for use in healthcare considering their privacy-violating limitation. On the other hand, IR sensors and radar sensors can be employed as non-intrusive devices, which will be outlined in depth. In regard to IR sensors, introductory studies exist, which lack evaluation and discussion of layout arrangements, number of sensors, multi-subject activities, and model generalisation. Furthermore, micro-Doppler radars are considered for human activity recognition. The current literature identifies only two studies concerned with unsupervised learning for micro-Doppler radars for the studied purpose. Hence, unsupervised learning is the topic of interest of this thesis.

In this chapter, the machine learning techniques for human activity recognition will be discussed based on the aforementioned categorisations. Handcrafted techniques are traditionally used by providing inexpensive computational time. On the other hand, the emergence of deep learning methods allowed their applicability to large, high-dimensional datasets with exceptional performances. Supervised learning methods for classification are used predominantly for human activity recognition, while feature extraction is either supervised or unsupervised. Unsupervised methodologies are rare for this purpose as described above.

A review of the evaluation metrics will also be performed for both supervised and unsupervised methods. Accuracy, k-fold CV, Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), confusion matrix, sensitivity, specificity, Receiver Operating Characteristic Curve (ROC), Precision Recall Curve (PRC), and F-score will be discussed for evaluating supervised methods. In terms of unsupervised methods, sum of squared errors (SSE), Silhouette score, Davies-Boulding index, Dunn's index, and Calinski-Harabasz index will be outlined. These metrics can be used for both evaluation of the clustering results and finding the number of clusters.

In addition, all sensing devices will be reviewed based on their key characteristics such as ease-of-use, comfort, cost, and size. Moreover, user acceptance of sensors and robots will be conducted to outline any differences. Important enablers and barriers for the introduction of aforementioned devices will be investigated initially. Then, studies for user acceptance by older adults will be evaluated including both prior interaction studies and short to medium-term interaction studies.

The literature review chapter is organised as follows: Section 2.2 will review both intrusive devices (Section 2.2.1) and non-intrusive devices (Section 2.2.2). Then, Section 2.3 will evaluate the techniques as hand-crafted feature extraction strategies (Section 2.3.1), deep learning (Section 2.3.2), supervised (Section 2.3.3) and unsupervised (Section 2.3.4). Evaluation metrics will be presented in Section 2.4 for both supervised methods (Section 2.4.1) and unsupervised methods (Section 2.4.2). The sensing devices will be reviewed in Section 2.5. Section 2.6 presents user acceptance of sensors (Section 2.6.1) and robots (Section 2.6.2) by older adults. Finally, Section 2.7 will conclude the literature review with the most important highlights.

## **2.2 Sensing Technologies**

### **2.2.1 Intrusive Technologies**

Cameras are an obvious solution for human activity detection, which have been widely used in gaming, healthcare, and security [40]. Thanks to the growing resolution (number of pixels per unit area) and frame updating rate, the video streams contain rich spatio-temporal information, which reveals the detailed and implicit features of human activities. For example, detecting respiration from slight periodical movement of nose and chest [41, 42] and extraction of walking gaits features [43]. Working together with the depth sensor, the camera is also widely used for facial articulated assessment for identification [44]. On the other hand, taking of the identification information, generates privacy issue that blocks the deployment of camera in residential area for continuous monitoring. In addition, the high definition and fast frame-per-second (FPS) rates result in resource-demanding data stream in network that limits large-scale deployment for daily in-home activity recognition.

Wearable sensors refer to the miniaturized kinetic sensors, which are used to capture human signals or sense their activities. Examples are accelerometers, gyroscopes or gravity sensors [45, 46] that are often embedded in the wearable devices like smartwatches and fitness trackers. These sensors accurately and timely measure the dynamic of body parts displacement, orientation and tilt, and are low-cost solution for monitoring patients with activity difficulties, assessing athletes' movements and performance [47]. However, the discomfort for wearing [48] and limited battery life [49] block the continuous use of them in daily life.

Application of acoustic sensors for activity recognition can be categorised into two genres: echo-based; spectrum feature based. The former genre employs similar signal processing techniques as radar, treats the reflected acoustic wave from the target as a distorted and delayed copy of the acoustic source and uses the Doppler shifts as the metric to interpret the subject activities [50]. The later genre detects activities via the feature of sounds from the event sources including falling, glass breaking, door closing, and others [51]. Their main limitation is the fact that basic to medium acoustic sensors suffer from a large number of false positives due to environmental sounds such as winds. On the other hand, the highly precise acoustic sensors capable of providing stable accuracies are often expensive [52].

To conclude, modern camera devices often suffer from subject identification privacy concern. This poses a problem for their implementation in private areas considering the negative user acceptance. While wearable technologies provide beyond human activity recognition – including vital signs monitoring – they are usually constrained by low battery life and patients forgetting or feeling uncomfortable wearing the device. In terms of acoustics, their signal is 1D and compared to 2D camera images, they are cheaper in terms of processing time. However, they are very prone to be mixed with lots of environmental sounds and also cannot detect health conditions that happen in silence without any sound. Considering the limitations of the aforementioned devices, IR sensors and RF radars become the point of interest for our research. Both devices obtain unobtrusive data, where subject identification is avoided. In addition, the success of IR sensors and micro-Doppler radars has been shown in the literature considering their high accuracy rates.

## **2.2.2 Non-Intrusive Technologies**

### **2.2.2.1 IR Sensors**

Heretofore, the pyroelectric infrared (PIR) and thermal camera are two types of popular IR sensors. While PIR sensors have been applied for fall detection [53], the binary output from the sensor only delivers coarse activity indication and also suffers from the false alarm caused by varying environmental brightness. The thermal cameras obtain the high resolution thermal pixels that can be used for constructing subject body profile and activity recognition. However, the use of thermal camera has also the privacy and cost issues. Therefore, in this thesis, characteristics of LRIR sensor are investigated that have the potential to provide low-cost, non-intrusive, and high-accurate activity recognition solution while preserving the privacy [17].

IR sensors are a type of thermopile IR sensors, which detect the IR rays (also known as IR energy) [54]. The thermopile sensors use small thermocouples deployed on a silicon chip. These small thermocouples can absorb the IR rays and produce a signal output. Passive IR sensors are known to have applications for non-intrusive human detection and recognition. These sensors deliver thermal images based on temperature emitted from subjects and items. As such, multiple human recognition is possible as well as human tracking. The main advantage of passive IR sensors is their non-intrusiveness compared with traditional cameras. Hence, they can be deployed in private areas such as bedrooms and bathrooms.

An important disadvantage of passive IR sensors is their sensing distance, which is usually with a distance range of up to several metres. Considering this constrain, these thermal devices can be employed in small to medium areas only. In addition, passive IR sensors provide extremely low resolution thermal images, where small fractions of the human body are difficult to be recognised [17]. Nevertheless, the low resolution can play the role of an advantage considering the inexpensive computational time. Another particular concern is the presence of ambient temperature when positioned in the field of view of the IR sensor. Sources of ambient temperature are usually heaters, air conditioners or warm weather.

Grid-EYE (AMG8833) sensor is a passive IR sensor produced by Panasonic [55]. The sensor has 60° field of view and provides 8 × 8 resolution outputs. The sensor distance is within 5-7 metres, while the refresh rate is 1 or 10 FPS (frames per second). Grid-EYE's temperature range of measuring object is 0°C to 80°C for the high gain type, while the same range is within -20°C to 100°C for the low gain type.

While human activity recognition is a constantly evolving field of machine learning, passive IR sensor arrays have also been proposed for this aim. The main areas of this field include applications for smart healthcare and eldercare. The applications of passive IR sensor arrays are still novice, which contributes to the lack of research on certain topics surrounding the issue. These topics include sensor positions, number of sensors, detection of multiple-subject activities, noise alleviation, sensor displacement, and model generalisation.

The majority of current studies deploy a single ceiling-mounted passive IR sensor for a number of single-subject activities [16, 17, 20]. These activities include standing, sitting, walking, falling, and others. An evident lack of research exists concerning optimal sensor position and number of sensors. In addition, multiple-subject activities are less prevalent than single-subject activities, which leads to a lack of comparison. While ceiling-mounted sensors can be useful for activities such as walking and falling, they pose difficulties for activities conducted in the same location. The number of deployed sensors is also not studied, which has important applications for multiple-subject activities. This is due to the fact that multiple-subject activities lack symmetry when captured by two side sensors. As such, optimal sensor selection can be applied in similar scenarios to discover the optimal sensor based on the most correctly classified multiple-subject activities.

An introductory study for IR sensor-based human activity recognition was focused specifically on recognising falling for eldercare applications [16]. A single wall-mounted sensor was deployed to classify fall or non-fall. Physical feature selection of four features was proposed. The four selected features included: 1) the number of consecutive frames with detected motion, 2) maximum number of pixels that temperature variance changed during step 1), 3) maximum temperature variance during step 1), and 4) distance of a maximum temperature pixel prior and after activities. Selected features were then classified with k-Nearest Neighbours (k-NN) with Euclidean distance as a measure. The authors used an older version of Grid-EYE with a sensing distance of up to 5 metres. The same authors extended their introductory study to cover more activities in addition to falling: no event, stopping, walking, and sitting [17]. However, the experimental settings remained identical to their first study. In addition, the same types of physical features were extracted for subsequent recognition. The extracted features were then analysed and classified using Support Vector Machine (SVM). The recognition results revealed 100%, 100%, 94.8%, 99.9%, and 78.6% for falling, no event, stopping, walking, and sitting respectively. Despite the limited number of samples, it can be concluded that sitting had the worst recognition result. This is in line with the fact that ceiling-mounted sensors are capable of detecting motion-related activities very accurately. However, when presented with an activity such as sitting, the recognition performance drops.

A similar study [20] improved the extracted physical features from Mashiyama's papers. By using the same experimental scenario with a single ceiling-mounted sensor, the authors achieved 97% testing accuracy for detecting humans instead of using human activities. In addition, the authors used the physical features from Mashiyama's work for comparison, which led to 70% accuracy.

Contrary to previous papers, a different study [18] used a single front far-infrared sensor with  $16 \times 16$  resolution. A dataset with various activities was collected including sitting, walking, extending arms, and others. The authors proposed thermo-spatial histogram feature extraction contrary to thermal histogram and spatial histogram. The proposed methods improved the recognition performance of the both previous state-of-the-art methods.

Noise alleviation methods for IR data are only seen in two studies [20, 21]. The authors in the first study [20] proposed a Kalman filter-based method for reducing noise in sensor recordings. While the Kalman filter method has shown robustness, it is difficult to be applied to real world scenarios due to its assumptions that both system and observation models are linear. In addition, the authors applied the existing background subtraction method to separate the target from the background. Background subtraction method was also applied in the second study [21]. For this purpose, clear background frames with no targets should be collected and subtracted from frames with targets and background noise. The background subtraction method assumes a static background, which has shown unsuitability in real world scenarios. This is due to the fact that real world scenarios are mostly dynamic occluded by multiple objects and factors. To the best of our knowledge, no previous papers consider the

alleviation of periodic noise from 2D spatio-temporal maps. In addition, no previous papers consider the use of sparse techniques for alleviating the effect of Gaussian noise and no algorithm was developed for the periodic noise effect seen in 2D spatio-temporal maps that will be further explained in Chapter 3.

### 2.2.2.2 Radar Sensors

Radio sensors refer to the active and passive radars working on different spectrum segments. The 2.4 GHz passive Wi-Fi radar and variants have generated big impact on recognizing the indoor body gesture and respiration, even in the through-the-wall condition [56]. Taking advantages of the wider bandwidth, higher sensitivity on target motion and diverse signal modulations, UWB (around 10GHz band) [57] and mmWave radars (on 24, 60 and above 70GHz bands) [58] achieve higher activity recognition performance and detect accurate bio-information such as respiration and heartbeat. But, the radio sensors have their own limitations. The high performance variance brought by interference and geometry uncertainty, makes the radio sensors difficult to be generalized in practice.

Doppler radars are a category of RF radars, which utilise the Doppler shift (effect) to detect motion-related activities [59]. Such activities include walking, running, jumping, and others. The Doppler shift is related to activities involving motion, which cause a shift in the frequency. As such, this phenomenon can be used to capture and detect these activities. Doppler radars can be used for both detecting the activities and locating the position. They work by sending a transmitted pulse and receiving echo. Therefore, the change of the frequency is measured, which corresponds to motion-related activities.

Doppler radars have shown suitability for human activity recognition due to their deep penetration, non-intrusiveness, and reliable accuracy rates. In addition, micro-Doppler radars have been used for respiration monitoring [60], and gait pattern analysis [61]. Regarding human activity recognition, micro-Doppler radars have been used for healthcare purposes [23, 24, 60, 62].

Most of the current studies for Doppler radar based human activity recognition are based on supervised learning, which requires the knowledge of the true labels. Hence, the frameworks are supervised, which consist of unsupervised feature extraction followed by supervised classification. While unsupervised learning can be used in scenarios with poor labelling, it has not received enough attention.

The current literature for Doppler radar-based unsupervised human activity recognition identifies only two papers. The first attempt to solve the unsupervised scenario for Doppler activities data is seen in a previous study [23]. The introductory paper used the existing Elbow method for distinguishing the number of activities. The number of activities were seen as unknown as the ground truth labels were seen as not present. The Elbow method is a visualisation technique, where a different number of potential candidate numbers of activities are given. For each, the method calculates the mean square error (MSE) and a noticeable decline

indicates the selected number of activities. The method can often be seen as ambiguous and it was not compared with other number of clusters estimation techniques. Then, state-of-the-art physical feature extraction followed by Hidden Markov Models (HMM) classification was applied. The achieved average classification accuracy was 69%.

The same authors extended their study with a more sophisticated analysis [24]. Log-likelihood HMM was implemented as a supervised feature extraction method. The features were then clustered with the unsupervised clustering methods K-Means and K-Medoids. In addition, comparison was performed for different state-of-the-art feature extraction methods including physical feature extraction, Singular Value Decomposition (SVD), and Principal Components Analysis (PCA). The overall accuracies were 79% and 80% for K-Means and K-Medoids respectively using the newly derived feature extraction method, which presents an improvement to their previous work.

## 2.3 Analysis Techniques for Non-Intrusive Sensing Technologies

The techniques for analysis of data derived from remote non-intrusive sensors are based on two categorisations: 1) hand-crafted methods, and deep NN methods; (2) supervised methods, and unsupervised methods. In addition, the described techniques can be used for human activity recognition purposes based on different criteria. More specifically, hand-crafted features can be utilised in small to medium datasets, while deep NN approaches are seen as useful for larger datasets (including more activities and participants). Furthermore, supervised approaches are applicable for human activity systems with knowledge of true labels. On the other hand, unsupervised methods will allow the clustering of a wide range of daily activities occurring in real settings, where only crucial activities such as a fall or immobility can be labelled.

### 2.3.1 Hand-Crafted Techniques

Hand-crafted feature extraction methods are common in human activity recognition applications for IR data and Doppler radar data. These methods offer cheap computational time combined with stable recognition performances. Two common categories of feature extraction methods are prevalent in the literature: physical features, and SVD/PCA.

A common method for unsupervised feature extraction used for both IR and radar data is selecting physical features from the thermal images and Doppler spectrogram respectively. In regards to IR data, four thermal image feature were selected by Mashiyama et al. in their study [16], as described above, for detecting falls. Similar physical features were extracted in previous studies [19, 20]. In terms of Doppler radar data, the pioneering study by Kim and Ling [59] proposed the following six Doppler spectrogram features: 1) torso Doppler frequency; 2) total bandwidth (BW) of the Doppler signal; 3) offset of total Doppler; 4) the

BW without micro Dopplers; 5) normalized standard deviation of Doppler signal strength; 6) period of the limb motion. As observed, the method retained a very limited number of features compared to the original high-dimensional data. Still, the recognition performance revealed more than 90% testing accuracy using the state-of-the-art classification method such as SVM.

Another group of methods used for feature extraction of IR data and Doppler radar data are PCA and SVD. Both PCA and SVD have the capacity to decompose the original features to low-dimensional principal components, which retain the most of the data variance. In regards to Doppler radar data, variations of PCA have been applied in the pursuit of a more accurately measured covariance matrix. The authors in a previous study [63] applied L1 norm PCA opposed to standard PCA, which led to improved classification performance. 2DPCA has also been explored as a variation of PCA, which accepts 2D matrices instead of vectors [64]. 1D-PCA requires the input to be vectorised, which often leads to vectors with high-dimensionality. As such, the large covariance matrix is difficult to evaluate and requires more computational time in comparison with 2DPCA. Previous work for Doppler radar data has shown improved testing accuracies with 2DPCA feature extraction. In addition, unsupervised multilinear PCA was combined with Linear Discriminant Analysis (LDA) and Shallow Neural Networks (SNN) [65]. This work for Doppler radar data marked the first attempt to include a 3D representation data of joint-variable signal of Doppler data in order to retain the original dependencies. Results were optimistic compared to PCA and 2DPCA.

Unlike previous studies for IR-based human activity recognition, thermo-spatial sensitive histogram was used in a single study [18], which was based on feature point extraction seen in many high-resolution vision-based applications. In addition to IR applications, the frequency-based method Discrete Cosine Transform (DCT) was applied to extract both temporal and spatial features from each sequence [21].

In summary, previous works mostly utilised physical features, and SVD/PCA including their variations for feature extraction. However, the IR images individually or concatenated into 2D spatio-temporal maps along with Doppler spectrograms can be treated as images. As a result, image texture analysis methods, that have not been used previously, can be employed. The previous solutions for human activity recognition also lack the use of novel methods for feature extraction as opposed to the off-the-shelf techniques.

### 2.3.2 Deep NN

The extracted features using handcrafted methods have increased over time as the methods are applied to more complex problems. Therefore, this leads to expensive computational time for both training and testing steps, which is problematic in real world scenarios. The current literature presents a single study concerned with the application of deep NN approaches to IR data for human activity detection [22]. More precisely, the authors proposed a 3D Convolutional Neural Network (CNN) architecture with achieved 97.22% overall accuracy.

This recognition performance served as an improvement in comparison with the state-of-the-art feature extraction and classification approaches used for the same data in a previous study [21]. On the other hand, Doppler radar based applications for human activity recognition have a rich history of deep learning approaches. These approaches include deep feature extraction or a combination of deep feature extraction and deep classification.

The main deep architectures deployed in Doppler radar based human activity recognition applications include CNN, Recurrent Neural Networks (RNN), Autoencoders (AE), and Hybrid deep models. CNNs and Stacked AE (SAE) compared to SVM for classification were used [66]. Despite the deep models' complexity, SVM showed a slight improvement to the recognition performance for the first dataset used in the study. In addition, CNNs were used with three layers each having 20 filters for classifying human activities [67]. The deep CNN architecture reported more than 90% average accuracy performance for human activities detection.

A hybrid CNN-Long Short-Term Memory (LSTM) architecture was deployed for detecting human activities using Doppler radars [68]. The proposed architecture combined the advantages of hierarchy-based CNNs and temporal feature extractors of LSTM. The hybrid architecture showed improved performance compared to both hand-crafted and deep learning techniques such as SVM, CNN and CAE. In addition to hybrid architectures, an RNN-LSTM architecture was deployed for detecting six human activities [69]. The inability of RNN to process long sequences is compensated with the introduction of the LSTM structure to the architecture. Recognition results revealed improvement using the proposed architecture in comparison with deep CNN.

AEs are common unsupervised NNs, which consist of an encoder and a decoder. The decoder aims to reconstruct the original data representation given the reduced dimensionality from the encoder. A stacked AE was used for recognising human activities including falls derived from Doppler radar [70]. The unsupervised feature extraction structure was followed by a softmax classifier, which required the labels as input. Similarly to this study, convolutional autoencoder (CAE) for human activity recognition of Doppler data was recently deployed [71]. The proposed architecture utilised the advantage of CNN in the sense of feature detector and the unsupervised feature extraction of AE. Similarly, it was followed by the supervised softmax classifier.

In conclusion, there have been limited studies on the use of deep NN strategies for IR-based human activity recognition and more specifically their use has been limited to the simple CNN architectures. However, other more advance architectures such as combinations of CNN with LSTM for automated feature extraction and prediction, or deep AEs for dimensionality reduction have not yet been utilized. On the other hand, radar-based applications for the similar purpose have a richer history of using deep NN architectures. However, the majority of them fall in the supervised learning category opposed to the unsupervised learning, where the latter is one of the topics addressed in this research.

### 2.3.3 Supervised Learning

Supervised learning methods learn from a large number of training examples, each with a given ground-truth output known as label in the case of classification or target in the case of prediction problems [72]. For classification, given a number of concrete classes, each training example is assigned a label. So that, each data sample is given as an input-label pair consisting of a feature vector and a label. On the other hand, regression problems are defined by a given real-valued output instead of a label. Supervised classification includes both traditional machine learning methods and deep NN. Human activity recognition applications with IR sensors and Doppler radars usually employ supervised classification based on hand-crafted techniques or deep NN. The features for such systems can be extracted based on supervised or unsupervised methods, where the latter is more common.

Reduced Rank Linear Discriminant Analysis (RRLDA) is a common supervised feature extraction technique [73]. The goal for the method is to minimise the within-class distance and maximise the between-class distance. An important constraint of the method is the Gaussian assumption of data. As such, non-Gaussian data can lead to poorer results using RRLDA. Supervised PCA is another popular method for supervised feature extraction [74]. The supervised variation of PCA estimates the PCs by their maximum dependence on the response variable.

Probably the most common supervised classifier is SVM. It was invented in early 1960s [75]. Improved later, the method has shown substantial results for classification, especially with the use of kernels [76]. The use of kernel in its objective function allows classification of datasets with non-linear nature. Another widely used supervised classification method is k-NN. It was developed in 1951, which follows a basic rule for nearest neighbours [77]. The method most commonly uses Euclidean distance as a metric, while other distances can also be applied. As described earlier, both SVM and k-NN have been previously used for IR sensor and Doppler radar based activity recognition applications. That is due to their simplicity and efficiency. The two classification approaches have shown stable accuracies, which have been comparable to the more complex deep NN approaches. Additional methods include Random Forest (RF) and Logistic Regression (LR). RF represents an ensemble of a number of decision trees, where voting takes place. Inspired by Amit and Geman [78], the researcher in a later study [79] proposed RF by showing the importance of individual tree classifiers and the connections between them. LR was originally used as a binary classifier indicating the probability of a certain class or event. Later, the LR method was successfully implemented for multi-class problems [73].

Deep learning exploration of supervised techniques begins with the development of a single-layer perceptron in 1958 [80]. The network consists of only input and output layers lacking any hidden representations. As such, the single-layer perceptron has shown applicability to linear problems. However, the network failed at non-linear problems such as the XOR function, which required a hidden layer.

Considering this major shortcoming of single-layer perceptrons, multi-layer perceptrons (MLPs) were invented in the 1970s [81]. These networks introduced a hidden layer and had the ability to separate non-linear data. Backpropagation based on gradient descent optimization was introduced as the main method for training MLPs. The backpropagation method consisting of two steps, forward pass and backward pass, allows the network to learn, and has now become the most common network training method. Since MLPs faced strong opposition in the face of simpler methods such as SVM and k-NN, the field was temporarily abandoned [82]. However, the emergence of deep learning allowed these networks to receive more attention. An important breakthrough was the invention of faster GPUs, which tangibly reduced the training computational time for deep learning architectures. Since then, deep NNs have proven their recognition advantage compared to SVM and k-NN as well as their capability of gradual improvement with the addition of more training data.

In summary, supervised learning solutions for human activity recognition have been widely researched and proposed. These include both traditional machine learning architectures and deep NN architectures, where the latter found popularity later leading to extreme emergence in the literature. Classical machine learning architectures are less computationally expensive and more easily implemented. Despite the emergence of deep learning, these classical architectures still remain popular and have shown significant results, especially for smaller training datasets. That is also the case for IR sensors data. Most previous research works have used these groups of techniques and only one previous research was concerned with supervised deep NN classification. Therefore, in this research, deep NN architectures will be employed and compared with other supervised strategies in order to find the most optimum solutions for IR data.

### 2.3.4 Unsupervised Learning

Unsupervised learning is a major field of machine learning, which does not require the true labels as an input [73]. This, on the other hand, can negatively affect the classification performance as only the data samples are used as an input. However, unsupervised frameworks are desirable when labelling large datasets is too costly. There are a variety of unsupervised learning subfields, where the most widely known is cluster analysis or also known as data segmentation. Since unsupervised learning does not require the labels, in data segmentation the data samples are designated in different groups / clusters based on their inputs and characteristics. The goal for cluster analysis implies that data samples from the same cluster are more related to one another. On the other hand, data samples from different clusters should be more distinct.

Reviewing the previous studies show that there have been few studies, where unsupervised clustering methods were used for human activity recognition. Especially, there are not any cases for IR data and in terms of Doppler radar based human activity recognition, only two papers were found. As described in Section 2.2.2.2, the two studies are focused on unsupervised clustering by using hand-crafted techniques in comparison with the plethora

of supervised classification approaches for the same purpose. On the other hand, the feature extraction was mainly performed unsupervised for radar data. For example, physical feature extraction or SVD/PCA have been used. Considering (1) the diverse number of daily activities and the fact that few of them might be important for recognition (2) the difficult labeling of large number of recorded data, unsupervised strategies are important to tackle these problems. Therefore, in this work, their use for radar sensors data will be addressed.

Three types of clustering methods exist: combinatorial methods, mixture modelling, and mode seeking. Currently, only combinatorial clustering methods have been used in the area of human activity recognition using Doppler radars. Combinatorial methods for clustering only consider the input variables for each data sample excluding any underlying probability [83]. The goal is to assign each data sample to only one cluster based on similarity. An important constraint of combinatorial methods for clustering is the number of clusters  $K$  to be known priori and  $K < N$ , where  $N$  is the total number of samples. The methods use distance, most commonly Euclidean distance to assign the data samples to each cluster. Two common clustering approaches are K-Means and K-Medoids, which have been applied previously for unsupervised Doppler radar based activity clustering [24].

Hierarchical clustering methods omit the constraint of user to provide the number of clusters [84]. In hierarchical clustering, the lowest level assumes each data sample as a distinct cluster, while the highest level represents one large cluster, where all data samples are assigned. Two categories of hierarchical clustering exist based on level starting point: agglomerative clustering and divisive clustering. Agglomerative clustering begins from the lowest level by merging clusters based on similarity at each step. On the other hand, divisive clustering starts from the large cluster at the highest level and divides it to more clusters. Hierarchical clustering is a less researched area of clustering and has been largely constrained by the lack of precise objective functions. A major development in hierarchical clustering was the proposal of cost function minimisation by Dasgupta [85]. Inspired by Dasgupta's work, researchers in a later study [86] analysed the definition of a "good" objective function by considering a number of functions for both similarity graphs and dissimilarity graphs. Therefore, the definition of a precise objective function will allow the implementation of improved approximation algorithms.

Mixture modelling methods are clustering methods based on probability. Multivariate Gaussian Mixture Models (GMM) are the most common mixture methods used in the literature [87]. These methods assume a number of Gaussian distributions. Each distribution can be viewed as a single cluster or group. The goal is to define a probability for a specific data sample to belong to a certain distribution. In addition, Poisson distribution [88] and Mises-Fisher distribution [89] have enjoyed some popularity for a number of applications including document classification, text and gene expression. Expectation-Maximisation (EM) algorithm is commonly applied in cases of missing or incomplete data.

Mode seeking methods are a less researched field of clustering methods. Mean shift clustering and Patient Rule Induction Method (PRIM) are two popular mode seeking meth-

ods [73]. The first approach for mode seeking methods proposed in 1975 uses mean shift for density function estimation [90]. Mean shift is used as a non-parametric approach for positioning the modes. The algorithm assigns each data sample to the closest modes. The modes in this approach can be viewed as clusters. On the other hand, the PRIM starts from the highest level, where all data samples belong to one large cluster or box. The feature space in this method is known as a “box”. To complete terminology, “face” is used at each step of the method, which is usually the highest box mean. At each peeling step, some of the data samples are removed based on the face. This results to a specific minimal number of data points in the box. Therefore, the process is repeated until a satisfactory number of boxes is achieved. However, the mean shift has been highly constrained by its expensive computational time. Later, k-NN was proposed, which improved the computational time [91]. In addition, to improve the performance of the method on high-dimensional data, a novel method has been proposed to directly estimate the ratios of density derivatives [92].

Unsupervised learning methods can be categorised as manual or automated methods for feature extraction and decision making. For manual techniques, prior information about the experiment’s conditions and environment are required [93]. In spatio-temporal feature extraction techniques, the signal strength and angle of measurement can affect the filtering window size and scaling or choice of basis function. On the other hand, automated methods do not require prior knowledge about the experiment and remove the need for manual settings [94, 95]. Nevertheless, the computational time required for training automated methods is noticeably more expensive compared to manual methods.

Another group of unsupervised methods are based on NN. The idea of competitive learning, opposed to error-correction learning such as backpropagation, was first introduced in the 1970s [80]. Competitive learning implies that the output neurons compete with one another to be activated. In addition, only one output neuron can be active at a time. Later, Kohonen introduced Self-Organising Maps (SOMs), which are based on the notion of competitive learning [96]. The SOMs are largely based on the human brain’s cerebral cortex – an area responsible for the sensory inputs such as visual or auditory information. SOMs are used for dimensionality reduction of high-dimensional data to a relatively low-dimensional output. In SOMs, the networks learn by shifting its weights to the active neurons. More specifically, only the active neurons are allowed to learn.

AEs are a powerful category of unsupervised neural networks for dimensionality reduction, commonly used for denoising [97]. Their main disadvantage is the fact that they learn local parameters of each data sample and include only a reconstruction term. To address these limitations, Variational AEs (VAEs) were introduced recently [98]. VAEs include a regularisation term and introduce a probability distribution for each data sample. The probability distribution is normally a standard distribution with zero mean and variance of one. In this sense, it ensures that data samples belonging to the same class are given similar representations in the Euclidean region. As described in Section 2.2.2.2, variations of AEs including CAE and SAE have been used for unsupervised feature extraction for Doppler radar based

human activity recognition. While these deep approaches are more complex, they have provided improvements compared to their hand-crafted competitors. In terms of IR sensors, while unsupervised approaches are common for feature extraction, they usually fall in the hand-crafted category.

In conclusion, unsupervised methods have not been used widely for human activity recognition despite their advantages. Therefore, the existing literature lacks a comparison between different unsupervised methods for clustering. While the manual data labelling is expensive, unsupervised approaches can provide clustering of activities in the absence of labels. Since, in practice, recognition of only few groups of extreme health condition are of interest, unsupervised clustering results can be used together with few extreme activity labels in this case. Unsurprisingly, the disadvantage relates to the fact that the unsupervised methods are only guided by the characteristics and variations of data samples and, this constraint usually leads to poorer accuracy in comparison with supervised architectures. However, a semi-supervised framework based on clustering of activities followed by the use of the labels of extreme condition can be considered.

## 2.4 Evaluation Metrics for Sensing Technologies

This section discusses metrics for evaluating activity recognition models. In particular, the generalisation ability is studied, which is a crucial capability of every model. Generalisation is the model ability to predict new data to a reasonable level. Most commonly, the data in machine learning is divided to three sets: training data, validation data, and testing data [73]. The manner of data split is usually by random order. Training data are used for training the model, while validation data are used for finding the optimal model's parameters. On the other hand, the optimum model is tested with the use of the remaining testing set. The metrics can be distinguished in two categories: 1) supervised metrics; and 2) unsupervised metrics.

### 2.4.1 Supervised Metrics

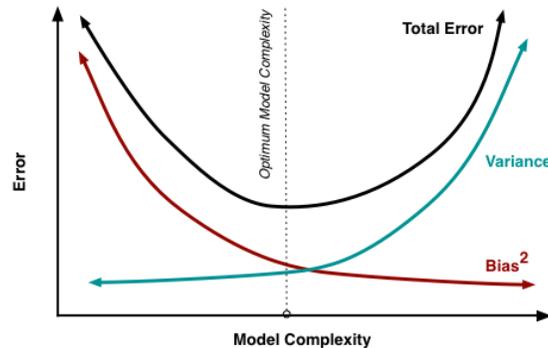
Classification accuracy is one of the most common metrics for both binary and multi-class supervised learning problems [99]. The metric is highly applicable for balanced human activity recognition datasets. It is calculated by dividing the number of correct predictions by the total number of predictions as given in the following equation:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

where  $TP$  and  $TN$  refer to true positives and true negatives respectively, while  $FP$  and  $FN$  denote false positives and false negatives respectively.

As such, the accuracy varies in the range of 0 to 1 or 0% to 100%. An important note for accuracy is the equal number of data samples for all classes. Hence, if the overall accuracy is considerably high, it can be assured that the model performance over all classes is reasonable. The train accuracy shows the model performance over the same data used for developing the same model and usually is the highest accuracy. The validation accuracy shows the optimum accuracy that is achieved on the validation process as the model parameters were varied. On the other hand, the test accuracy shows the model performance over the unseen data during the training. A well-trained generalized model should show very close accuracies over the train, validation and test data.

The model accuracy is a function of model complexity [100]. Poor train and test accuracies can be due to a high bias in the model. A high level of bias indicates a low complex model that is under-fitted. For example, when the number of variables are not enough and more information in the form of additional features or variables are required. It also might be due to the lack of enough samples. On the other hand a model can suffer from over-fitting issue that is an indicator of high complexity. An over-fitted model accuracy is very good on the training and is low in the case of testing data. Such model is not generalized to work on unseen future data. The aim of validation process and model selection strategies is to find an optimum model that does not suffer from any of the over- and under-fitting issues. This is called the bias-variance trade-off, which is illustrated in Figure 2.1 based on error.



**Figure 2.1:** Bias-variance trade-off graph based on error [101].

One of the common techniques for tuning the optimum model parameters and selection of the best model parameters is K-fold CV [102]. It is also widely used for estimation of the prediction error. Usually, K-fold CV is deployed especially in cases with reasonably large datasets. A whole number  $K$  is selected, which denotes the number of subsets of the original data. Most commonly, 5 or 10 are applied for  $K$  in CV. By using  $K = 5$  CV, five presumably equal subsets of the data are used. At each step, the model is fit to the remaining  $K - 1$  subsets of the data. Therefore, prediction error is estimated for each step with a total number of  $K$  steps. Finally, the total prediction error is calculated based on the average of errors over the  $K$  steps. For the purpose of model selection, several candidate parameters of the model are varied over the CV loop. Then, the dimension of the average error array is equal

to the number of parameters and based on that, optimum model parameters corresponding to minimum average error are found.

A common question is the correct selection of  $K$ . The value for  $K$  should be chosen in a way that the samples in the validation fold are reasonable in number. That allows validating the trained model without having bias or variance issues. In condition that a large number of samples compared to the number of variables exist,  $K$  can be chosen as a large number e.g. 10. However, when the number of samples are limited,  $K$  should be small, so that the validation fold will not be so limited in the number of samples. The authors in a previous study [102], who considered  $K = 5$  CV, observed that the classifiers improved their performance by increasing the dataset to 100 observations. In addition, increasing the dataset to 200 can provide only a minor improvement to the previous result. It can be concluded that bias is of a particular concern for smaller applications. With  $K = N$  CV, the training sets are extremely similar, which explains the higher variance in the validation errors over the folds. However, the results are unbiased for this case, which is an advantage. Considering the computational complexity, it is obvious that  $K = N$  CV is more expensive than  $K = 5$  or  $K = 10$  CV.

Bayesian information criterion (BIC) is a statistical method for model selection based on maximisation of log-likelihood function estimation [102]. The method, also known as Schwarz information criterion (SIC) was proposed by Schwarz in 1978, by using a Bayesian approach [103]. BIC is similar to Akaike information criterion (AIC), which is another statistical method for assessing a model's quality [104]. More specifically, BIC and AIC are given as:

$$\begin{aligned} BIC &= \ln(n)k - 2 \ln(\hat{L}) \\ AIC &= 2k - 2 \ln(\hat{L}) \end{aligned} \tag{2.2}$$

where  $n$  refers to the number of data observations, while the number of parameters estimated by the models is denoted by  $k$ . In addition, the maximisation of the log-likelihood function of the model is given by  $\hat{L}$ .

A common property for both BIC and AIC is the ability to penalise models, which are given a higher number of parameters. This is important due to the fact more parameters can lead to over-fitting. As such, these techniques provide a good trade-off between model complexity and model goodness. However, it is worth noting that BIC provides heavier penalties if the model's complexity increases. In addition, the model with minimum BIC corresponds to the model with the maximum posterior probability. On the other hand, AIC computes the difference between the data likelihood and the true likelihood. Similarly, the AIC with the lowest value is the optimum model. A particular limitation of BIC is the fact that it imposes a selection of simpler models due to its penalties nature. Considering this limitation, AIC usually tends to choose more complex models due to its penalty function.

Confusion matrices, also known as contingency tables or tables of confusion, are used for analysing classification accuracy per class, which are both applicable to binary and multi-

class problems [105]. In addition, confusion matrices serve as means for identifying problematic classes by providing a cross-section of actual class versus predicted class. The tables of confusion provide the true positives, true negatives, false positives, and false negatives. By having this detailed analysis in hand, commonly misclassified classes can be identified and studied.

Two additional important evaluation metrics are sensitivity, also known as recall, and specificity [106]. Both methods are commonly used in the field of medical test diagnosis for binary problems. The sensitivity and specificity are calculated as follows:

$$\begin{aligned} \text{Sensitivity/Recall} &= \frac{TP}{TP + FN} \\ \text{Specificity} &= \frac{TN}{TN + FP} \end{aligned} \quad (2.3)$$

ROC [107] and PRC [108] are visualisation methods, which illustrate the model at all thresholds. Similarly to sensitivity (recall) and specificity, ROC and PRC are commonly used for binary problems. More specially, the ROC curve represents a plot between true positives rate (TPR), also known as sensitivity or recall, and false positives rate (FPR). The definition of FPR is given as follows:

$$FPR = \frac{FP}{FP + TN} \quad (2.4)$$

The PRC aims to show the direct representation of precision and recall (sensitivity). This is extremely useful in cases of imbalanced data, where the number of negatives dominates over the number of positives [109]. As such, the user will be more interested in discovering the positives rather than the negatives. A common example is imbalanced medical data, where the majority of patients do not have the disease in question due to its low prevalence among the population. Precision is given by:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.5)$$

The F-score provides the Harmonic mean of precision and recall [110]. It is discovered that the F-score provides better results for projects with higher sensitivity. On the other hand, projects with higher specificity will suffer when using this measure. The method is commonly used for handling imbalanced classification data. It is defined as follows:

$$F - \text{score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.6)$$

where recall and precision were given in equation (2.3) and equation (2.5) respectively.

In summary, classification accuracy and confusion matrices are utilised in this thesis. This is due to the fact that the thesis involves multi-class problems with balanced data. In addition, confusion matrices are used to identify the misclassification between two or more classes that provide further insight about the problems.

## 2.4.2 Unsupervised Metrics

Unsupervised learning's main property is the absence of true labels [83]. As such, the discussed metrics such as classification accuracy, K-fold CV, BIC, AIC, confusion matrix, sensitivity, specificity, ROC, PRC, and F-score are not applicable. On the other hand, unsupervised metrics exist, which are primarily based on the two important characteristics of cluster analysis. These are the minimisation of data points distances belonging to the same cluster/class, and maximisation of the data points distances, which belong to different clusters. As distances are considered, different types for distance computation can be used. For the studies in this thesis, Euclidean distance is applied. Two-dimensional Euclidean distance is given as follows:

$$d(p, q)^2 = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2} \quad (2.7)$$

where  $p$  and  $q$  have the coordinates  $p_1, p_2$ , and  $q_1, q_2$  respectively.

In addition to the unknown true labels in unsupervised learning applications, the number of clusters is also unknown. As such, the described methods below can serve two purposes. Firstly, they can show the quality of a model, which can be used for model selection of unsupervised methods. Secondly, they can be used in a combination with a clustering method such as K-Means and K-Medoids for number of clusters estimation [111].

Sum of squared errors (SSE) can be used as a metric for identifying the clustering quality [112]. More specifically, SSE estimates the sum of squared Euclidean distances between an observation and the centre of the cluster, also known as centroid. The centroid is usually represented by a central vector and it is used by a wide range of methods such as the popular clustering methods K-Means and K-Medoids. For K-Means, the centroid is not necessarily a data sample. On the other hand, K-Medoids, which can be implemented as a similar distance-based clustering method, requires the centroid to be a data sample. SSE is given as follows:

$$SSE = \sum_{i=1}^n (x_i - \bar{x}) \quad (2.8)$$

where  $n$  refers to the total number of observations,  $x_i$  represents a single observation, and  $\bar{x}$  is the cluster centroid/centre.

Silhouette score is probably the most common technique for estimating the quality of a clustering model [113]. In addition, the method can also be used for graphically estimating the fit of a data point to a cluster, and number of clusters estimation. Silhouette coefficients considers two important values: the mean distance between a sample and all other samples in a cluster; the mean distance between the same sample and all other samples belonging to the second nearest cluster. A higher value for Silhouette coefficients is desirable.

Davies-Bouldin index is another clustering evaluation metric, which considers the similarities between clusters [114]. Differently from other metrics, a smaller value corresponds to dissimilar clusters, which is desired in cluster analysis.

Dunn's index is a popular method for evaluating clustering methods [115]. It refers to the minimum inter-cluster distance, which is divided by the maximum cluster size. Logically, the method has constraints in cases with a higher inter-cluster distance and a small maximum cluster size. Hence, this can lead to a large value, which can be misleading in some cases.

Calinski-Harabasz index is a method for evaluating clustering methods [116]. It is given as a ratio of the between-cluster sum of squares and within-cluster sum of squares. As such, Calinski-Harabasz index measures both separation by evaluating the maximum distance between clusters centroid and compactness of the data points belonging to the same cluster.

There are cases similar to our proposed project, where the true labels are known, but unsupervised learning is applied for analysis. By having the true labels and estimated cluster assignments, a number of techniques can be applied to match them. This is needed due to the fact that cluster assignments order do not necessarily follow the original order of the true labels. The most popular methods for this purpose are homogeneity, completeness, and V-measure [117]. Two important constraints are noticeable for homogeneity and completeness. In regard to homogeneity, each cluster needs to contain only members of a single class. Completeness implies that all members of a given class have to be assigned to the same cluster. Due to these constraints, the results can be misleading and unexpected since they are linked to accuracy. V-measure, on the other hand, can provide more accurate results, but it also has an important constraint. If the true labels and predicted labels are swapped, the result will be identical.

To conclude, a number of clustering methods for finding  $K$  and estimating the “goodness” of clustering results are used in this thesis. As the true labels are known, supervised metrics can be used such as accuracy and confusion matrix considering that the problem is multi-classified with balanced data.

## 2.5 Evaluation Criteria for Sensing Technologies

The aforementioned sensor devices are evaluated in terms of comfort, ease-of-use, energy, and cost in this section. Their metrics are seen from the perspective of the end-users, which are the older adults in this context. A major review study identified that older adults showed willingness to accept a new technology mostly based on the perceived health and environment benefits [48]. In addition, cost and perceived need were found crucial for adopting a new technology by older adults [118].

The rapid development of modern technologies has allowed the dissemination of more affordable, cost-effective, and portable sensors [119]. Hence, the smart sensors can relieve

the need for staff monitoring, while providing important features such as human activity recognition, vital signs monitoring, or gait pattern analysis. While the user acceptance of ambient sensors is constrained by cost and maintenance, the acceptance of wearable sensors is more complex [120].

In terms of comfort, wearable sensors can be directly evaluated as they interfere with the user. Ideally, wearable sensors should be with minimal weight and size, while causing a less intrusive sense to the user's skin. The most common wearable system is the combination of an accelerometer with a gyroscope for human activity recognition [45]. Wearable sensors can be attached to different parts of the human's body including most commonly wrist, chest, and waist. Wrist-attached sensors can recognise basic activities, while motion activities are best detected by waist-mounted sensors, which cover a larger part of the human body. [121]. Additional body parts for sensor attachment include foot and thigh for leg-related activities [122]. Considering all other studied sensors, as they are less-obtrusive and interaction-free devices, older adults have generally shown moderate to high user acceptance [26, 27]. The comfort of wearable sensors can be regarded as mixed considering a study with 28 subjects [123]. On the other hand, wrist-worn wearable sensors for activity tracking were highly admired by the participants group ( $n = 20$ ) in a different study [120]. As observed, the number of participants in each study was mostly limited, which can potentially lead to bias in the results.

The ease-of-use is an important factor for user acceptance. In the context of sensor systems with applications in the eldercare, installation [119], device maintenance and management [124], and interaction [125] are of particular concern. In terms of installation and device maintenance, it was reported that the majority of sensor systems require professional installation and frequent maintenance. The interaction factor mainly affects private home-based older adults and elderly care homes' staff. As identified earlier, a study has outlined the need for training requested by care home staff when approached with potential interaction of any new technology [126]. Furthermore, care home staff have particularly expressed concerns with operating and maintaining technologies due to lack of special knowledge.

Cost-effectiveness is an important factor for user acceptance of sensor systems, which includes device cost, infrastructure components cost, installation and maintenance cost [127]. However, wearable sensors are required for each older adult, which, on the other hand, increases the overall cost. In addition, it was reported that multiple sensors are needed for each person to detect more complex activities [122]. Acoustic, IR, and radar sensors have also shown a cost decrease with the mass production of such technologies. While acoustic sensors are usually inexpensive, a broad number of devices is needed for optimal accuracy. Hence, this can increase the overall cost of the system [122]. A single radar sensor can be required in one room due to their deep penetration and high distance range [128]. Likewise, IR sensors do not cause huge economic burden, but they often suffer from low distance range [16]. In this regard, more IR sensors can be required to cover one large room.

Energy consumption is crucial in the context of portable devices used for older adults. Low battery life for wearable devices has been identified as an important barrier [129]. The higher battery life corresponds to a bigger in size battery, which then poses a different issue. IR sensors have been identified as low power consumption devices [122]. Micro-Doppler radars are shown to leverage the existing RF bursts in the environment, which lowers the used power consumption [130]. Wi-Fi radars, on the other hand, are high power consumption devices, which limits their long-term usage [122].

## 2.6 User Acceptance of Sensors and Robots

The introduction of sensors [131], and robots [132] in the eldercare has been emerging in the recent years. Ideally, user acceptance of any technology should be studied prior the development and introduction. Nevertheless, user acceptance of such technologies among residents of care homes has largely been ignored. In terms of sensors, while the studied solutions in this thesis are less intrusive, factors influencing their user acceptance are still minimally researched. Considering the very diverse real settings with a number of furniture, occupants, and heating source, an IR sensor or a Doppler radar can be mounted to the body of a robot for human activity recognition. As such, the robot can provide interactive sensing of subjects, while its user acceptance among the residents is crucial.

Current studies are focused on identifying the potential enablers and barriers to the introduction of (1) sensors and (2) robots in the care home settings or private home. A small qualitative study [126] investigated the enablers and barriers for older adults ( $n = 9$ ) and care home staff ( $n = 24$ ) towards sensing systems. The major enabler identified by the older adults group was safety enhancement. On the other hand, the care home staff expressed concerns for lack of knowledge to use the technology. Besides emergency sensors for safety enhancement, sensors capturing daily activities for subsequent analysis were not addressed due to the lack of discussions for the potential benefits.

The most recent research, which included analysis of 12 qualitative studies for socially assistive robots, was focused on enablers and barriers for robots. The study identified four enablers and three barriers [133]. The enablers included enjoyment, usability, personalisation, and familiarisation. On the other hand, technical problems, limited capabilities, and negative preconceptions were found to be the major barriers. Similar study focused explicitly on the zoomorphic robot Paro identified the social stigma as a primary barrier [134].

### 2.6.1 User Acceptance of Sensors

Sensors in general are emerging in almost every field of daily life. As such, a number of studies have outlined the general user acceptance of these devices among the end-users mostly in healthcare applications [135, 136, 137]. A major review discussing 31 studies for sensors user acceptance in different fields of healthcare found overall acceptance of sensors in

the sense of telemedicine research [138]. However, the individual studies had a number of differences in relation to sample size, time of interaction, and healthcare field, which posed issues for drawing conclusions.

The literature investigating the user acceptance of sensors among the older population is very limited. A pilot study ( $n = 13$ ) focused on private home-based residents found that sensors are desirable in cases where no interaction is required [25]. The authors of the study stressed that no prior interaction of the study participants with sensors was conducted. As such, the views, perceptions, and attitudes of the older adults were less affected by previous experience. In regards to potential benefits of sensor systems, fall detection was regarded important by the participants. Likewise, a similar study ( $n = 11$ ) found the capability of detecting falls crucial [26]. This finding corresponds to the interaction-free preference as participation by the older adults is not needed. Since both studies included focus groups based at private homes, they found the sensors as a potential tool for prolonging their independent stay at home. In addition, cost was found to be a major concern for the participants in these studies. This is due to the fact that they live independently and might be required to cover the cost by their own means. Wearable sensors were evaluated explicitly in a small study ( $n = 5$ ) with mainly positive responses [28]. However, the same study only evaluated this technology failing to provide comparison between different types of systems. Nevertheless, a Switzerland-based study ( $n = 34$ ) reported low to moderate positive user acceptance for a wireless sensor system [29].

To conclude the findings of the limited literature, older adults highly value technologies for emergency situations, which omit the need for any specific interaction. Independently living participants identify cost as a major barrier if expected to be covered by themselves. In addition, more investigation is needed with potential comparison of sensor systems with robots with a physical presence.

## 2.6.2 User Acceptance of Robots

The use of robots in human social spaces is becoming increasingly common. Ideally, human perceptions and opinions towards the introduction of any new technology should be considered by developers prior to the design and modelling stage [139]. This will ensure that a newly implemented technology leads to fewer errors and less likelihood of providing misleading information, which, in its turn, will make the technology more usable and satisfactory. The term ‘robot’ is defined as a machine—especially one programmable by a computer—capable of carrying out a complex series of actions automatically. Robotics involves design, construction, operation, and use of robots. Because many robots are now accepted in homes and wider society, consideration of their acceptance by the elderly population can be made purely based on their existing characteristics [140].

Due to the fact that robots are regarded as a new type of technology, several countries have surveyed ‘general acceptance’ by their populations. A 2012 survey of citizens in European Union (EU) countries [141] found that, although acceptance varied between countries,

robots were generally accepted by the majority (70%) of EU citizens, with men more likely to demonstrate a positive view (76% versus 65% of). In terms of age, those between 15 to 24 years old reported 79% acceptance of robots, with this falling to 62% among citizens aged 55 and over. Higher education level, interest in technology and science, and personal experience with robots were also key factors that increased user acceptance. In terms of the fields and/or institutions where robots should be used, the vast majority of EU citizens considered that robots should perform jobs that are too hard and dangerous for humans. Priority areas were space exploration, manufacturing, military and security, and search and rescue. The least accepted fields were care (children, elderly and disabled), leisure and education. Overall, 60% of EU citizens believed that robots should be banned in the area of care although 76% agreed with the statement that robots are good for society because they help people. A very high proportion of people (91%) agreed that careful management of robots is needed. This was closely connected to perceived impact on employment; 70% considered robots could steal people's jobs and only 39% agreed that robots will boost job opportunities.

A similar study has examined the views of American citizens towards robots [142]. This more recent research mainly focused on the potential for automation to impact on jobs in various sectors, with only 33% "enthusiastic" about the concept of human replacement. Mirroring the European findings [141], the majority of Americans (85%) favoured use of robots in settings that are too dangerous and unhealthy for humans. Similarly, the least accepted businesses for automation were healthcare and education.

Classical user acceptance models such as the TAM only consider the perceived ease-of-use and usefulness of a new technology. For robots, however, other hedonic principles have been identified such as enjoyment and attractiveness, especially when considering social robots [143]. Physical attractiveness is therefore another important property which could appeal to users in a positive way. A further study investigating humanoid robots acceptance ( $n = 15$ ) found that although the majority of such robots were criticised by the participants, some small humanoid robots were highly admired. [144].

The literature focused on user acceptance of robots among elderly residents is richer compared with studies for sensors user acceptance. However, the majority of studies investigate the attitudes towards a specific robot with a physical presence following short to medium term interaction. Hence, this potentially results in bias in elderly people's perceptions and attitudes. To begin with, the robot's orientation in the context of elder care has been studied ( $n = 71$ ) [145]. The study mainly identified the lack of conversation of potential activities and their benefits related to robots interaction. More specifically, the study stressed on how potential benefits of robots should be related to real world scenarios. Similarly to the studies identifying the barriers, another study reported the limited capabilities of such agents as pivotal [34].

### 2.6.2.1 Robot's Physical Appearance

The robot's physical appearance is crucial as it presents the first impression. It can be categorised in the following categories: android, human-like, mechanical humanoid, zoomorphic, and mechanical. An important concept in the field of robot's physical appearance is the term "uncanny valley" coined by the Japanese professor Mori [146]. The uncanny valley identifies that mechanical humanoid robots are highly likely to cause familiarisation among people. On the other hand, when humanoid robots possess human's attributes such as skin, hair, and nails, this can easily lead to revulsion among people.

Anthropomorphism (translation from Greek: "human form") is particularly relevant when robots resemble humans and are used in social tasks. Attribution of human characteristics to non-human entities means that people's expectations towards robots are principally formed based on their own perceptions, emotions and beliefs [147]. This is closely linked to user acceptance, with the assumption that if robots resemble humans physically, mentally and technically more closely, the chances of acceptance increase. Research about the reasons behind attribution of human characteristics to robots concludes that it makes a non-human entity more familiar, friendly, and predictable for people [148].

A France-based study ( $n = 25$ ), which included all physical appearance categories for evaluation among older adults [30], has found correlations with Mori's uncanny valley. The most preferred category was mechanical humanoid, followed by zoomorphic (mechanical animal-like, and animal-like) and mechanical robots. Android and human-like robots were ranked at the bottom of the table.

These findings are in line with two other studies, which evaluated the user acceptance of mechanical humanoid robots Kompai ( $n = 11$ ) [31] and NAO ( $n = 19$ ) [32]. The first study reported the views of Kompai following one-month interaction. While the interaction reported mainly negative responses, the physical appearance was highly admired by the residents. The short interaction with NAO reported high preferences for its physical appearance. In regards to zoomorphic robots, a study ( $n = 42$ ) compared a 1) zoomorphic robot cat, with a 2) physical monitor [33]. The findings revealed stronger acceptance for the zoomorphic robot.

To conclude, android and human-like robots might easily "fall" in Mori's uncanny valley, which is highly undesirable in the context of eldercare. The studies found that mechanical humanoid robots are the most preferred category followed by animal-like robots, while mechanical robots find less sympathy.

### 2.6.2.2 Robot's Role and Capabilities

The role of robots is highly related to the capabilities they possess. Robots can be distinguished in two main categories in terms of their role: 1) physical robots, and 2) socially-assistive robots [132]. Physical robots are known to provide manual tasks such as cleaning,

cooking, washing, environmental navigation, and others. On the other hand, socially assistive robots directly interact with people by providing companionship. Generally, it is concluded that older adults prefer physical robots with little to no interaction [149]. The requirement for no interaction has a strong correlation with the user acceptance of sensor systems, where the same attitude is found. Still, socially assistive robots play an important role for meaningful interaction and have found their applicability in the eldercare context.

Studies in the literature identified both positive responses for physical robots with minor interaction and socially-assistive robots. A robot concerned with helping older adults to drink water ( $n = 10$ ) found positive attitudes [34]. On the other hand, a socially assistive mechanical humanoid robot Matilda ( $n = 115$ ) found strong preferences for the activity dancing [35]. The same study reported the willingness (60%) of the older adults to participate in a group activity with Matilda. Similar findings were reported by a study ( $n = 46$ ) related to the social robot Brian [36]. The majority of the participants valued the social interaction with the robot.

The literature for socially-assistive robots consists of studies, which included comparison between two types of robots: 1) social robots, and 2) less social robots. The more social robot used in the experiments usually possessed stronger communication skills compared to its less social competitor. More specifically, communicative robots were preferred to non-communicative robots reported by a focus group ( $n = 40$ ), which included interaction with two versions of a robot [37]. The results show that older adults felt more comfortable and expressive with the more social version of the robot. A further study by the same authors ( $n = 40$ ) reported increased enjoyment, sociability, and social presence by the more communicative robot compared with a non-communicative version [38]. A similar study ( $n = 28$ ) in Japan developed two versions of a wheelchair robot based on its social capabilities [39]. Likewise, the findings reported the same pattern of views among the older adults.

To conclude, physical robots are highly preferred as they require little to no interaction. Still, social robots have found their presence in the context of eldercare by the introduction of many similar agents. The studies mainly reported positive responses, while communicative social robots were more admired than non-communicative robots.

## 2.7 Chapter Conclusion

The above discussion outlined a number of sensing devices for the purpose of human activity recognition in eldercare and healthcare. In addition, the analysis strategies were reviewed. With the rapid increase of ageing population worldwide, the need for supporting elderly populations with cost-effective technologies is constantly growing. However, some of the discussed devices suffer from intrusiveness, which is highly undesirable in the context of healthcare and blocks their usage.

IR sensors were discussed in depth by focusing on their non-intrusiveness and relative cost-effectiveness. These devices possess great potential for human activity recognition, but the current literature lacks in-depth evaluation of different layouts, multi-subject activities, number of sensors, and sensor displacement. Currently, IR sensors are shown to detect activities such as walking, sitting, standing, falling with more than 90% accuracy rates. In addition, noise alleviation methods and machine learning techniques require further research. Micro-Doppler radars are extensively researched with supervised learning methods. However, supervised learning is expensive in terms of labeling time, while unsupervised learning is an appealing area for research.

The current techniques for human activity recognition were outlined with the major developments in the field of machine learning. Evaluation metrics for both supervised and unsupervised methods were presented and discussed. In addition, evaluation was presented for sensing devices with key factors discussed. Finally, user acceptance and evaluation criteria of both sensors and robots were presented, which are crucial for any technology introduction in eldercare.

# Chapter 3

## IR Sensing Array Based Human Activity Recognition

### 3.1 Introduction

The previous chapter was aimed at reviewing the current findings in terms of human activity recognition with both LRIR sensors and micro-Doppler radars. The existing techniques for these devices were categorized based on handcrafted feature extraction strategies and the automated feature extraction based on deep learning methods. Furthermore, supervised and unsupervised modeling strategies as well as the evaluation metrics were reviewed. In addition, the previous studies on user acceptance of both sensing technologies and robots with a physical presence were reviewed.

This chapter is focused on the interpretation of LRIR data for human activity recognition. As outlined in the previous chapter, the current studies suffer from a number of limitations. In particular, noise reduction methods for 2D spatio-temporal maps are not widely researched, which is pivotal for the proposed work in this chapter. In addition, multiple sensor layouts, sensors positions, multi-subject activities, and model generalisation in cases of sensor replacement as well as leave-one-subject-out analysis are not evident in the current literature. The proposed contributions of this chapter are summarised as follows:

- A periodic noise reduction method aimed for the 2D spatio-temporal maps of the LRIR streams, which is based on DFT with model selection strategies evaluated with K-fold CV.
- A holistic understanding and interpretation of IR data for human activity recognition including analysis of multiple layouts, number of sensors, multi-subject activities and model generalisation.
- Methods for interpolation and extrapolation aimed at equalising the number of frames per IR sensor recordings.

- A human activity dataset *Coventry-2018* derived from IR sensing arrays with multiple capabilities, which address the shortcomings of previous datasets in the literature.

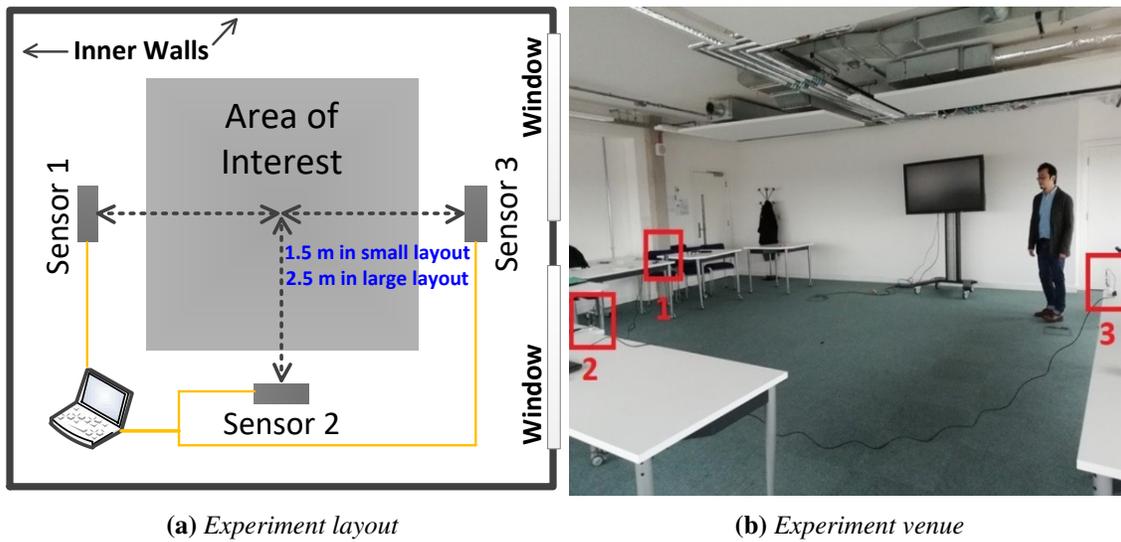
The next section of this chapter will introduce the proposed dataset *Coventry-2018* as well as the model verification dataset *Infra-ADL2018*. Then, dimensionality reduction with sparse methods applied only on a subset of *Coventry-2018* dataset will be considered. In addition, this analysis will include noise tolerance, where sparse methods are known to outperform their non-sparse equivalent techniques. Secondly, the remaining analysis will be concerned with periodic noise reduction as well as analysis of the aforementioned limitations in previous research. This stage of analysis will use the entire *Coventry-2018* dataset including different subsets of it, and *Infra-ADL2018* dataset.

## 3.2 Data Description

### 3.2.1 Infrared Human Activity Recognition Dataset *Coventry-2018*

*Coventry-2018* dataset was collected in Faculty of Engineering, Environment and Computing, Coventry University, United Kingdom in March 2018. Three Panasonic® Grid-EYE (Model: AMG8833) sensors, described in Section 2.2.2, were used for conducting the experiment. The three sensors included two side sensors and one front sensor corresponding to position in regard to the area of interest. The three AMG8833 evaluation boards were connected to a host laptop via universal asynchronous receiver-transmitter (UART) interfaces. The evaluation boards were synchronised with a refresh rate of 10 FPS. All sensor recordings were saved in a \*.csv format. The 10 FPS was chosen for: *i*) keeping update rate for capturing the detailed states transition during the activity cycle; *ii*) reserve enough time for sensor reacting to the changes in temperature and obtain accurate thermal reading. The configuration of the AMG8833 evaluation boards and controlling of the data streams were coordinated by National Instruments LabVIEW®.

Two layouts of sensors positioning were used in this dataset: 1) small layout; 2) large layout. In small layout, the three sensors were positioned equally 1.5 metres from experiment area, designated for performing the activities. Regarding large layout, the distance of the sensors was increased to 2.5 metres. All three sensors were elevated approximately 1 metre from the ground for both layouts. Both layouts ensured that the subject was in coverage of sensing system. The exact sensors positioning can be observed in Figure 3.1a, while Figure 3.1b depicts the experiment venue with the three sensors and a subject. Regarding the thermal concerns, the pixels of the IR array correspond to the measured temperature. In the beginning of the experiment, the temperature was measured at 18° centigrade. The small layout experiment data was collected during these conditions. Later, the temperature gradually rose to 21° centigrade, which marked the data collection in the large layout scenario. The participants were wearing normal early spring clothes to be compliant with a real world scenario.



**Figure 3.1:** Layouts and experiment scenario. (a), Small and large experiment layouts; (b), The real experiment scenario.

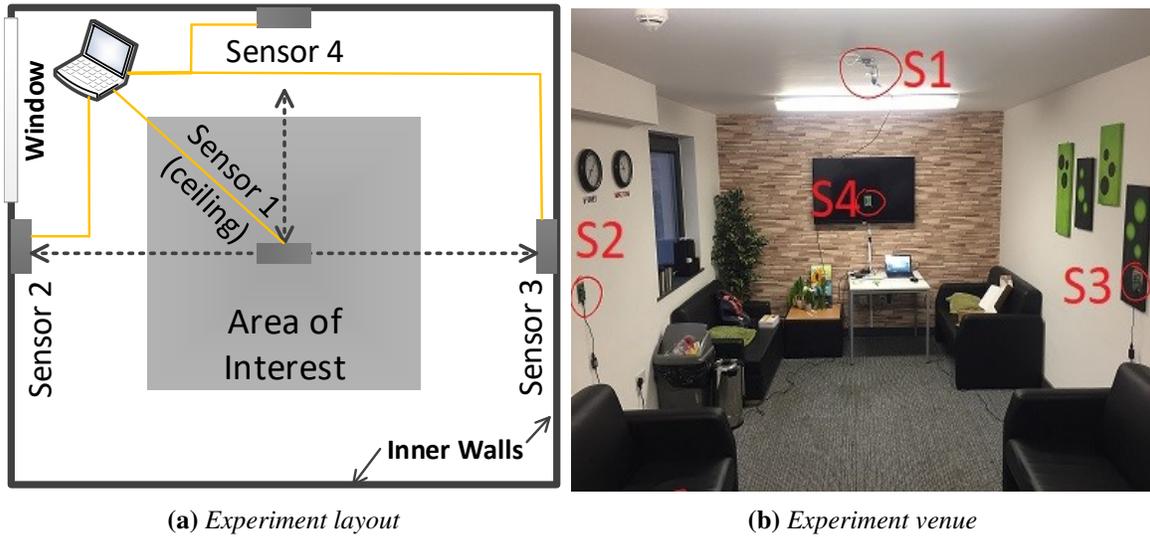
The dataset is composed with activities distinguished in two categories: a) single-subject activities; b) double-subject activities. An important note is that double-subject activities were collected only in large layout. On the other hand, single-subject activities were collected in both layouts. Three participants (one female and two males) with diverse body sizes contributed for capturing the activities. Each single-subject activity was repeated 10 times by each participant in small layout (AS1~AS8) and large layout (AL1~AL8). Regarding double-subject activities, the three possible double combinations of three subjects similarly performed 10 repetitions of each activity in the large layout (B1~B7). This leads to 480 single-subject activity outputs (240 outputs in small layout; 240 outputs in large layout) and 210 outputs of double-subject activities as observed in Table 3.1.

**Table 3.1:** Two categories of activities in Coventry-2018

Activities	1-subject		Activities	2-subject
	Small	Large		Large
Sit Down	AS1	AL1	Both Sitting	B1
Stand Still	AS2	AL2	Sitting & Moving	B2
Sit Down; Stand-Up	AS3	AL3	Sitting & Standing	B3
Stand Up	AS4	AL4	Random Moving	B4
Move Left, Right	AS5	AL5	Both Standing	B5
Move For, Backward	AS6	AL6	Standing & Moving	B6
Walking Diagonally 1	AS7	AL7	Walking Across	B7
Walking Diagonally 2	AS8	AL8		

### 3.2.2 Infrared Human Activity Recognition Dataset *Infra-ADL2018*

*Infra-ADL2018* dataset was collected in Bristol Robotic Lab (BRL), University of West England, United Kingdom in 2018. Similarly to *Coventry-2018*, Grid-EYE Panasonic® sensors were used for conducting the experiment. There exist four key differences of *Infra-ADL2018* compared to *Coventry-2018*: *i*, a ceiling-mounted sensor is deployed in addition to the two side sensors and one front sensor; *ii*, besides single-subject activities and double-subject activities, triple-subject activities are presented; *iii*, nine participants contributed for dataset collection; *iv*, only one layout exists in the dataset. The experiment layout and experiment venue are observed in Figure 3.2a and Figure 3.2b respectively.



(a) Experiment layout (b) Experiment venue  
**Figure 3.2:** Layout and experiment scenario. (a), Experiment layout; (b), The real experiment scenario.

Table 3.2 depicts the name of activities designated in three categories. The nine single-subject activities were repeated each three times from nine subjects (243 outputs per sensor). In regard to double-subject activities, there exist 10 activities, which were repeated three times by each of eight combinations (240 outputs per sensor). Finally, the two triple-subject activities were repeated similarly three times by three combinations (18 outputs per sensor).

**Table 3.2:** Three categories of activities of *Infra-ADL2018*

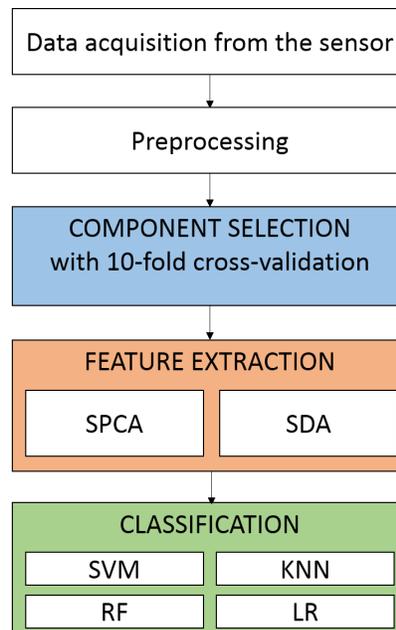
<b>1-subject</b>	Walking LR; Walking RL; Walking Away; Walking Toward; Falling; Stand to Sit; Sit to Stand; Sitting Still; Standing Still
<b>2-subject</b>	Walking Opp Direction; Walking Same Direction; Sitting; Standing; Sitting+Walking Front; Sitting+Walking Behind; Standing+Walking Front; Standing+Walking Behind; Sitting+Standing; Falling+Walking
<b>3-subject</b>	Free Movement; Stand Still

### 3.3 Data Analysis Based on Sparse Techniques

The data derived from the sensors has extremely high-dimensionality of  $40 \times 64 = 2560$ , where 40 are the number of frames and 64 refers to vectorised  $8 \times 8$  2D IR images. Considering the limited number of samples, this leads to an ill-posed problem and overfitting. The current section outlines the analysis of using sparse methods for supervised and unsupervised feature extraction of high-dimensional data from *Coventry-2018*. Only single-subject data from this dataset are used. The sparse methods are compared with their non-sparse equivalents in terms of recognition accuracy as well as additive noise tolerance.

#### 3.3.1 Methodology

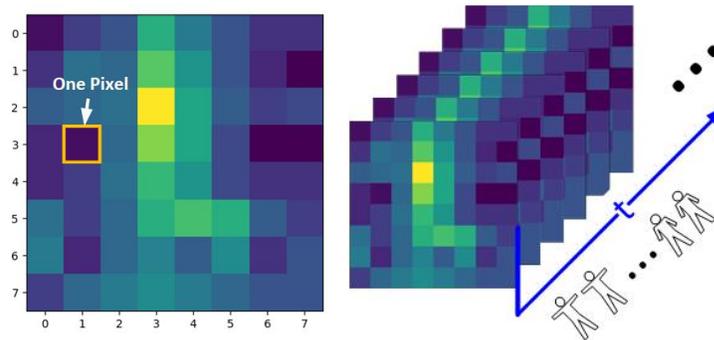
Figure 3.3 describes the steps for the data analysis based on sparse techniques. Once the data are collected and pre-processed, component selection is achieved using 10-fold CV. Feature extraction is implemented using SPCA and SDA, described later. Therefore, the extracted sparse features are classified with four off-the-shelf methods: SVM, k-NN, RF, and LR.



**Figure 3.3:** Methodology steps for data analysis based on sparse methods.

##### 3.3.1.1 Data Acquisition

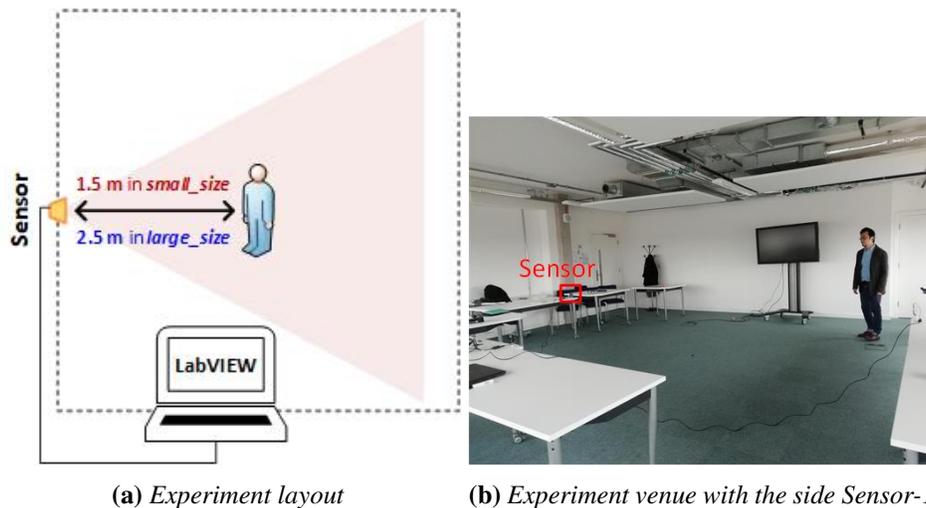
A subset of the *Coventry-2018* dataset based on one of the three sensors is used for the sparse techniques analysis. The LRIR sensor employed in this study is used for *Coventry-2018* dataset. As explained earlier, the Panasonic Grid-EYE (AMG8832) sensor was used, where each of the  $8 \times 8$  pixels represents temperature value derived from the environment. A single frame is visualised in Figure 3.4a for the activity sitting. A 10 FPS is selected observed in Figure 3.4b.



(a) An  $8 \times 8$  thermal pixel matrix      (b) A stream of different frames

**Figure 3.4:** Example of (a) a single frame of an IR image and (b) a stream of a number of frames.

Considering the fact that only a subset of *Coventry-2018* dataset was used, the large layout is selected as observed in Figure 3.5a. The experimental area is positioned 2.5 metres from the only used side Sensor-1. The sensor was positioned approximately 1 metre from the ground. The actual experimental venue with one of the participants are shown in Figure 3.5b.



(a) Experiment layout

(b) Experiment venue with the side Sensor-1

**Figure 3.5:** Layout and experiment scenario for Sensor-1. (a), Experiment layout; (b), The real experiment scenario.

As explained earlier, the IR sensors detect the temperature of the environment including ambience, subjects, and different heat sources. Considering subjects, the normal human's temperature is approximately  $36^{\circ}\text{C}$ , which is usually considerably higher than the ambience in a room. The dataset *Coventry-2018* was collected in March 2018, where a temperature in the range of  $18^{\circ}\text{C}$ - $21^{\circ}\text{C}$  was observed. The three participants who volunteered for creating the dataset were all wearing normal daily clothes, which provides realism to the scenario.

### 3.3.1.2 Data Preparation

For the sparse techniques analysis, the eight (8) single-subject activities of the *Coventry-2018* dataset from all three participants are used. They are listed in Table 3.3.

In this work, each of the single-subject activities is recorded separately. While segmentation techniques are possible for separating the activities in continuous streams, this is not the case for the current work. In addition, each activity is performed 10 times by each participant considering the fact that the participant is positioned 2.5 metres from the area of interest in the case of large layout.

The *Coventry-2018* dataset was described earlier in Section 3.2.1. Considering the current data analysis with sparse methods, only single-subject activities are used captured by the side Sensor-1. More specifically, eight single-subject activities exist, which were repeated 10 times by each of the three participants. This leads to a total of 240 samples. Obviously, different activities require different amounts of time for execution. Therefore, it was computed that the average duration among all activities was 4 seconds or 40 frames considering the 10 FPS. Interpolation and extrapolation methods are used for equalisation, which are described later. This leads to a cuboid of shape  $240 \times 40 \times 64$ , where 64 refers to the  $8 \times 8$  pixel matrix. For data analysis, the cuboid shapes are transformed to a 2-dimensional space of  $240 \times 2560$ .

**Table 3.3:** *Investigated Activities*

	Activities		Activities
A1	Sit Down	A5	Move Left, Right
A2	Stand Still	A6	Move Forward, Backward
A3	Sit Down; Stand Up	A7	Walking Diagonally 1
A4	Stand Up	A8	Walking Diagonally 2

Obviously, the extremely high-dimensionality of the data compared to the limited number of samples can lead to overfitting. To address the issue, sparse methods are applied for feature extraction, which are later compared to their non-sparse equivalents. Sparse methods are known to identify unimportant features and set them to zero, which is their main advantage. Hence, the results can provide means for analysis and discussion.

## 3.3.2 Data Analysis

### 3.3.2.1 Sparse Feature Transformation Techniques

**3.3.2.1.1 Sparse Principal Components Analysis (SPCA)** PCA is one of the leading dimensionality reduction methods in the literature. The method identifies the directions, also referred to as principal components, capturing the most variance. As such, a limited number of principal components can be selected explaining more than 90% of the data variance [150].

The non-sparse PCA has a main disadvantage, where each principal component uses a linear combination of all features. As such, it is difficult to identify problematic and insignificant features with low or none contribution for explaining the data. To address the issue, SPCA is a regression-type problem, which uses a quadratic penalty as seen in equation (3.1). In its most common implementation, the sparse method uses elastic net by imposing L1 (Manhattan distance) and L2 (Euclidean distance) penalties. The used penalties set unimportant features to zero by leaving only features with contribution to data variation with non-zero values. As such, SPCA provides interpretation and analysis of the data features.

Let  $X = [x_1, \dots, x_n]$ , and  $x \in \mathbb{R}^m \sim p(x)$  be a random variable, where  $p(x)$  is a distribution. Initially,  $A = V[v_1, \dots, v_K]$  as the first  $K$  loadings of ordinary PCs. In first step, the first  $k$  PCs  $A = [\alpha_1, \dots, \alpha_K]$  are considered fixed and the following elastic net objective is solved for  $\beta_j, j = 1, \dots, K$ :

$$\beta_j = \arg \min_{\beta} (\alpha_j - \beta)^T X X^T (\alpha_j - \beta) + \lambda \|\beta\|^2 + \lambda_{1,j} \|\beta\|_1 \quad (3.1)$$

Then in the second step,  $B = [\beta_1, \dots, \beta_K]$  are considered fixed and the SVD of  $X^T X B = U D V^T$  is computed and  $A = U V^T$  is updated. The first and second steps are iterated until convergence. Finally, the eigenvectors are normalized as ordinary  $\hat{v}_j = \frac{\beta_j}{\|\beta_j\|_j}, j = 1, \dots, K$ . Further information in this case can be found in [151].

**3.3.2.1.2 Sparse Discriminant Analysis (SDA)** Linear Discriminant Analysis (LDA) is a widely used method as both a Gaussian classifier and a supervised dimensionality reduction method. The dimensionality reduction method known as Reduced Rank LDA (RRLDA) assumes that the  $Q$  classes are positioned in an a space with dimension  $Q - 1$ . The assumption of Gaussian distribution allows a limited number of components to be selected as specified. Considering this, at most  $Q - 1$  discriminant features can be extracted, which is a significant drop in the dimensionality. The RRLDA method computes the within-class covariance matrix  $W$  along with the between-class covariance matrix  $B$ . The Fisher's *canonical variables* or *discriminant coordinates* allow the transformation of data to a lower dimensional space [73]. More specifically, the Fisher's aims to maximise the Rayleigh quotient given by:

$$\max_a \frac{a^T B a}{a^T W a} \quad (3.2)$$

where  $a$  is *canonical variable* or *discriminant coordinate*. The method focuses on increasing the distances between the classes means (centroids) for improving separability.

In cases with  $n \ll p$  condition as in this study, where  $n$  refers to the number of samples, and  $p$  denotes the number of features, LDA cannot be applied directly. This is due to the fact that the within-class covariance matrix  $W$  of the features is singular. To address the issue, sparsity is implemented using the L1 (Manhattan distance) norm used in the lasso objective function. More specifically, the method calculates the sparse transformations  $\beta_q, q = 1, 2, \dots, Q - 1$  to transform the original data  $X$  to a new low dimensional space

$X\beta_{1:m}$ ,  $m \leq Q - 1$ . In addition, the  $1 : m$  notation refers to the selection of the first  $m$  vectors. In this study, each  $\beta_q$  vector length is 2560, which corresponds to the total number of raw features. The objective function uses the L1 norm penalty as shown below:

$$\min_{\beta_q} \left\{ \frac{1}{n} \|Y\theta_q - X\beta_q\|^2 + \gamma\beta_q^T\Omega\beta_q + \lambda\|\beta_q\|_1 \right\} \quad (3.3)$$

where  $Y = n \times Q$  presents a matrix of dummy variables for the  $Q$  classes, while  $Y_{iq}$  serves as an indicator showing whether the  $i^{th}$  observation belongs to the  $q^{th}$  class. In addition,  $\theta_q$  denotes a  $Q$ -vector of scores. The original data matrix is transformed to a lower dimensional space  $Z = X\beta_{1:m}$ ,  $m \leq Q - 1$  using the sparse matrix. To address the aforementioned issue with singularity in the within-class covariance matrix  $W$ ,  $\Omega$  is defined as a positive definite matrix. In addition,  $\lambda$  and  $\gamma$  are non-negative tuning parameters. For further information, the readers are referred to [152].

### 3.3.2.2 Classification Methods for Activity Detection

In this work, four traditional classification methods are used - SVM, RF, k-NN, and LR. All four techniques are supervised and have shown suitability for multi-class classification problems.

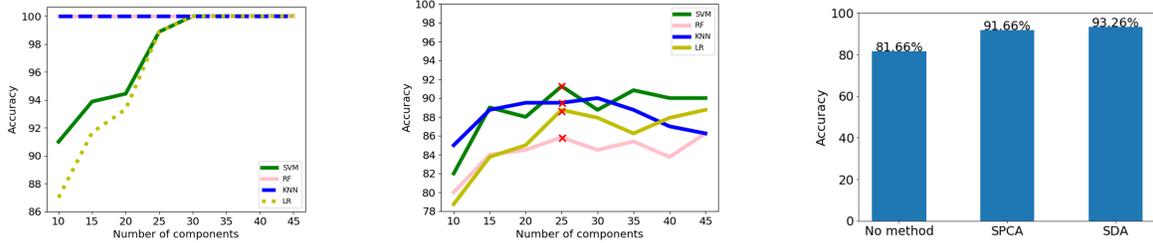
SVM uses hyperplanes to divide the input space by providing separation between the classes [153]. In cases where the classes are non-linearly separable, a number of kernels can be adapted to these problems. The second considered method RF represents an ensemble of decision trees used for making predictions [79]. Considering the resulting predictions of each decision tree for classification, the majority output class is selected. Next, k-NN adapts a simple concept of computing distances between the data samples [154]. Euclidean distances are used in this work. When a new data sample is considered, it is assigned to the class belonging to its nearest neighbour or majority of nearest neighbours. The nearest neighbour is based on the calculated nearest Euclidean distance. Finally, LR models the probabilities of a data sample to belong to a specific class [73]. The classification technique ensures to keep to sum-to-one constraint of the probabilities and their range of  $[0, 1]$ . While LR was initially used for binary problems only, it has been later adapted to multi-class problems.

To ensure fair distribution of data samples in both training set and testing set, the data was randomly divided. In addition, 75% were used for training, while the remaining 25% were designated for testing. This leads to  $180 \times 2560$  training set, and  $60 \times 2560$  testing set. To select a number of principal components for SPCA, 10-fold CV was applied. This resulted in selected 25 principal components. In terms of SDA, the maximum possible number of components  $Q - 1$  is selected, which is 7 in this case. Therefore, the described above four classification methods are used: SVM, RF, k-NN, and LR. Subsequently, a scenario with additive positive Gaussian noise is implemented, which resembles the overall increase in the room temperature in a real scenario. Since sparse methods are known to set certain features to zero, they are expected to perform better compared to their non-sparse equivalents.

### 3.3.2.3 Sparse Dimension Reduction and Model Selection

As described above, 10-fold CV is used for selecting the optimum number of PCs  $\hat{c}$ . A candidate set of number of sparse principal components is [10, 15, 20, 25, 30, 35, 40, 45]. Training and validation accuracy maps are computed shown in Figure 3.6a and Figure 3.6b respectively. Both accuracy maps are based on the four classification methods described earlier. As observed, 25 SPCA components are selected since the prominent peaks refer to 25 components in the validation map. Therefore, the transformation to the reduced space is performed based on  $Z_{tr} = X_{tr}V_{1:\hat{c}}$ ,  $\hat{c} = 25$ . The  $1 : \hat{c}$  notation shows the first 25 eigenvectors. Considering the original 2560 features, the selection of only 25 SPCA components represents a significant drop in the dimensionality as observed in Figure 3.7.

In terms of SDA, a sparse  $\beta_{2560 \times 7}$  is obtained, where  $Q - 1 = 7$ , and used for transforming data  $Z_{tr} = X_{tr}\beta$ .

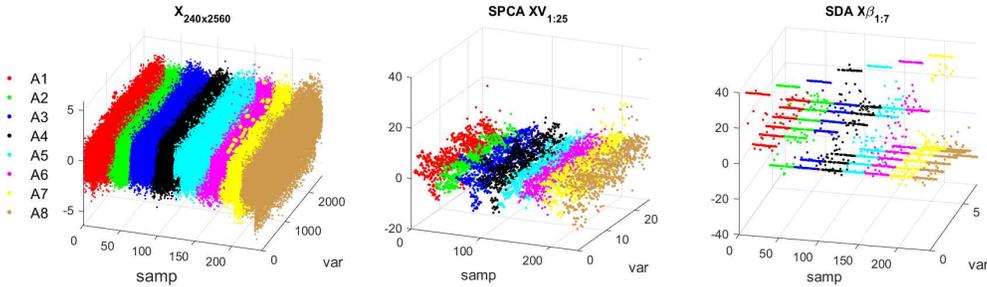


(a) Average of training accuracy over 10 folds of CV loop, used for SPCA component selection

(b) Average of validation accuracy over 10 folds of CV loop, used for SPCA component selection

(c) Comparison of SVM accuracy using original features, transformed features based on SPCA and SDA

**Figure 3.6:** Accuracy of SVM, RF, KNN and LR on both training and testing sets for SPCA and SDA.



**Figure 3.7:** Visualisation of the original data matrix and the transformed samples using SPCA and SDA.

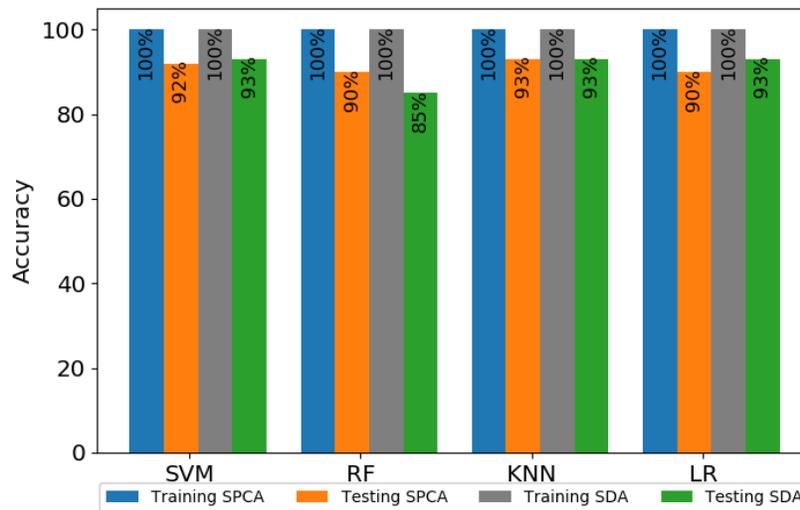
### 3.3.2.4 Classification Analysis

The effectiveness of the sparse dimensionality reduction methods is further analysed. This is achieved by visualising the eight (8) different classes in their original dimension as well as their reduced representations with SPCA and SDA as observed Figure 3.7. As illustrated, the general variations of the data samples as well as the variations between different classes

are increased, which is a positive indicator for improved classification. To prove this, Figure 3.6c shows the classification accuracy using the raw features compared with the sparse transformations with SPCA and SDA. As expected, the classification accuracy is improved significantly.

### 3.3.3 Experimental Results

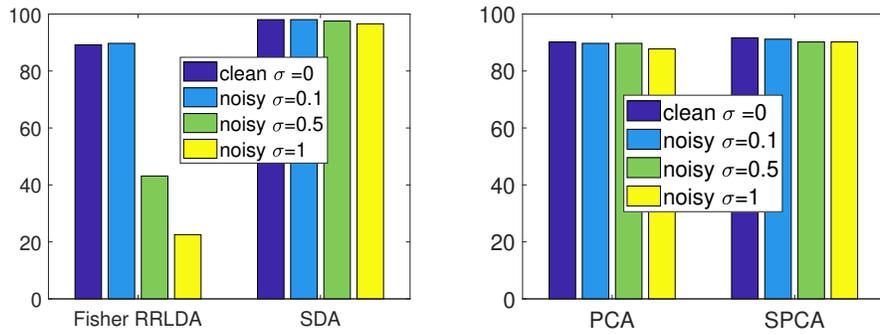
The four classification methods have been applied to both the training set and the testing sets. Figure 3.8 reveals the accuracies with all combinations of feature extraction strategies and classification methods. The training accuracy is exactly 100% for all combinations. In terms of testing accuracy, SDA outperforms SPCA with the exception of RF. This is attributed to the fact that SDA is a supervised method. To conclude, these results are only based on the large layout. In order to evaluate the effect of the subject's distance from the sensors, the comparison experiment discussed in Section 3.4.2.5 demonstrates the results based on the two layouts in *Coventry-2018* dataset. Model generalisation in terms of layout in two scenarios is further discussed in Section 3.4.2.6.



**Figure 3.8:** Accuracy of SVM, RF, KNN and LR on both training and testing sets for SPCA and SDA.

#### 3.3.3.1 Noise Analysis

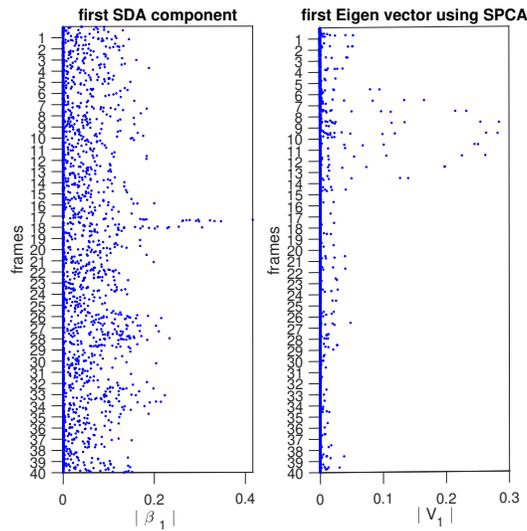
As discussed above, sparse methods are known to set certain features to zero, which can potentially lead to a higher tolerance of noise. To explore this, randomly distributed positive Gaussian noise with  $\mu = 0$  and  $\sigma = 0.1, 0.5, 1$  is added to the testing data only and the classification performance for the testing data is computed. The sparse methods are compared with their non-sparse equivalent techniques - PCA and RRLDA. The results in Figure 3.9 are expected as the sparse methods show higher tolerance to noise.



**Figure 3.9:** Comparison of the clean and noisy test data classification performances for RRLDA and SDA as well as PCA and SPCA.

### 3.3.3.2 Visualisation of Sparse Components

In terms of the sparse methods’ ability to set unimportant features to zero, it is important to further analyse this property. Therefore, the absolute value of the first SDA component  $\beta_1$  and the first SPCA eigenvector  $V_1$  are visualised in Figure 3.10.

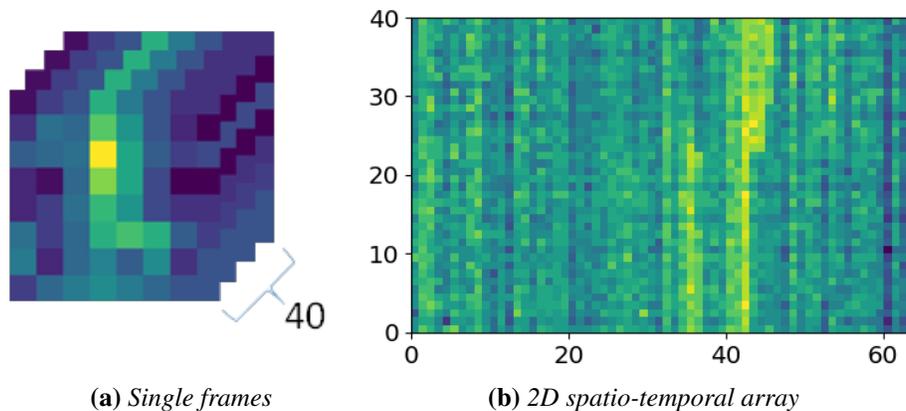


**Figure 3.10:** Visualisation of the first sparse component of SDA  $\beta_1$  and first sparse eigenvector of SPCA  $V_1$ .

As observed, the largest non-zero coefficients over 40 frames of activity are shown to be distributed in the central frames although not as strongly. This may match to the fact that the participants are less active in the beginning and ending of an activity. However, the first and last frames also show contribution to the activity variations.

### 3.4 Noise Analysis and Classification Modeling for Holistic Understanding of LRIR Data

This section considers the entire *Coventry-2018* dataset for a number of experiments as well as *Infra-ADL2018* dataset for techniques verification and analysis. More precisely, first data pre-processing with frame equalisation algorithms are described in detail. Second, noise reduction from the 2D spatio-temporal maps is considered. Figure 3.11 shows the formation of 2D spatio-temporal maps from the individual  $8 \times 8$  frames. Next, traditional machine learning architectures are applied for recognition performance as well as a deep NN architecture. Furthermore, small layout and large layout data are compared to test the models' sensitivity towards the sensors' distances. Similarly, single-subject activities and double-subject activities are then compared. Moreover, model generalisation is achieved, where the models are trained on the small layout data and tested on the large layout data, and vice versa. Finally, model generalisation in terms of unseen subjects is evaluated in a leave-one-subject-out fashion, where the models are tested on data from subjects not used for training.

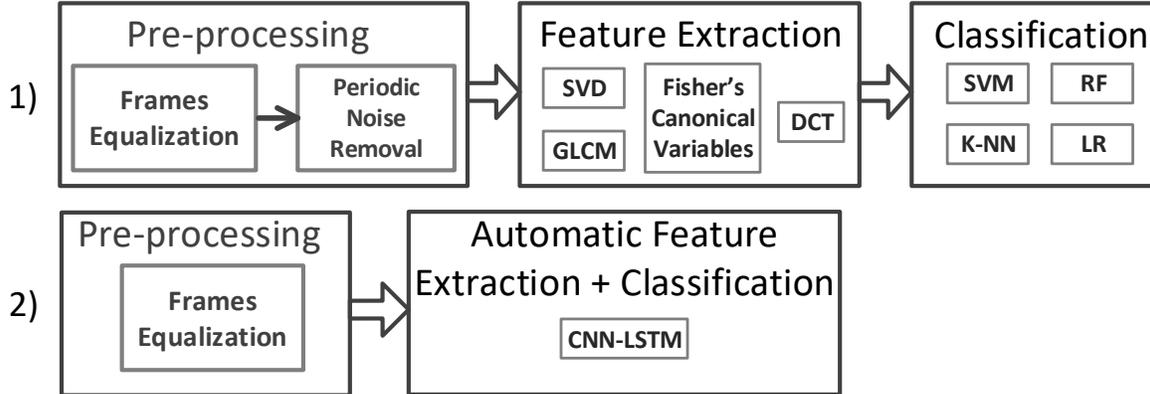


**Figure 3.11:** (a) Examples of  $8 \times 8$  pixel frames for the activity sitting, and (b) the vectorised spatio-temporal 2D array over 40 frames and 64 pixels.

#### 3.4.1 Data Analysis Framework

This section focuses on the feature representation of the LRIR data. Since the acting duration is not equal for different activities and participants, the durations of records in both *Coventry-2018* and *Infra-ADL2018* show diversity. As described in Section 3.3.1.2, the records are equalised before passed for feature extraction for both machine learning and deep learning architectures. The 2D spatio-temporal maps suffer from periodic noise as shown in Figure 3.11. Therefore, a periodic noise removal algorithm is introduced to alleviate the effect of noise in the developed 2D spatio-temporal arrays after frame equalisation. This improved the activity detection. It is important to note that the periodic noise removal method is applied only to the machine learning architectures. In regards to the deep learning architecture, the  $40 \times 8 \times 8$  inputs are used instead of 2D spatio-temporal arrays. Hence, periodic noise

removal is not needed for this architecture. The above mentioned two operations are called pre-processing. The overall data flow for the two categories of architectures can be generally described by Fig. 3.12.



**Figure 3.12:** The data processing flow in this work, which shows the two undertaken strategies.

### 3.4.1.1 Data Pre-Processing

**3.4.1.1.1 Frames Equalisation and Vectorisation** In order to implement the feature extraction and training processes, the records are equalised to 4 seconds (equivalent to 40 frames), which is the average duration of the all the activities in *Coventry-2018* dataset. Appropriate methods based on interpolation and extrapolation are proposed in this case. For the records longer than 4 seconds, extrapolation is used by removing frames at regular intervals. In the case of records shorter than 4 seconds, a new frame is added between already existing two frames. The newly added frame is an average of the two. The proposed methods for interpolation and extrapolation are described in Algorithm 1 and Algorithm 2 respectively.

The main idea for interpolation algorithm is to randomly select frame indices from the existing sequence of frames and order the indices in ascending order. Then for each index, the average of the corresponding frame and the following frame is computed to generate a new frame which is placed between them in the new sequence. This strategy is compared with random selection and insertion of the frames within the streams. The proposed strategy shows better accuracy compared to the random repetition of some frames in the sequence.

Regarding the extrapolation method, the main idea is to delete frames from equal intervals. As shown in Algorithm 2, a step size is used to remove some frames. Depending on the length of the frames  $N$ , two different step sizes are calculated; if  $N$  is large so that, the floored result of  $step_1 = \frac{N}{N-d}$  is one, an empty array  $C$  is formed first. The array  $C$  is used to keep the frame indices that should be removed. Then, series of frame deletion indexes are computed via two nested loops. The outer loop variable  $i$ , counts all frame numbers, skipping only the first one. The inner loop variable  $j$  starts from 0 up to  $step_2 - 1$ , that generates

small values to be added to  $i$ , in order to skip some frames. The summation of  $i + j$  generates the indexes for deletion. The iteration within the nested loops will stop if the required number of deletion indexes  $N - d$  is achieved. Otherwise, if  $N$  is not large compared to the desired number of frames  $d = 40$  so that  $step_1 > 1$ , the deletion indexes are found simply based on one loop. The frame index selection equation  $z = i \times step_1 - i$  is used in this case. Since both  $i$  and  $step_1$  are integers,  $z$  can never become a non-integer value.

---

**Algorithm 1: Interpolation algorithm for frames equalisation**


---

**Input:** Set of the acquired  $8 \times 8$  IR frames  $A$ ; Desired frames number  $d = 40$

**Algorithm:**

1. Compute  $N = length(A)$ .
2. Initialise an empty array  $C$ .
3. Initialise  $C = A$ .
4. Select  $d - N$  number of random values (without replacement) in range  $0, 1, \dots, N - 2$  and sort them in ascending order in a set ( $rs$ ).
5. Set index counter  $k = 1$ .

- (1) **For**  $i=0, \dots, d - N - 1$ 
  - $C[rs(i) + k] = avg(A[rs(i)], A[rs(i) + 1])$
  - $k = k + 1$

**End Loop 1**

**Output:** Interpolated array  $C$ .

---



---

**Algorithm 2: Extrapolation algorithm for frames equalisation**


---

**Input:** Set of the acquired  $8 \times 8$  IR frames  $A$ ; Desired frames number  $d = 40$

**Algorithm:**

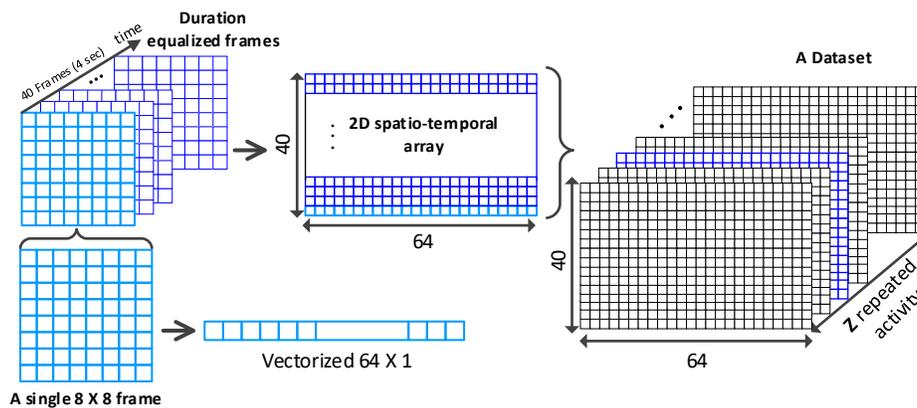
1. Initialise  $N = length(A)$
  2. Compute  $step_1 = int\left(\frac{N}{N-d}\right)$ ;  $step_2 = int\left(\frac{N}{d}\right)$ .
    - If**  $step_1 = 1$ :
      - Initialise an empty array  $C$
      - (1) **For**  $i=1, \dots, N-1$  **with step size**  $step_2 + 1$ 
        - (2) **For**  $j=0, \dots, step_2 - 1$ 
          - if**  $length(C) < N - d$ :
            - Append frame number  $i + j$  to  $C$
      - End Loop 2,1**
      - Sort  $C$  in descending order
      - (3) **For**  $i$  in  $C$ 
        - Delete a frame with index  $i$  from the array  $A$
      - End Loop 3**
    - End Loop 1**
    - Else:**
      - (1) **For**  $i=0, \dots, N - d - 1$ 
        - Delete a frame with index  $z = i \times step_1 - i$  from array  $A$
      - End Loop 1**
- Output:** Extrapolated array  $A$ .
- 

In later stages of this PhD, it was found that the above extrapolation algorithm has a lim-

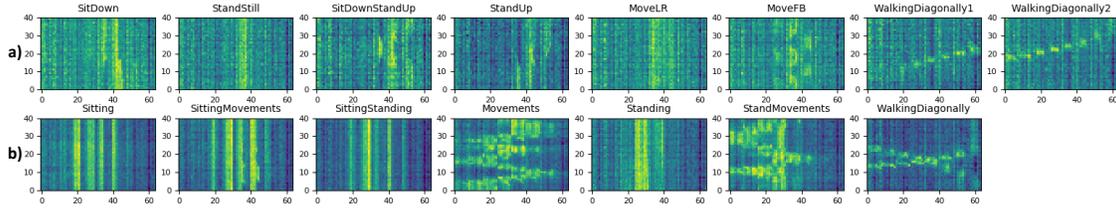
itation for some frame numbers. More specifically, in some cases when the criterion inside the two nested loops stops the index generation, it was observed that most of the selected frames for removal were from the earlier part of the stream and some last indexes were remained. This means this selection strategy has favoured earlier frames for deletion for some frame streams. Then, it is recommended to perform the frame removal based on analysis of frames data variation so that, some frames from the smooth sequence of frames with less variation to be selected. Despite of the described limitation, comparison of the used strategy with random selection and exclusion of the frames, showed some improvements. The proposed strategy worked better based on the achieved results as described later in Section 3.4.2.1. The same equalisation strategy is applied on *Infra-ADL2018* dataset.

After frame equalisation, the algorithms yield 40 frames per activity with  $8 \times 8$  temperature pixels. Considering the low resolution of IR images and low level of correlation between local pixels, the individual pixels can be considered as independent variables. Therefore, each  $8 \times 8$  frame is vectorised. This is performed for all frames of each activity. Thus, the equalised 4-second records of each activity, are re-formed into a  $40 \times 64$  2D spatio-temporal array as observed earlier in Fig. 3.11 in Section 3.4. This 2D array can be treated as an image. In the analyses performed in this study, the 2D spatio-temporal arrays are formed for both *Coventry-2018* and *Infra-ADL2018* datasets.

In the case of *Coventry-2018* dataset, the equalised and vectorised data consist of a cuboid of size  $240 \times 40 \times 64$  for single-subject activities in small layout and large layout, and a cuboid of size  $210 \times 40 \times 64$  for double-subject activities in large layout. This is illustrated in Fig. 3.13. To simplify the writing and avoid confusion, the cuboid size is considered as  $\mathbf{Z} \times 40 \times 64$ .  $\mathbf{Z} = 240$  is used for the single-subject data (in both small and large layout) and  $\mathbf{Z} = 210$  for the double-subject data (only in large layout). The example visualisation of the equalised and vectorised LRIR data are shown in Fig. 3.14. They are the 2D spatio-temporal maps used for activity recognition in this Chapter.



**Figure 3.13:** Data vectorisation and dimension description. One sample 2D spatio-temporal array is visualised in blue color in the middle.



**Figure 3.14:** The (a) one-subject and (b) two-subject vectorised  $8 \times 8$  frames of activities (64) over 40 equalised frames.

**3.4.1.1.2 The Proposed Noise Removal Algorithm** In order to improve the activity classification accuracy, having high quality data is important. The equalised  $40 \times 64$  2D spatio-temporal maps formed for each activity show periodic noise effect in both horizontal and vertical direction. This can be observed in Fig. 3.18a. The periodic noise includes both horizontal and vertical stripes. The reason can be due to the low resolution of the frames that cannot encode small local variations. Then, the same value is assigned for adjacent pixels with small level of variations. For example, the horizontal noise stripes are formed due to a similarity of vectorised sequential frames. That can occur for example when an activity includes slow movements or still condition. On the other hand, the vertical noise stripes correspond to some unchanged pixels over all 40 frames, such as pixels that have never been occluded by the subjects. As it can be seen, these effects are prominent in the arrays and the classification tests demonstrated that they can negatively affect the classification accuracy in both training and validation stages.

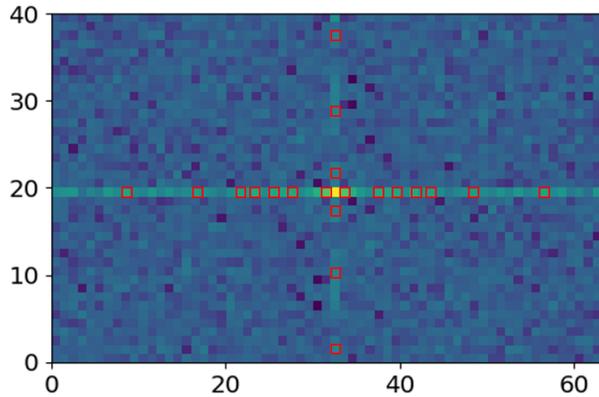
In order to alleviate the effect of the noise, 2D Discrete Fourier Transform (DFT) is used. It is a digitized version of the Fourier Transform, so that, it contains sampled frequencies of image rather than all continuous frequency information. The 2D DFT of an image array is a 2D array of the same size. Each pixel in this 2D array corresponds to the coefficient in a specific frequency. They show the main spectral information of the image. The equation for 2D DFT is shown in (3.4).

$$F[u, v] = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X[m, n] e^{-j2\pi(u\frac{m}{M} + v\frac{n}{N})} \quad (3.4)$$

In equation (3.4),  $M$  and  $N$  show the number of rows and columns of  $X$  and  $X[m, n]$  is one pixel in the original image, which is the 2D spatio-temporal map in this work.  $F(u, v)$  is the element in row  $u$  and column  $v$  of the output array  $F$ . This 2D DFT equation computes each element,  $F(u, v)$ , in spectral domain by multiplying the spatial image  $X$  with the corresponding sinusoidal and cosinusoidal base functions and summing up the result. For example,  $F(0,0)$  describes the information in zero frequency called DC. If  $u$  and  $v$  are set to zero in equation (3.4), the DC component is calculated as the average of the gray levels in the 2D spatio-temporal map. On the other hand,  $F(M-1, N-1)$  is the highest frequency coefficient. In this work, 2D DFT is calculated for each  $40 \times 64$  2D spatio-temporal map, which

is illustrated in Figure 3.13. Though,  $F(u,v)$  is a complex value matrix, the amplitude of the complex coefficients are considered for further processing. Due to the large dynamic range of the DFT coefficients, the logarithm function is applied, yielding a 2D power spectrum.

Next, the resulting power spectrum in FFT domain is visualised, as shown in Fig. 3.15. Due to the Hermetian symmetry property of DFT, the elements in  $F[m, n] = F(M - 1 - m, N - 1 - n)$  so that, there is symmetry about  $(\frac{M-1}{2}, \frac{N-1}{2})$ , the central pixel in the 2D DFT array. For more information, the readers are referred to [155]. The coefficients in the central horizontal and vertical lines in this 2D spectrum are considered for periodic noise removal, as previously used in other works [156, 157, 158]. As expected, there exist noticeable peak pixel values in both directions of stripes, that indicate periodic effect of noise. The peaks are also symmetric with respect to the central point of the 2D DFT array. The horizontal periodic noise corresponds to the peaks in vertical stripes coefficients and the vertical noise is connected to the peaks in horizontal stripe as shown in Figure 3.15.



**Figure 3.15:** The power spectrum with prominent peaks rounded in red.

In order to reduce the noise, a threshold needs to be defined to identify the peaks and change their values to a lower level over the central stripes. The DC component is not considered as it represents zero frequency. Two early issues about the choice of threshold value and changing the peak values are addressed.

The first issue is the strategy about considering the threshold value for each horizontal and vertical direction. The two possible choices to consider are *i*) a unique constant threshold value for each direction based on the overall mean of the horizontal and vertical stripes over all 2D power spectrum arrays or *ii*) a random variable based on the individual power spectrum arrays statistics for the threshold value. For the first choice, the mean values over horizontal and vertical stripes (excluding the DC component)  $\mu_h$  and  $\mu_v$  for all the three sensors (of *Coventry-2018* dataset) are calculated. Under the null hypothesis that all power spectral arrays have similar mean values (over the stripes) for each of the horizontal and vertical directions, the one-way analysis of variance (ANOVA) test is performed described in detail in Section 3.4.2.2.1. The p-value rejects the null hypothesis, and therefore, general

thresholds (for the two directions) based on the stripes mean values cannot be used for all data. Therefore, the threshold values are customised for each 2D power spectrum array over the two directions.

The second initial issue is about the reduction strategy to replace the peak values. The two values considered are 0 and statistics based on mean and standard deviations  $\mu_h \pm \sigma_h$  and  $\mu_v \pm \sigma_v$ , where the two latter showed better accuracy. Therefore, for the peaks replacement strategy, the statistical approach using the average stripe values is considered.

After addressing these two initial issues, a supervised strategy was developed for finding the best parameters for periodic noise removal. For this aim, the threshold parameters  $T_h$  and  $T_v$  for each horizontal and vertical direction, are formulated based on the mean and the standard deviation of all the coefficients in the corresponding stripe as follows:

$$T_h = \mu_h + i_1 * 0.5 * \sigma_h; T_v = \mu_v + i_1 * 0.5 * \sigma_v \quad (3.5)$$

where  $i_1$  is an integer value within the range  $\{-2, 8\}$  that is learned based on a supervised strategy. In addition, as there exists a number of pixel pairs greater than the threshold, it needs to be decided how many of them should be reduced to achieve the highest accuracy. The number of pairs of pixels to be changed depends also on the threshold value. In other words, the smaller the threshold, the bigger the number of pixel pairs violating the threshold. Overall, the parameters,  $T_h$  and  $T_v$  and number of pairs of pixels  $Num_h$  and  $Num_v$ , that influence the quality of noise removal, are found based on a supervised learning strategy in order to achieve optimum accuracy.

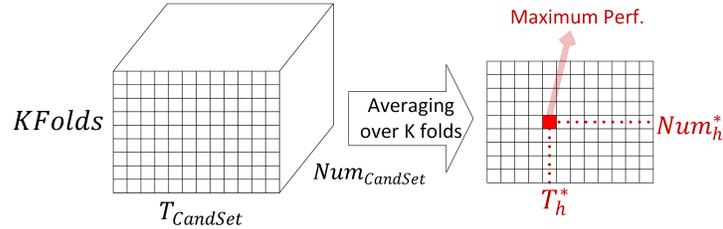
The supervised learning strategy is based on the K-fold CV model selection technique. So that, 10-fold CV is used to find the optimum parameters  $T_h^*/T_v^*$  and  $Num_h^*/Num_v^*$ , for each of the horizontal and vertical directions. For this aim, candidate sets of values are formed for these parameters  $T_{CandSet}$  and  $Num_{CandSet}$  first. The  $T_{CandSet}$  is formed based on the variable  $i_1$  as shown in equation (3.5) and the  $Num_{CandSet}$  is formed based on the maximum pairs of pixels in each direction. Thus, based on the CV loop, a 3D array of validation accuracy of size  $K \times length(T_{CandSet}) \times length(Num_{CandSet})$  is calculated.

In each CV iteration, one candidate threshold value from  $T_{CandSet}$  is used to find the pairs of peaks that violate the threshold (in the corresponding direction). Then, one candidate number of pixel pairs is selected from ( $Num_{CandSet}$ ) and used to select a number of highest peaks. The selected peaks values are reduced to  $\mu_h$  or  $\mu_v$  depending on the direction. Next, based on the 2D inverse DFT (iDFT) function shown in equation (3.6), the de-noised images are calculated in spatial domain.

$$X(m, n) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F[u, v] e^{j2\pi(u \frac{m}{M} + v \frac{n}{N})} \quad (3.6)$$

The image with alleviated horizontal noise is merged with the image with reduced vertical noise, by averaging the two matrices of inverse transform. After repeating this for all training and validation images, they are used to perform activity recognition in the next step.

Then, the recognition performance for both train and validation sets is calculated. This process is iterated inside the three nested loops for all *feasible combinations* of the three parameters, namely,  $K$  (of CV loop),  $T_{CandSet}$  and  $Num_{CandSet}$  (of the candidate sets). That explains the reason for the aforementioned dimension of the performance arrays to be three for both training and validation. The three dimensional array is also referred as training and validation cube. Figure 3.16 shows the 3D cube of validation accuracy as well as the averaged 2D array of validation performances over the  $K$  folds. In the validation and training cubes, only the elements corresponding to the *feasible combinations* of candidate thresholds and number of pixel pairs, are filled. This is explained further in the following.



**Figure 3.16:** Illustration of the 3D cube of validation accuracy and the 2D average validation array obtained by averaging over  $K$  folds of the 3D cube for selection of  $T_h^*$  and  $Num_h^*$  parameters.

Not all combinations of the candidate thresholds and number of pairs of pixels are feasible for noise removal. Because, the number of violating peaks of pixel pairs changes based on the threshold value so that, the smaller threshold increases them and the bigger threshold value reduces their number. Then, there might not exist any candidate pair of peaks for some combinations of the  $T_{CandSet}$  and  $Num_{CandSet}$ . As a result, the elements of the validation cube corresponding to those *infeasible combinations* of the candidate thresholds and the number of pixel pairs are empty.

As shown in Figure 3.16, by averaging the 3D cube of validation performances over the  $K$ -folds, 2D validation array of performance is obtained. The size of this 2D array is  $length(T_{CandSet}) \times length(Num_{CandSet})$  for each direction. The 2D array shows the variation in performance for different feasible combinations of the candidate parameters  $T_{CandSet}$  and  $Num_{CandSet}$ , shown as two heat maps in Figure 3.17 for the two directions. Then, the highest performance in the 2D arrays corresponds to the optimum parameters  $T_h^*/T_v^*$  and  $Num_h^*/Num_v^*$ . Using these parameters, the vertical and the horizontal noises are removed, resulting in two de-noised images after applying 2D iDFT transform. The two de-noised images are merged by averaging. This is performed for all training images in the data set, making them ready for activity recognition in the next step. Algorithm 3 shows the procedure, which is used on the training data only. Thus, the required parameters,  $T_h^*/T_v^*$  and  $Num_h^*/Num_v^*$  are derived and applied to the test images. A sample noisy image and the result after de-noising is illustrated in Figure 3.18.

---

**Algorithm 3: DFT-Based Periodic Noise Removal**


---

**Input:** Training data  $(X_{tr}, Y_{tr})$ , The number of folds  $K$

**Algorithm:**

1. Divide data into  $K$  training and validation sets:  $xtr_k, ytr_k, xv_k, yv_k$ .
2. Repeat the following steps for both horizontal and vertical stripes, represented as horizontal/vertical for showing all parameters.
3. Form the  $T_{CandSet} = 2, \dots, 8$  and  $Num_{CandSet} = 1, \dots$ , Maximum number of pixel pairs.
4. (1) **For**  $k=1, \dots, K$

(2) **For**  $i_1 = 1, \dots, length(T_{CandSet})$

(3) **For**  $i_2 = 1, \dots, length(Num_{CandSet})$

(4) **For**  $j=1, \dots, \text{number of images in } xtr_k \text{ and } xv_k$

-Compute the power spectrum coefficients of the  $j^{th}$  image,

-Compute  $\mu_{hj}/\mu_{vj}, \sigma_{hj}/\sigma_{vj}$ <sup>1</sup>,

-Find pixel pairs in horizontal/vertical stripes higher than:

$$T_{hj_{i_1}} = \mu_{hj} + T_{CandSet}(i_1) * 0.5 * \sigma_{hj}$$

$$T_{vj_{i_1}} = \mu_{vj} + T_{CandSet}(i_1) * 0.5 * \sigma_{vj},$$

-Reduce the first  $i_2^{th}$  greatest pixel pair values to  $\mu_{hj}/\mu_{vj}$ <sup>1</sup>

-Compute the de-noised image using 2D iDFT,

**End Loop 4**

-Use the de-noised image sets  $xtr_{dk}$  and  $xv_{dk}$  to classify the

activities and calculate the element  $(k, i_1, i_2)$  in training and

validation performance arrays,  $Prf_{tr}$  and  $Prf_v$ .

**End Loop 3,2,1**

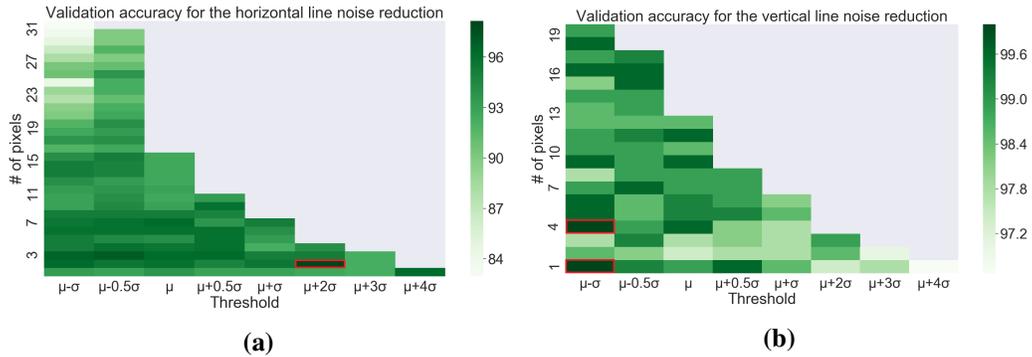
5. Average the  $Prf_{tr}$  and  $Prf_v$  arrays over the  $K$  folds

6. Find the highest validation accuracy corresponding to the optimum

threshold  $T_h^*/T_v^*$  and number of pixel pairs  $Num_h^*/Num_v^*$ .

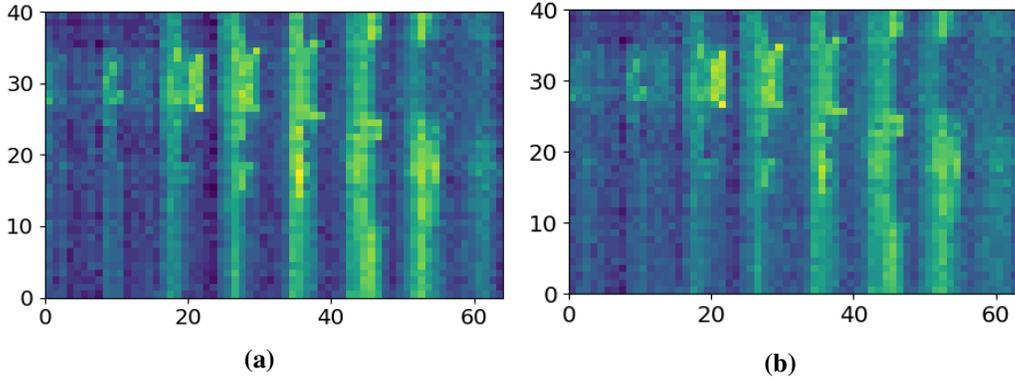
**Output:**  $T_h^*/T_v^*$  and  $Num_h^*/Num_v^*$

---



**Figure 3.17:** Heatmaps of average CV validation accuracy for selection of optimum  $T$  and  $Num$  parameters for (a) horizontal (b) vertical stripes.

<sup>1</sup>The sign / is not used for mathematical fraction, it shows both cases.



**Figure 3.18:** (a) Original image and (b) noise reduced image.

### 3.4.1.2 Feature Extraction Methods

In this work, two main categories of feature extraction approaches based on variation analysis and texture analysis are employed on the dataset for classification of activities. The variation analysis methods include the principal components based SVD and linear discriminant analysis based Fisher's canonical variables. The texture analysis method includes the frequency domain analysis based on DCT and spatial domain analysis based on GLCM.

**3.4.1.2.1 Singular-Value Decomposition (SVD)** As stated earlier, the individual pixels can be considered as independent variables. Then, the original feature space dimension in this view can be considered as the number of pixels in one slide of data (for example, the blue slide in Figure 3.14, which is  $40 \times 64 = 2560$ . 2560 is much higher than the sample size (240 for single-subject samples, 210 for double-subject samples in Coventry-2018 dataset). Having much smaller samples compared to the high dimension of the feature space can result in an ill-posed problem, models with high variance and overfit issue. Then, feature extraction for dimension reduction can alleviate this issue and improve the classification accuracy. SVD is a commonly used method for feature extraction [159, 160]. The original cuboid data matrix  $\mathbf{Z} \times 40 \times 64$  is first reformed to  $\mathbf{Z} \times 2560$  by vectorising the  $40 \times 64$  arrays. This new data form is denoted by  $S$ . Here  $\mathbf{Z}$  is the number of the selected samples for training. According to SVD, the array  $S$  can be decomposed as:

$$S = UDV^* \quad (3.7)$$

where  $U$  is a  $\mathbf{Z} \times \mathbf{Z}$  unitary matrix and  $V$  is a  $2560 \times 2560$  unitary matrix. The columns of  $V$  are the orthogonal eigenvectors describing the main directions of the data and  $D$  is  $\mathbf{Z} \times 2560$  matrix with non-negative real numbers on the diagonal which are known as singular values and often are sorted in descending order.  $(\cdot)^*$  shows the conjugation transpose. The value of the singular value indicates the level of variation along each eigenvector. In this work, the first  $k$  singular values describing 95% of the main variation of data are selected. Using these eigenvectors, the original array of data  $S$  is transformed into a much lower dimensional space, by applying  $Z_{SVD} = SV_{1:k}$ . In this work,  $k = 45$  eigenvalues are used describing more than 95% of the variance of the data. This means the feature dimension is reduced

from 2560 to 45. That is a linear combination of all variables used to transform data to a much smaller orthogonal space.

**3.4.1.2.2 Fisher’s Canonical Variables** One of the supervised dimension reduction strategies is feature extraction based on Fisher’s canonical variables. It is based on maximising the Rayleigh quotient that finds applications in feature transformation for improving the between-class separability [161, 162, 163]. Rayleigh maximisation was formulated previously in equation (3.2) in Section 3.3.2.1.2, where the sparse equivalent of the method, namely SDA, was described.

As described earlier, based on this supervised method, the data is transferred to a new orthogonal feature space so that, the Euclidean distance between the features in each class is minimized, while their distance to the features of the other class is maximized. The class centroids  $C$  of the data  $S$  lie in an affine subspace of dimension  $C - 1$ . Thus, only the first  $C - 1$  eigenvectors corresponding to the highest discriminative directions are considered. That is a considerable drop in the dimension, if the original data lies in high dimensional space, which is the case for the current data (2560). In the case of *Coventry-2018* dataset, if both single-subject and double-subjects activities can be used to shape a 15 class problem, the first 14 eigenvectors,  $l = 1, 2, \dots, 14$  are used so that, the transformed matrix is calculated as  $Z_l = v_l^T S$ , where  $v_l = W^{\frac{1}{2}} v_l^*$ . The use of the first  $C - 1$  eigenvectors in Fisher’s method results in a reduction in the overlap between the features of different classes and enhances the classification model. The above maximisation, applies an eigendecomposition on  $W^{-1}B$  to find eigenvectors that maximise the generalised Rayleigh quotient objective function. Fisher’s method is close to the RRLDA technique [163]. RRLDA also projects data into a lower dimension. Having  $C$  number of classes, with class centroids matrix  $M$  of size  $C \times n$  in  $n$ -dimensional input space, first the class centroids are mapped using the common within-class covariance  $W$ . That is performed by eigendecomposition of  $W$  so that, the transformed matrix of class centroids  $M^* = MW^{-\frac{1}{2}}$  is calculated. RRLDA assumes a Gaussian distribution of data and this transformation of class centroids minimises the overlap of the Gaussian data. This new subspace can be further decomposed into successively optimal subspaces in terms of centroid separation. For this aim,  $B^*$  (the between-class covariance of  $M^*$ ) is decomposed as  $B^* = V^* D_B V^{*T}$ . The columns of  $V^*$  are the eigenvectors  $v_l^*$  sorted based on their variances described by their corresponding eigenvalues. They define the coordinates of the optimal subspaces, while  $D_B$  is a diagonal matrix with the corresponding eigenvalues.

**3.4.1.2.3 Discrete Cosine Transform (DCT)** For texture analysis, the 2D spatio-temporal maps of size  $40 \times 64$  are considered as images with unique pattern of textures for each type of activity. Then, based on this view, image texture analysis strategies are employed.

For texture analysis in frequency domain, images can be represented as frequency coefficients sorted in a two dimension representation based on 2D DCT transform. Depending on the nature of the image, having smooth variation of colours or sharp changes, the higher

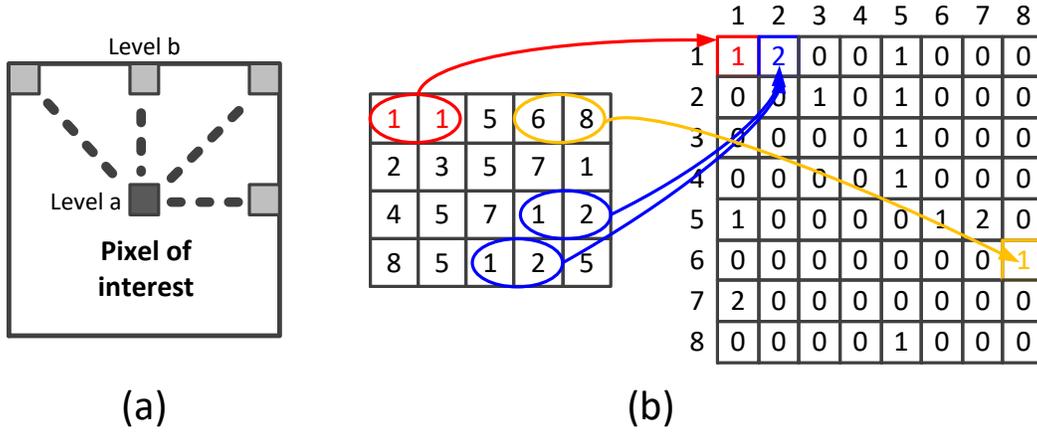
value DCT coefficients are distributed in lower or higher frequencies respectively. The DCT method is similar to DFT, as it transforms a signal or an image to the frequency domain. However, it sorts the frequency information of the image from lowest (DC) to highest. In this study, 2D DCT is used to extract spectral information from the 2D spatio-temporal maps shown in the middle part of Figure 3.14. From mathematical perspective, DCT represents an image as a sum of sinusoids of varying magnitudes and frequencies [164]. In fact, it is a real, orthogonal, linear transformation that can be represented based on equation (3.8):

$$F[p, q] = \frac{2}{\sqrt{MN}} a_p a_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f[m, n] \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N} \quad (3.8)$$

where,  $F[p, q]$  values are the DCT coefficients at row  $p$  and column  $q$ ,  $f[m, n]$  is the element in row  $m$  and column  $n$  of the image matrix and  $M = 40$  and  $N = 64$ .  $0 \leq p \leq M - 1$  and  $0 \leq q \leq N - 1$ . Moreover,  $a_p = \frac{1}{\sqrt{2}}$  if  $p = 0$  and it is 1 otherwise. Similarly,  $a_q = \frac{1}{\sqrt{2}}$  if  $q = 0$  and its value is 1 otherwise.

The resulting 2D DCT map has the same size as the original image, which is  $40 \times 64$  in this work. The selected coefficients in this work are from the lower order coefficients due to the higher energies observed there. Therefore, six significant coefficients are extracted in a zigzag pattern from the top left corner of the 2D DCT map. The coefficients are then concatenated into a feature vector. In this work, the images are divided to 40 sub-patches. 2D DCT is applied to each sub-patch for subsequent feature selection. Therefore, six coefficients are extracted from the top left corner in a zigzag pattern from each sub-patch. The resulting feature vectors are of size  $\mathbf{Z} \times 40 \times 6 = \mathbf{Z} \times 240$ .  $\mathbf{Z}$  varies depending on the number of data samples in large layout or small layout as described earlier.

**3.4.1.2.4 Grey-Level Co-occurrence Matrix** GLCM is a common technique for texture analysis and is used to quantify the spatial alternation of pixel values [165]. For example, how often a pixel with the similar intensity  $L_a$  occurs at a distance  $d$  to another pixel with the intensity  $L_b$  in horizontal, vertical or diagonal direction as shown in Fig. 3.19. Due to the nature of the  $40 \times 64$  images in this work, this method is applied on the images of the cuboid  $\mathbf{Z} \times 40 \times 64$  shown in Figure 3.14. GLCM uses decomposition levels denoted as scales for reducing the intensities in an image. Depending on the range of intensities in an image, a number of scales are defined. Figure 3.19 shows an example of the formation process of a GLCM matrix based on horizontal occurrences at  $d = 1$ . The gray scales are between 1 to maximums 8 in this case. Then, the GLCM square matrix of the same dimensional size is formed. The image pixels are quantized based on the discrete scales and the GLCM matrix is filled for each direction.



**Figure 3.19:** (a) Forming GLCM matrices in four directions i.e.,  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ; (b) Computation of GLCM matrix based on horizontal occurrences at  $d = 1$  for an image.

Next, four second order statistics metrics are calculated from GLCM matrix: Correlation, Contrast, Dissimilarity, and Energy that are defined in equation (3.9):

$$\begin{aligned}
 \text{Correlation} &= \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i,j)}{\delta_i \delta_j} \\
 \text{Contrast} &= \sum_{i,j} (i - j)^2 p(i,j) \\
 \text{Dissimilarity} &= \sum_{i,j} |i - j| p(i,j) \\
 \text{Energy} &= \sum_{i,j} p(i,j)^2
 \end{aligned} \tag{3.9}$$

where  $p$  is the grey co-occurrence matrix,  $i, j$  denote row and column indexes,  $\mu_i, \mu_j, \delta_i$  and  $\delta_j$  are the means and standard deviations of  $p_x(i)$  and  $p_y(j)$ , so that  $p_x(i) = \sum_{j=0}^{G-1} p(i,j)$  and  $p_y(j) = \sum_{i=0}^{G-1} p(i,j)$  and  $G$  is the number of intensity scales used during GLCM matrix forming.

In this work, first the  $40 \times 64$  images are divided into 5 local patches of size  $8 \times 64$  and then, for each local patch, GLCM matrices are calculated for two different distances (one and three number of pixels) and three different angles  $0^\circ$ ,  $45^\circ$  and  $90^\circ$ . This results into  $5 \times 2 \times 3$  GLCM matrices. For each GLCM matrix, four GLCM statistics are calculated as described in equation (3.9). Then, for each  $40 \times 64$  image, 120 texture features are extracted based on GLCM analysis resulting into GLCM features of size  $\mathbf{Z} \times 120$  for the cuboid data sets  $\mathbf{Z} \times 40 \times 64$ .

### 3.4.1.3 Classification Methods for Activity Detection

In this work, four traditional classification methods are used: SVM, RF, k-NN, and LR, which were described earlier in Section 3.3.2.2. All four techniques are supervised and have shown suitability for multi-class classification problems.

### 3.4.1.4 Deep Neural Networks (DNN) for Activity Detection

DNN models can learn a hierarchy of features from low-level features to build high-level ones. Therefore, the process of feature extraction is automated when there is a lack of empirical or hand-crafted features. In this work, the DNN model is used for analysis of streams of LRIR images for human activity recognition. As the LRIR data corresponding to a given activity can be treated as the sequence of the IR frames, the mixed convolutional and recurrent NN model is explored. The CNN layers are used to interpret the features of each frame, while the LSTM layer is used to explore the sequential relation between the frames.

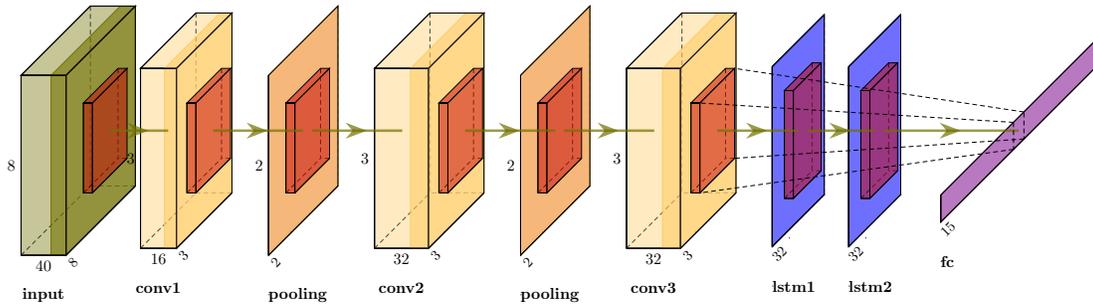
The CNN-LSTM belong to the class of models that are both spatially and temporally deep. They are flexible to capture spatial features from each time frame data and also capture their temporal variations. Then, they are capable to utilise both spatial and temporal features for prediction over sequences of varying inputs frames [166]. For this aim, they utilise CNN layers in early parts of their architecture to extract deep spatial features from input data such as video frames that have sequential nature. The sequential deep features are fed into the LSTM layer to learn temporal variations and perform prediction.

In this work, three convolutional and two LSTM layers are designed for the CNN-LSTM architecture. Then, the input are samples each consisting 40 frames of size  $8 \times 8$  and their corresponding activity label. The model robustness to the number of layers is validated in Section 3.4.2.3. The convolutional layers are used to extract spatial features from the LRIR streams, while the LSTM layers learn the temporal variations of the sequence of features. The developed classification model is capable of identifying the activities. The number of epochs used until model convergence is 1000, while the batch size is 32. The developed CNN-LSTM architecture with the hyperparameters is represented in Fig. 3.20. The first convolutional layer has 16 filters with  $3 \times 3$  size. Regarding the second and third convolutional layers, 32 filters with the same size are used. The two LSTM layers have both 32 filters.

## 3.4.2 Evaluation and Results

The datasets in this study are divided into training (75%) and testing (25%). The testing accuracy is used as an evaluation metric. In this section, the following experiments are presented:

- Noise-reduction test - the test is performed on *Coventry-2018* dataset to outline the models performance on noisy data and noise-reduced data.

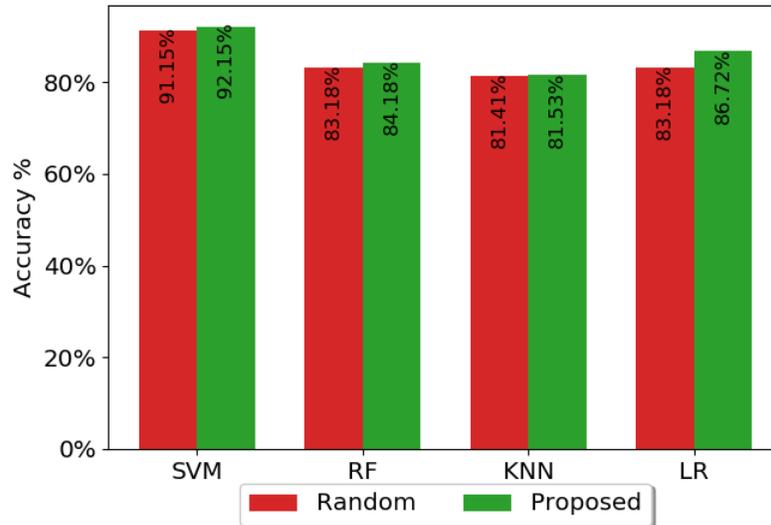


**Figure 3.20:** Illustration of the CNN-LSTM architecture.

- Comprehensive model comparison test - the test is performed on *Coventry-2018* dataset to discover the most optimum model. Then, the best model is applied to the total class problem as well as different subsets of the *Infra-ADL2018* dataset.
- Layout-sensitivity test - the test is performed on *Coventry-2018* dataset to compare small layout and large layout.
- Subject-sensitivity test - the test is applied on both *Coventry-2018* dataset and *Infra-ADL2018* dataset to report classification results on single-subject activities data and double-subject activities data.
- Model-generality in terms of layout - the test is applied on *Coventry-2018* dataset, where the models are trained on small layout and tested on large layout, and vice versa.
- Model-generality in terms of subjects - the test is applied on *Infra-ADL2018* dataset to report results on models trained with a fixed number of subjects and tested on subjects data not used for training.
- Optimum sensor test - all individual sensors and combinations of them are used to discover the most optimum sensor or combination solution.

### 3.4.2.1 Frames Equalisation

The proposed methods for interpolation and extrapolation in Algorithm 1 and Algorithm 2 respectively are compared with random methods for the same purpose. In terms of interpolation, random frames are selected from the existing frames, which are then added to the existing matrix. The random extrapolation method selects random frames to be deleted until the total number of frames becomes 40. For comparison, the proposed methods and random methods are tested on the 15-class problem in the large layout from the *Coventry-2018* dataset. More precisely, the raw data are used before applying periodic noise reduction and feature extraction techniques for this experiment. Figure 3.21 outlines the performance of the proposed methods against random techniques. The proposed methods exhibit an improvement of up to 3.54% for LR, which will later be evaluated as the most optimum classifier.



**Figure 3.21:** Comparison of the classification performances of random methods and proposed methods for interpolation and extrapolation before denoising (BD) and feature extraction using all four classification methods. The Coventry-2018 large layout data (15-class problem) of Sensor-1 was used.

### 3.4.2.2 Periodic Noise Removal

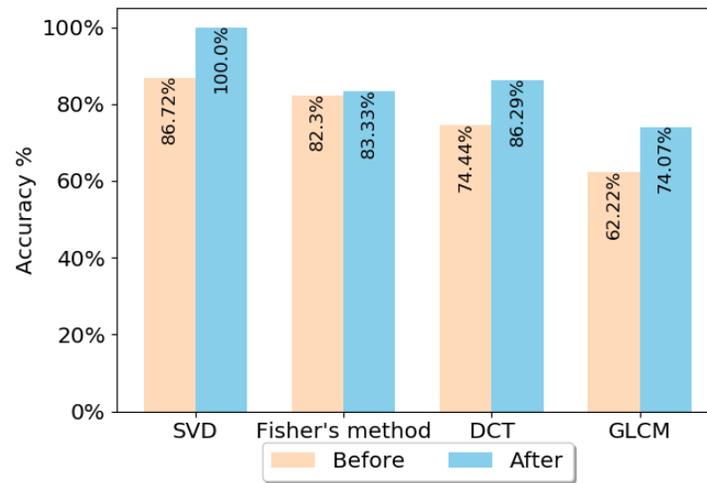
**3.4.2.2.1 ANOVA Test** In order to demonstrate the significant difference in the mean values over the horizontal and vertical stripes (excluding the DC component), one-way ANOVA is applied as described in Section 3.4.1.1.2. The null hypothesis is that the average of the horizontal and vertical stripes in power spectrum 2D arrays (excluding the DC component) is not significantly different for all 2D spatio-temporal maps. F-score and p-value are calculated for each sensor and direction (horizontal and vertical) as outlined in Table 3.4. The p-values are very small and the null hypothesis is rejected at 0.01 level. Thus, it is concluded that the threshold values based on average of stripes must be customized for each 2D array and a pre-defined threshold cannot be used for the same purpose.

**Table 3.4:** The f-score and p-value results of the ANOVA analysis for the average horizontal and vertical stripes (excluding the DC components) of the 2D power spectrum for all three sensors (S1, S2, S3), and a combination of them (S1+S2+S3) using large layout and 15 classes data from Coventry-2018 dataset.

	Horizontal line		Vertical line	
	f-score	p-value	f-score	p-value
S1	507.06	4.71e-259	68.51	1.67e-100
S2	418.67	3.89e-242	40.83	3.06e-70
S3	551.57	1.50e-266	81.58	5.02e-112
S1+S2+S3	832.71	0	137.26	1.64e-246

**3.4.2.2.2 Result of Periodic Noise Removal Cancellation** In order to evaluate the effect of the proposed periodic noise removal strategy, the classification results of data before pre-processing and after pre-processing are compared. For this aim, the *Coventry-2018* dataset for 15-class problem including both single subjects and double subjects activities of Sensor-1 was used. Sensor-1 is selected as the optimum sensor selection, which is evaluated later in this Chapter.

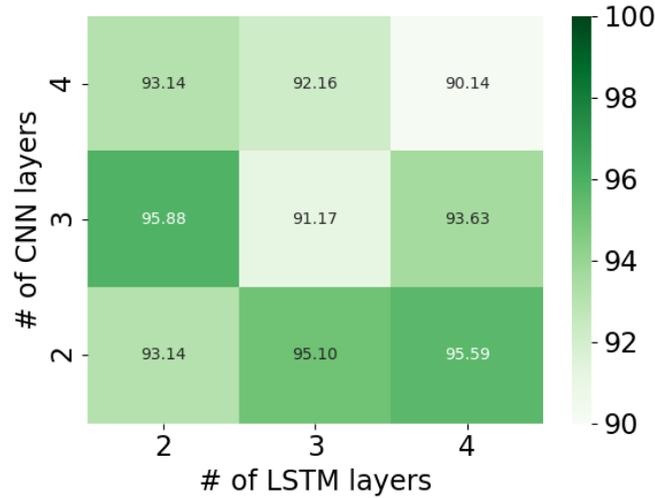
The four feature extraction methods: SVD, Fisher's method, DCT, and GLCM with LR are used to train classification models. The LR classifier was chosen because it was one of the best classifiers compared to others in terms of accuracy. This will be shown in next section. The models are trained using the raw data (before denoising) first. Then, the same modeling strategies are applied on the denoised data sets. The results of this experiment show an increase in accuracy. Figure 3.22 shows the result of classification. The *Coventry-2018* large layout Sensor-1 data was used for these tests. The results show improvement in the classification performance in all cases.



**Figure 3.22:** Comparison of the classification performances before denoising (BD) and after applying the DFT-based periodic noise removal algorithm (AD) using different feature extraction methods and LR classification. The *Coventry-2018* large layout data (15-class problem) of Sensor-1 was used.

### 3.4.2.3 CNN-LSTM Model Robustness Evaluation

In order to discover the most optimum model for the CNN-LSTM architecture, a number of combinations of CNN and LSTM layers are used. Average testing accuracy is used as a model evaluation metric. The models are tested on *Coventry-2018* dataset for large layout with all 15 activities. In particular, two, three and four layers are used for both CNN and LSTM. All nine combinations are evaluated for random splits of *Coventry-2018* dataset. Figure 3.23 reveals the testing accuracy for each model. As observed, 3CNN-2LSTM represents the model with the highest accuracy.



**Figure 3.23:** CNN-LSTM model evaluation using Coventry-2018 large layout data (15-class problem) of Sensor-1.

#### 3.4.2.4 Comprehensive Comparison of Recognition Techniques

This section compares the capability of different combinations of feature extraction and classification methods described in Section 3.4.1.2 as well as the CNN-LSTM in activity recognition. The idea is to evaluate the methods using the most challenging dataset at hand. Therefore, instead of evaluating based on the 8-class data (of single-subject activities) or 7-class data (using double-subject activities), the 15-class data including both single-subject and double-subject activities from the *Coventry-2018* large layout Sensor-1 are considered in this section. The hand-crafted feature extraction methods are SVD, Fisher’s Canonical Variables, DCT and GLCM. The investigated classifiers include SVM, RF, k-NN and LR. K-fold CV ( $K = 10$  in this work) is used in order to find the optimal parameters for the classifiers. Therefore, the selected dataset is divided randomly into training and testing for 10 times to perform the tests for several times (10) and the average and standard deviation of the results are reported. The percentage of the training and testing sets are 75% and 25% respectively.

Table 3.5 presents the results achieved based on different feature-classifier combinations as well as the CNN-LSTM model. The results show that the applied methodologies were successful to discriminate activities. The highest accuracy obtained based on the unsupervised variance-based SVD and CNN-LSTM model, where the former achieves 100% when using LR classifier. Although CNN-LSTM model is more complicated compared to SVD+LR model in terms of computation time and implementation to the embedded systems, its performance is not higher. Justification of the results will be given in Chapter 6.

Based on the classification results for *Coventry-2018* dataset, SVD feature extraction with LR classifier is selected as the optimum modeling strategy. Therefore, this modeling strategy is used for the rest of the tests in this study. Since in the *Infra-ADL2018* dataset

**Table 3.5:** Testing performances for all combinations of the feature extraction and classification methods as well as CNN-LSTM using the Sensor-1 large layout data set from Coventry-2018 for 15 classes of single subject and double subject activities.

	SVM	RF	KNN	LR	CNN-LSTM
SVD	96.66%±0.9	97.02%±1.04	88.88%±1.57	100%±0	
Fisher'method	87.03%±2.28	88.14%±0.52	86.29%±1.04	83.33%±1.6	
DCT	88.14%±2.28	92.21%±0.9	87.39%±0.5	86.29%±1.38	
GLCM	77.39%±2.09	82.22%±1.81	73.7%±2.09	74.07%±1.04	
CNN-LSTM					95.88%±1.1

among the three categories of activities for single-subject, double-subject and three-subject activities, the number of samples for the latter are very limited (18 samples per sensor), making it very difficult to train a model. Therefore, the selected optimum model was used to train classifiers for single-subject (9 classes) and double-subject (10 classes) activities using each sensor data separately, as shown in Table 3.6. In addition, all of the subjects data were used to develop models for classification of all activities (21 classes) using each sensor data. The achieved results show the robustness of the classification methods on this dataset.

**Table 3.6:** Testing performance of the classification models using SVD+LR for single-subject, double-subject, and all activities of Infra-ADL2018.

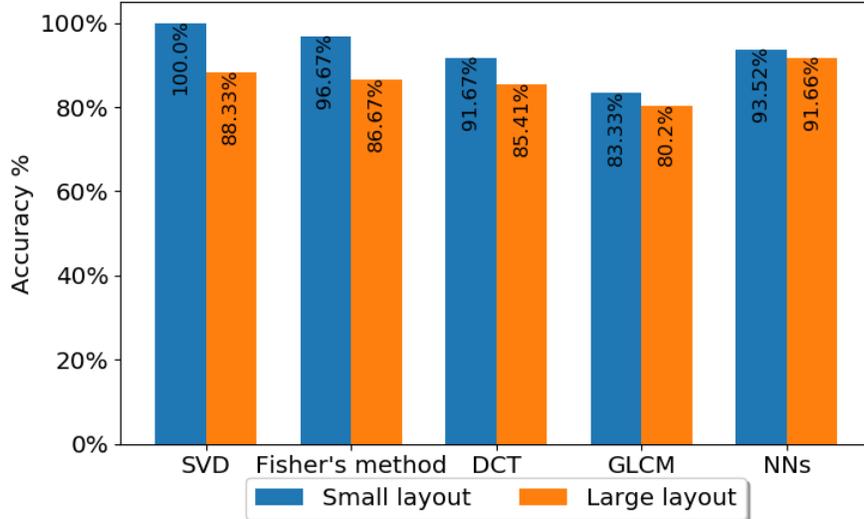
	Single-subject	Double-subject	All subject
Sensor 1	94.53% ±0.77	100% ±0	96.02% ±0.72
Sensor 2	95.62% ±0.77	100% ±0	100% ±0
Sensor 3	98.36% ±1.33	100% ±0	97.34% ±0.37
Sensor 4	95.08% ±1.33	100% ±0	95.49% ±0.37

Because of the robustness of classification models on individual sensors data, there is no need to use several sensors, which increases the hardware and software complexity. However, as shown in Table 3.6, the nine-class single-subject classification using individual sensors (Sensor-1, Sensor-2, Sensor-3 and Sensor-4) shows that the results deviated from 100%. Therefore, experiments for combination of two sensors data were performed. All six possible combinations of two sensors data including single-subject activities are formed. Using the SVD with LR classification, the average test accuracy was 100%±0 for all six models. Overall, considering this result and the results in Table 3.6, the use of Sensor-2, which is a side sensor, showing > 95% accuracy is the optimum choice.

### 3.4.2.5 Layout-Sensitivity Test Results

One of the concerns about the use of LRIR sensors is sensitivity in their distance to the target. In this section, the impact of distance when using the LRIR sensor for human target activity recognition is experimented. For this aim, the single-subject small layout (1.5 m away from area of interest) 240 × 40 × 64 data and the single-subject large layout (2.5 m away from area

of interest)  $240 \times 40 \times 64$  data are utilised. The performance of the four SVD, Fisher, DCT and GLCM feature extraction methods using the LR classifier as well as the CNN-LSTM model are compared. In Figure 3.24, the recognition performances are shown.



**Figure 3.24:** Comparison of the effect of small layout and large layout data in classification performance. The LR classifier with the four feature extraction methods are applied on the single-subject activities for small layout and the single-subject activities for the large layout as well as CNN-LSTM using Coventry-2018 Sensor-1 large and small layout data sets.

As expected, the accuracy for the small layout is higher as the subject body profile will cover more pixels. Nevertheless, the difference in the accuracy is not significant, which showcases that activities are correctly classified in both cases. That indicates the applied techniques including pre-processing, feature extraction, and classification are applicable for sensor settings placed in different distances with respect to the subject.

### 3.4.2.6 Model-Generality in Terms of Layouts

A classical machine learning problem is whether a model can generalize for different scenarios. In *Coventry-2018* experiments, two facts are worth to be noticed: *i)*, there are small and large layouts; *ii)*, The room temperature increased during the experiments when collecting the large layout data. These two facts can be utilised to test the models' generalization when having unseen layout condition after training. For this aim, two tests are performed:

1. **Extreme model generalisation:** The model which was trained using the single-subject Sensor-1 data in small layout was tested on the single-subject Sensor-1 data in large layout. The same experiment was also repeated for the large layout as train and small layout as test. The distance from sensors to the subjects is different. In addition, the average and standard deviation of the pixels for the small layout are  $17.47 \pm 1.21$ ,

while for the large layout, they are  $18.86 \pm 1.11$ . All combinations of feature extraction and classification methods as well as CNN-LSTM were tested. GLCM for feature extraction with LR for classification achieved the best accuracy in both experiments using Sensor-1. In the first experiment, the accuracy was 71%. In the opposite scenario, where the large layout is used for training and the model is evaluated on the small layout, the testing accuracy was 64.58% using the same feature extraction and classification strategy. (GLCM + LR) can still generalise and show intermediate levels of results of 71% for small layout training and large layout testing. This is a positive indication that such systems still work in the case of changes in the original settings such as displacement, etc.

2. **Mixed model generalisation:** The single-subject data from both small layout (240 samples) and large layout (240 samples) are mixed, which represents the moderate scenario. Therefore, the data are divided into training and testing. This ensures that samples from both layouts are distributed both in the training set, and testing set. Similarly to the previous experiment, all classification models were tested in the pursuit of the most optimum model. As a result, the most accurate model was the automatic feature extraction and classification with the deep CNN-LSTM architecture, achieving  $94.99\% \pm 1.66$  average testing accuracy. Considering the very high accuracy, this experiment shows that the approximately equal distribution of data from both layouts in all sets provides a very accurate representation of the data.

### 3.4.2.7 Sensitivity to the Number of Subjects Test Results

In this section, the sensitivity of the algorithms in detection of the activities with a different number of subjects are evaluated. For this aim, the single-subject dataset of size  $240 \times 40 \times 64$  and double-subject of size  $210 \times 40 \times 64$  both from *Coventry-2018* large layout settings of Sensor-1 are used for comparison. In addition, single-subject activities of size  $243 \times 40 \times 64$  and double-subject activities of size  $240 \times 40 \times 64$  from *Infra-ADL2018* using Sensor-2 are also compared. Fig. 3.7 describes the test accuracy of LR classifier when applied on SVD, Fisher's method, DCT and GLCM features as well as the CNN-LSTM structure. The graph shows that the double-subject activities outperform the single-subject activities for all five models.

### 3.4.2.8 Model-Generality in terms of Subjects

The dataset *Infra-ADL2018* includes samples from nine subjects. As such, leave-one-subject-out experiment is conducted to evaluate the model tested on unseen subject data. Similarly to the other experiments, the most optimum model SVD+LR is selected for training and testing the models. The data from the side Sensor-2 is used as it outperformed the other three sensors. Considering single-subject activities, all nine cases were considered so that, each time eight subjects were used for training and the remaining one was used for test. Regarding double-subject activities, there exist eight combinations of subjects for conducting the

**Table 3.7:** Comparison between single-subject activities and double-subject activities using Coventry-2018 large layout data of Sensor-1 and Infra-ADL2018 Sensor-2 data. LR was used together with all feature extraction methods as well as CNN-LSTM.

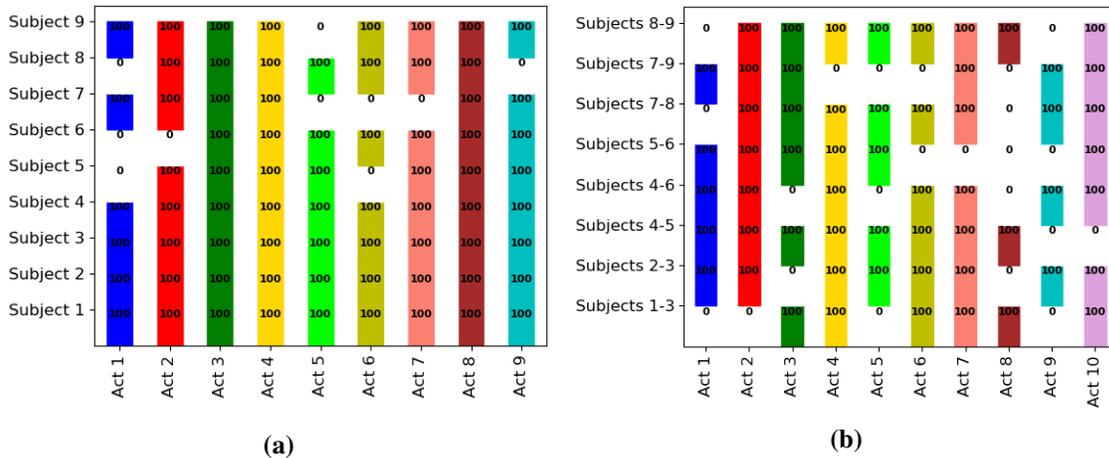
	Coventry-2018		Infra-ADL2018	
	Single	Double	Single	Double
SVD	88.88%±1.57	98.11%±0	97.63%±2.01	100%±0
Fisher	86.11%±1.56	91.82%±0.89	83.05%±2.78	97.21%±0.78
DCT	84.71%±0.98	92.25%±1.03	92.5%±0.9	100%±0
GLCM	83.33%±1.69	84.91%±1.12	68.02%±0.96	96.52%±.98
CNN-LSTM	96.29%±2.62	96.87%±2.2	95.13%±5.77	96.66%±6.66

activities. Therefore, all eight possible cases are presented similarly. The average testing accuracies with standard deviations are seen in Table 3.8.

**Table 3.8:** Testing performance of leave-one-subject-out model generalisation for single-subject activities and double-subject activities of Infra-ADL2018.

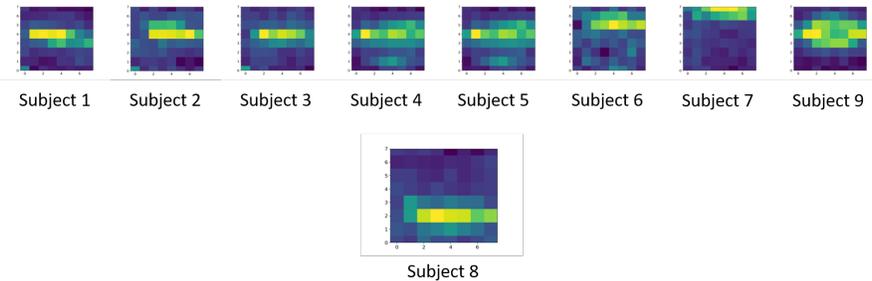
	Single-subject	Double-subject
SVD+LR	88.06% ±12.32	71.25% ±9.27

The results reveal considerably high standard deviation for both cases. The probable reason is the difference in the body characteristics and motion behavior of the unseen test subjects compared to the training set. The average activity recognition performance over three repetitions of each activity is visualised for each unseen test subject in Figure 3.25. The graph showcases the poorly classified activities.



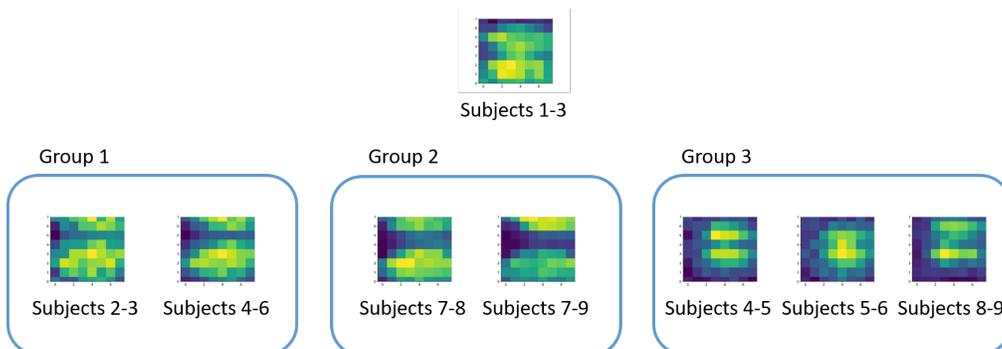
**Figure 3.25:** (a) Single-subject activities and (b) double-subject activities. The vertical axis show the unseen test subject and the numbers show the average test accuracy over three repetition of each activity.

As observed in Fig. 3.25a, Activity 3, Activity 4 and Activity 8 are the three single-subject activities, which are recognised for all three repetitions and nine subjects. The activities which are not recognised for the specific cases yield 0% average accuracy for the three repetitions. This shows that a subject pattern used for testing is likely to be noticeably different from the subject patterns used for training. To visualise this, one activity that was not recognized for an unseen subject is considered. Then, the average of some key frames is visualised for the unrecognized activity for all test subject. This helps to identify the difference from the unrecognized subject profile compared to the rest of the subjects. Our research discovered that the middle frames represent the essence of the activities, while the starting and ending frames represent mostly still environment as observed in Section 3.3.3.2. The following Figure 3.26 represents the patterns for Activity 9, which was not recognised for subject 8 as observed in Fig. 3.25a.



**Figure 3.26:** Patterns for activity 9 using all single-subjects.

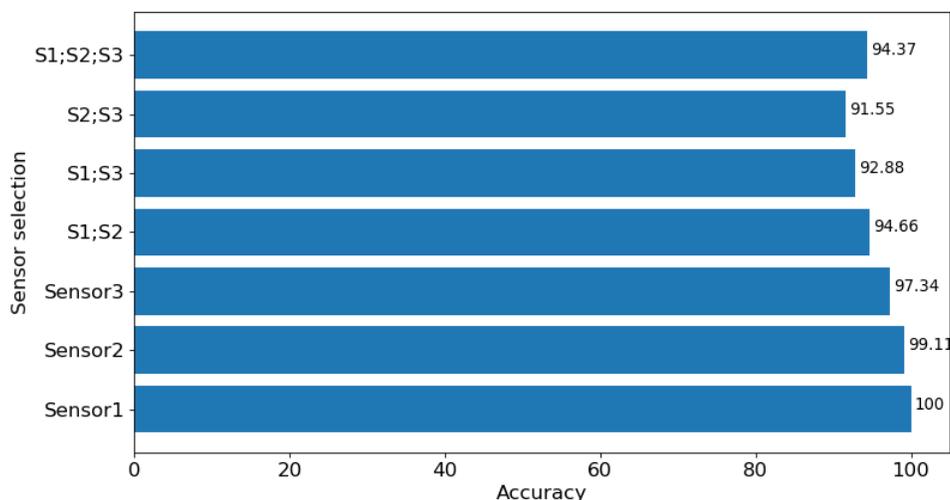
Figure 3.25b reveals the performance for double-subject activities in leave-double-subject-out experiments. Likewise, some of the activities were not recognised for all three repetitions or they were recognised for all three. To evaluate this, the same average frame approach is generated to discover whether the unrecognized activity of a double subject profile differs from the rest. Fig. 3.27 reveals Activity 2 patterns, which was not recognised for double subject 1-3 as seen in Fig. 3.25b. As observed, the patterns for double subjects 2-3 and 4-6 are similar. Likewise, the double subjects 7-8 and 7-9 represent the second group of similar patterns. Finally, the double subjects 4-5, 5-6, and 8-9 have similar profiles. However, none of these groups' patterns are similar to the only unrecognized double subject's pattern.



**Figure 3.27:** Patterns for activity 2 using all double subjects.

### 3.4.2.9 Optimal Room Setup and Sensor Selection

The room setup was part of a comprehensive discussion prior to any experiment conduction. For *Coventry-2018* experiments, instead of using one sensor attached to the ceiling of the room, it was decided to use three sensors in total. Then, it appeared that the most successful results come from Sensor-1 with a side view to the subjects. In order to identify the optimum sensor arrangement, the classification accuracy for different combinations of sensors are compared. The results of this comparison are shown in Figure 3.28.

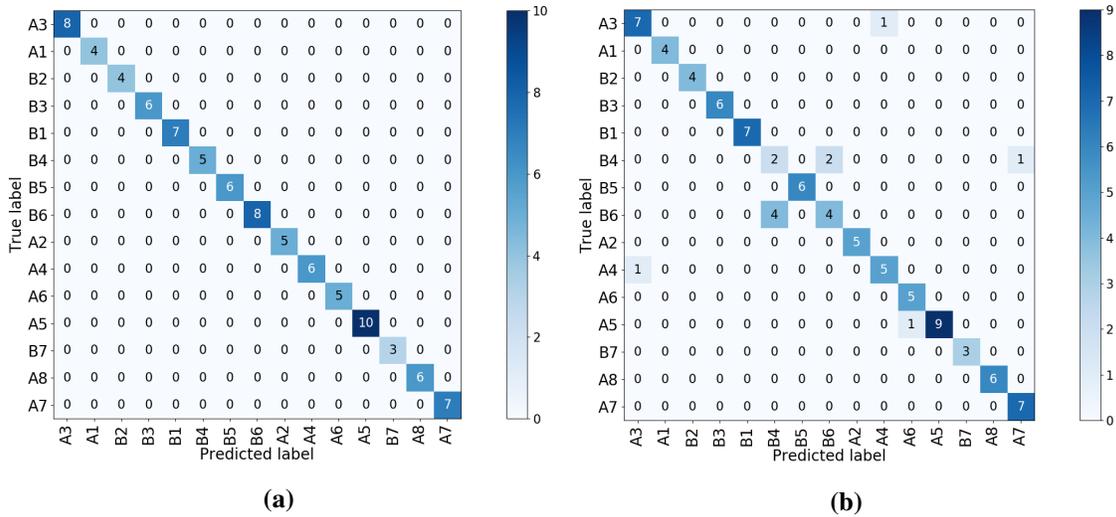


**Figure 3.28:** Classification accuracy for individual sensors and combinations of them for *Coventry-2018* dataset using SVD+LR on large layout for the 15-class problem including single and double subjects.

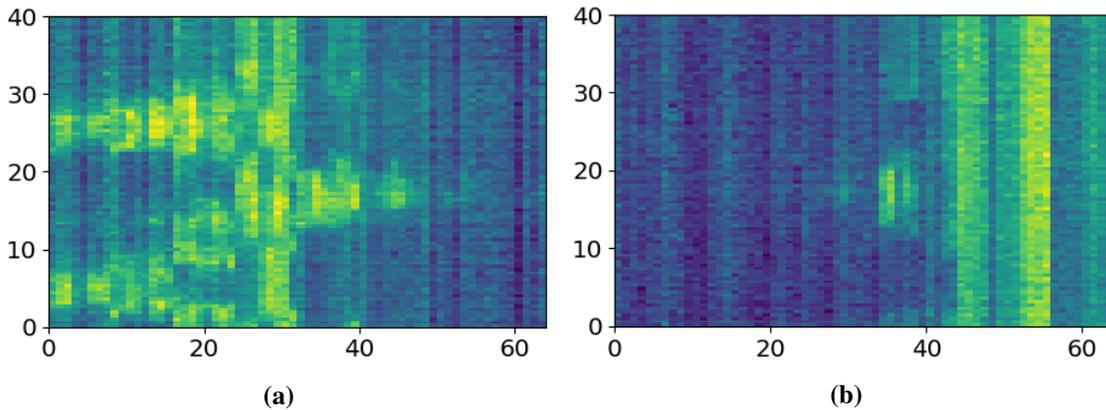
By using multiple sensors, a systematic review of the performance of each sensor can be performed. As seen in Figure 3.28, using Sensor-1 data, the best accuracy is achieved, showing that the other two sensors can be excluded from the experiments. Considering the fact that Sensor-1 is a side sensor, it can better capture activities such as sitting, where a front sensor or a ceiling sensor would fail. That simplifies both hardware and software requirements. Furthermore, as mentioned earlier, the temperature of the room rose from 18° centigrade to 21° over the experiments in the large layout. However, the classification models still remain accurate despite of the higher temperature in the room.

Another issue to investigate is the reason for the difference in the performance of the two side sensors, Sensor-1 and Sensor-3. While they are symmetric with respect to the subjects position, Sensor-3 shows lower accuracy compared to Sensor-1. The confusion matrices shown in Fig. 3.29 outline that the lower accuracy of Sensor-3 is due to the double-subject activities such as B4 (Small Movements) and B6 (Standing; Moving). This is due to the fact that these activities are not symmetric as visualised in Fig. 3.30. In a real world scenario there are many different range of activities and further studies are required about the optimum set of sensors position and orientation for accurate activity recognition in practical settings.

Furthermore, the confusion matrices in Fig. 3.29 prove that the pair-subject activities are problematic and this leads to lower performance of the data captured by Sensor-3.



**Figure 3.29:** Comparison between the confusion matrices of the two side sensors, Sensor-1 (a) and Sensor-3 (b), of Coventry-2018 dataset. The large layout with single and double subject data was used for training a model based on SVD+LR.



**Figure 3.30:** Comparison between Coventry-2018 Sensor-1 (a) and Sensor-3 (b) for “Standing + Small Movements” activity.

### 3.5 Chapter Conclusion

This chapter of the thesis presented a novel method for DFT-based supervised periodic noise reduction aimed for the 2D spatio-temporal maps of LRIR data. The performance of the method, compared with noisy data, achieved outstanding results of up to 13.28% for the model SVD+LR. Comparison between single-subject data and double-subject data from the two datasets, *Coventry-2018* and *Infra-ADL2018*, discovered double-subject activities as more correctly classified for all used models. Unsurprisingly, the most optimum model

SVD+LR showed better performance on small layout data, while its recognition accuracy dropped when using the large layout data from *Coventry-2018*. In addition to the contributions, novel methods for interpolation and extrapolation were proposed. Despite the limitations of the extrapolation method, the two strategies have shown an improvement compared to random selection methods.

The first novel strategy was in formation of the 2D spatio-temporal maps from LRIR streams followed by the proposed periodic noise reduction algorithm that improved the recognition accuracy. Despite the great efforts of previous research studies, no comparison for the different aforementioned scenarios was achieved as well as discovery of the most optimum machine learning or deep learning model. Frames equalisation was not discussed in the current literature, which lacked any proposed methods for the purpose of interpolation and extrapolation.

This work brings us closer to implementing a working LRIR system for applications in healthcare and eldercare. However, the main limitations of the presented study are mostly the limited number of samples for both datasets. In addition to this limitation is the fact that datasets were collected by younger adults, which may not correctly resemble the activities and movements of older adults.

# Chapter 4

## Unsupervised Doppler Radar Based Activity Recognition

### 4.1 Introduction

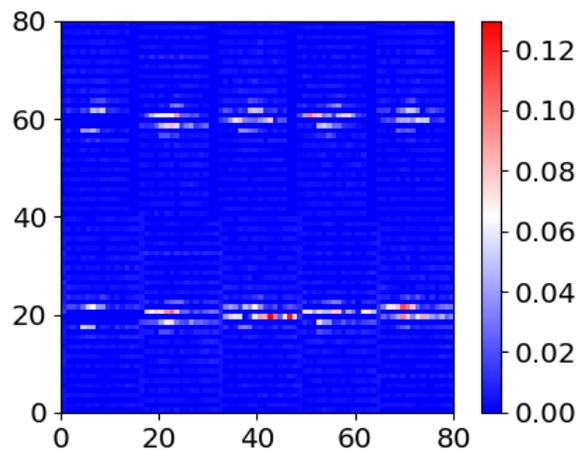
The previous chapter was focused on IR sensors for human activity recognition using supervised learning strategies. A number of methods were proposed for data pre-processing. In contrast to the previous chapter that was based on supervised learning strategies, in this chapter unsupervised learning strategies will be considered for human activity recognition using passive micro-Doppler radars.

Passive micro-Doppler radars have been introduced for human activity recognition [59], which led to a diverse number of applications, mostly focused on supervised scenarios. These radars are a preferred option for human activity recognition applications considering their privacy-preserving advantage, deep penetration, high distance range, and substantial accuracy rates [60], [62]. Micro-Doppler radars use the Doppler effect, a phenomenon caused by motion-related activities such as walking, running, jumping and others, which cause a shift in the frequency. Therefore, the shift in the frequency can be used for detecting these activities.

As described in Chapter 2, unsupervised learning applications for human activity recognition with micro-Doppler radars is a largely ignored area in the literature. However, such applications have advantages such as removing the need for manually labelling the data. Clustering methods are only guided by the variations and characteristics of the data samples, which are then grouped in clusters. Obviously, the absence of labels can negatively influence the quality of the clustering results. On the other hand, the models can be updated faster in unsupervised applications, while supervised applications are highly limited due to the need for labelling any new coming data. In a real world scenario in a care home, a diverse number of activities are performed including walking, sitting, drinking a cup of tea, reading a newspaper, and others. However, only critical activities are important such as a fall or immobility. Considering this, such activities can be labelled and recognised for future scenarios. How-

ever, this work is only focused on unsupervised activity clustering of Doppler radar data and the latter problem is not addressed in this thesis. In addition, unsupervised learning usually requires the use of techniques for estimating the number of clusters. This is due to the fact that subjects conduct a broad number of activities in a real world scenario. As such, embedding all activities in a pre-collected dataset for supervised frameworks is problematic.

The employed mmWave Doppler radar dataset in this chapter was equalised and normalised as it will be explained in detail in Section 4.2. Each data sample was vectorised to 6400 variables. In order to obtain 2D images for subsequent image analysis,  $80 \times 80$  reshaping was applied. Figure 4.1 reveals the patterns of the activity walking (1). It is obvious that some areas of the image present richer information, while others can be removed from the analysis. This notion will be used for the the proposed two unsupervised feature extraction methods based on local patching.



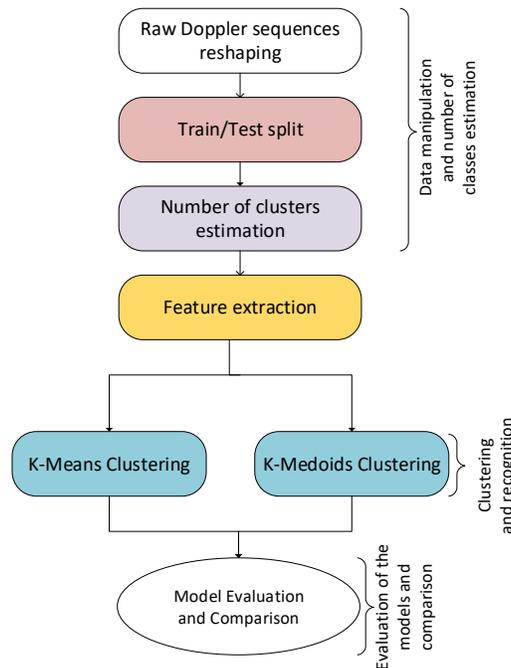
**Figure 4.1:** Example of an  $80 \times 80$  image of the activity walking (1).

Currently, only two studies are focused on the use of unsupervised learning strategies for Doppler radar-based human activity recognition [23, 24]. Considering their main limitation with intermediate level of recognition accuracy, this chapter aims to deliver unsupervised solutions using the mmWave Doppler radar dataset for the aim of activity recognition with improved results. By addressing the variations in 2D radar profiles, different image feature extraction methods can be used for activity detection. For this reason, in this thesis the author is investigating interpretation strategies for the 2D images. In particular, considering the high dimensionality of the radar data, two novel unsupervised feature extraction methods for 2D Doppler maps will be proposed. These two methods are based on analysis of image local areas using 2D DCT and entropy analysis. Furthermore, CVAE will be considered as a unsupervised architecture for feature extraction. This method as an unsupervised feature extraction strategy has not been applied to this data for previously. Additionally, for visualisation of high dimensional data in lower dimension, three state-of-the-art methods will be employed. This will be performed in two scenarios. Firstly, the raw data will be transformed to a 2D space, which will be visualised with the three methods. Secondly, the CVAE encoded

data will be then transformed similarly, and visualised. The aim is to show the visualisation capabilities of CVAE purely for illustration purposes.

## 4.2 Methodology

The unsupervised framework proposed in this study is described in Figure 4.2. As observed, the first step is to manipulate the data by reshaping. Therefore, the data are divided to train and test considering the fact that leave-one-subject-out fashion is implemented. Since the study's nature is unsupervised, the number of clusters/classes is unknown. Then, a number of methods for number of clusters estimation are employed. The next step is to perform feature extraction with both proposed and state-of-the-art methods. The proposed methods include local DCT-based method and local entropy-based method, which are compared with the existing CVAE, CAE, PCA, and 2DPCA. The state-of-the-art CVAE for feature extraction has not been used previously for human activity recognition with Doppler radars.

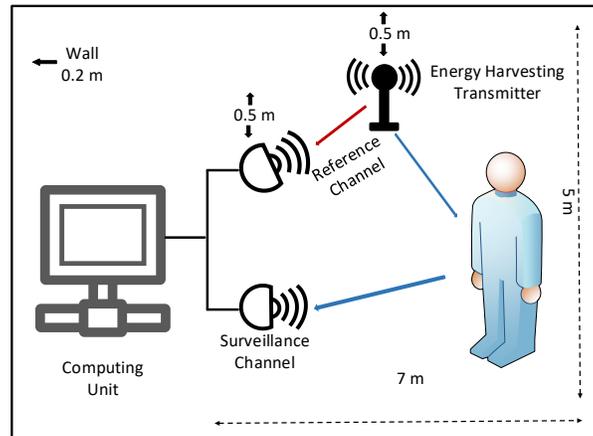


**Figure 4.2:** Flowchart, describing the overall analysis framework of the study.

### 4.2.1 Dataset Description

The micro-Doppler radar dataset *Doppler-Radar-2018* was collected independently by researchers at University of Bristol, UK. Figure 4.3 illustrates the experimental layout, where a  $5 \times 7$  metres room is observed. Energy Harvesting Transmitter (TX91501 POWERCASTER) radio source was used working on 915 MHz ISM band with 30dBm DSSS signal. The micro-

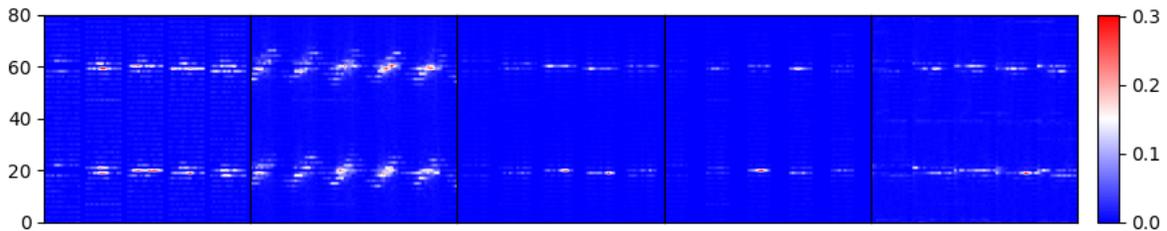
Doppler radar consisted of two channels: 1) surveillance channel; 2) reference channel. Both channels were synchronised using NI USRP 2920s and connected with directional antennas. As observed, the surveillance channel was pointed at the subject, while the reference channel was positioned 1 metre from the Energy Harvesting Transmitter. To obtain range-Doppler plots, a Cross-Ambiguity Function (CAF) was used. For further information, the readers are referred to [60].



**Figure 4.3:** Experimental layout for conducting the activities in the used Doppler radar dataset.

The dataset was created by four participants (one female and three male) in total with diverse body sizes. Each of the five activities, including (1) walking, (2) running, (3) jumping, (4) turning, and (5) standing, was performed 10 times by each participant. This resulted in 40 samples for each activity leading to a total of 200 samples.

The Doppler radar dataset used in this study was already pre-processed. More specifically, the total number of features per sample is 6400 (2 directions  $\times$  100 Doppler bins  $\times$  32 time index). In addition, the data is normalised referring to the fact that each feature's value is within the range of (0, 1). The vectorised data with 6400 features is therefore reshaped to 2D maps of size  $80 \times 80$ . This is implemented in order to obtain 2D images, which can be manipulated with image analysis strategies considering the local variations and patterns in each image. Figure 4.4 presents an example sample of each activity reshaped into 2D maps.



**Figure 4.4:** Example of an  $80 \times 80 = 6400$  image for each activity: (1) walking, (2) running, (3) jumping, (4) turning, and (5) standing.

The study is implemented in Python, where a number of libraries are used. For data pre-processing, pandas (version 1.0.5) and numpy (versions 1.19.1) are applied. The methods for number of clusters estimation with K-Means are implemented with scikit-learn (version 0.22.2). K-Medoids for clustering is applied using scikit-learn-extra (version 0.1.0b2). The deep learning architectures for CAE and CVAE are implemented and run with Tensorflow (version 2.2.0) and Keras (version 2.2.4). In order to implement visualisation results, matplotlib (version 3.2.2) functionalities are applied.

## 4.2.2 Number of Classes Estimation

The number of class/clusters in this study is unknown a priori considering the unsupervised scenario. To estimate the correct the number of clusters, a number of methods are applied in combination with K-Means including Elbow method, Silhouette analysis, Davies-Bouldin score, Dunn's index, and Calinski-Harabasz index.

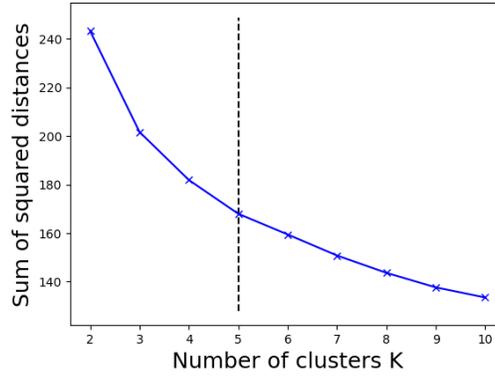
### 4.2.2.1 Elbow Method

The Elbow method is a visualisation technique aimed for finding the optimal number of clusters [167], [168]. The overall goal for the method is to minimize the intra-class variability and maximize the inter-class variability. In this work, the data samples are denoted as  $D = D_1, D_2, \dots, D_n$ , where  $n$  refers to the total number of samples. Additionally, the number of clusters is given by  $K$ , while the centroids are defined as  $\omega_1, \omega_2, \dots, \omega_K$ . The effectiveness of the method described by the distortion  $J$  is given by:

$$J(K, \omega) = \frac{1}{n} \sum_{i=1}^n \left( \min_{j=1}^K (D_i - \omega_j)^2 \right) \quad (4.1)$$

The Elbow method is implemented in combination with K-Means clustering, which is described later in this section. A candidate number of clusters  $K = 2, 3, \dots, 10$  are selected. Elbow method computes the sum of squared errors for the data samples in each cluster. The increase of  $K$  refers to the decrease of  $J$ . An optimal selection is where an additional increase in the number of clusters  $K$  does not change the squared sum of errors significantly. In other words, a further increase would result in over-clustering. The bend point is where the decrease in  $J$  is noticeable before reaching the actual  $K$  and smoother afterwards. Since Elbow method is a visualisation technique, the graph shows a significant decline before reaching the actual  $K$ . Therefore, with the further increase of the number of clusters  $K$ , the distortion  $J$  is very similar. Figure 4.5 shows the Elbow method for these data.

As illustrated  $K = 5$  is selected as the optimum number of clusters. This is the actual number of class in our data. However, Elbow method is a visualisation technique and the finding can be ambiguous in some cases. To strengthen the analysis, additional heuristics techniques for finding  $K$  are applied.



**Figure 4.5:** An Elbow test used to determine the number of clusters  $K$ .

#### 4.2.2.2 Silhouette Analysis

Silhouette analysis is probably the most popular method for estimating the number of clusters [113]. It is given by:

$$Silhouette = \frac{1}{n} \sum_{i=1}^n \frac{b(D_i) - a(D_i)}{\max\{a(D_i), b(D_i)\}} \quad (4.2)$$

where  $a(D_i)$  refers to the average distance between data sample  $D_i$  and all other data samples belonging to the same cluster. In addition,  $b(D_i)$  denoted the minimum average distance between data sample  $D_i$  and the remaining clusters. The Silhouette analysis aims to show the suitability of a data sample  $D_i$  to belong to a cluster. More specifically, it shows the “goodness” of clustering results and can serve as a parameter selection technique. The Silhouette score is within the range of  $(-1, 1)$ , where a lower values shows overlapping clusters, while the higher value refers to good clustering results.

#### 4.2.2.3 Davies-Bouldin Index

The Davies-Bouldin index is a method for estimating the number of clusters, which considers the worst clustering scenario. Therefore, unlike previous methods, a smaller value is desirable. The method is given by:

$$Davies = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} \frac{\Delta(C_i) + \Delta(C_j)}{\lambda(\omega_i, \omega_j)} \quad (4.3)$$

where  $\Delta(C_i) = \sum_{D_i \in C_i}^n \|D_i - \omega_i\|^2$  refers to the distance between each data sample  $D_i$  and the centroid of the corresponding cluster  $\omega_i$ . Therefore, the distance between the centroids of all clusters is given by  $\lambda(\omega_i, \omega_j) = \|\omega_i - \omega_j\|^2$ . The maximisation term computes the ratio of between-clusters for the  $i^{th}$  and  $j^{th}$  clusters. More specifically, all combinations of clusters are considered. The Davies-Bouldin method shows the clusters with minimum inter-cluster distance, where the data samples in each cluster have large spreads. This is the worst clustering scenario. As already specified, a smaller number of Davies-Bouldin method is desirable.

#### 4.2.2.4 Dunn's Index

Dunn's index is one of the oldest methods for estimating the number of clusters [115]. As with most methods concerned with clustering, the aim is to minimise the distance of data samples in each cluster, and maximise the distance between data samples from distinct clusters. The method is given by:

$$Dunn = \frac{\min_{1 \leq i < j \leq K} \lambda(\omega_i, \omega_j)}{\max_{1 \leq k \leq K} \Delta(C_k)} \quad (4.4)$$

where the distance between cluster centroids  $\omega_i$  and  $\omega_j$  is given by  $\lambda(\omega_i, \omega_j)$ . In addition, the distance between data samples belonging to the same class is given by  $\Delta(C_k)$ . Since the numerator seeks to find the distance between clusters, a maximum value is desired. On the other hand, a minimum value is desired for the denominator aiming to compute the intra-cluster distance. Similarly to Silhouette analysis, a higher value for Dunn's index is needed.

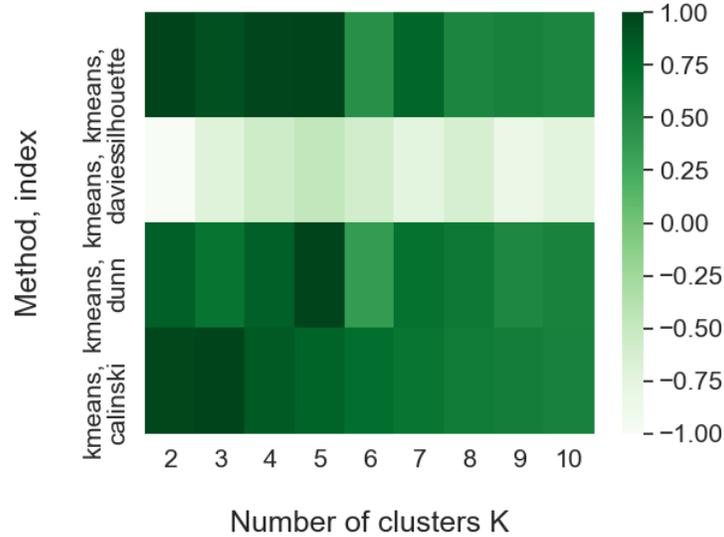
#### 4.2.2.5 Calinski-Harabasz Index

Calinski-Harabasz index is a popular method for estimating the number of clusters and clustering quality [116]. It computes the ratio of between-cluster dispersion and within-cluster dispersion. The specific formulation of the method, also known as Variance ratio criterion, is given as follows:

$$CH = \frac{B(K)}{W(K)} \times \frac{n - K}{K - 1} \quad (4.5)$$

where  $B(K) = \sum_{k=1}^K n_k \|\omega_k - \omega\|^2$ , while  $W(K) = \sum_{k=1}^K \sum_{i=1}^{n_k} \|D_i - \omega_k\|^2$ . The total number of points is denoted by  $n$ , while  $K$  refers to the total number of clusters. The global centroid is given by  $\omega$ , while  $\omega_k$  is the centroid of a cluster  $k$ . In line with the majority of clustering methods, a higher value is desirable for Calinski-Harabasz index.

Similarly to the Elbow method implementation, a candidate set of clusters is given  $K = 2, 3, \dots, 10$  for K-Means clustering. The heatmap in Figure 4.6 shows the results for each heuristic technique, where a darker colour refers to a higher value. As already discussed, the Davies-Bouldin index refers to the worst clustering scenario, where the lowest value shows the best clustering. Therefore, the inverse of this method is used in the graph in order to be in line with the other methods. The selected number of clusters is  $K = 5$  for Silhouette, Davies and Dunn's index, while the Silhouette coefficient is very similar for  $K = 4$  and  $K = 5$ . Only Calinski-Harabasz index shows  $K = 3$  as the most optimum number of clusters. Overall,  $K = 5$  is selected as the most optimum number of clusters by the majority of methods.



**Figure 4.6:** Silhouette score, Davies score, Dunn's index and Calinski-Harabasz index used to determine the number of clusters  $K$ .

### 4.2.3 Feature Extraction Methods

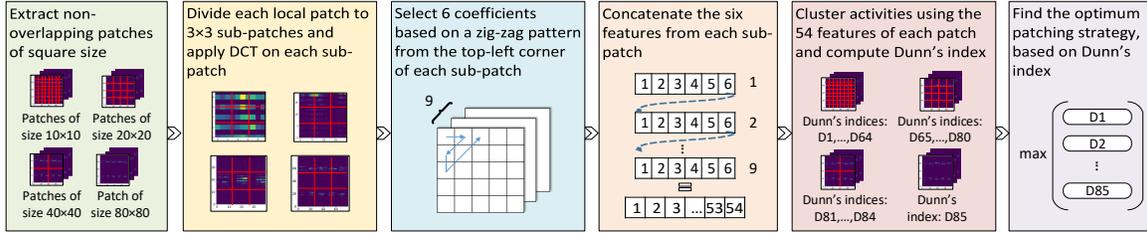
In order to extract features from the usually high-dimensional micro-Doppler radar data, researchers have used two main approaches. More specifically, physical feature extraction is commonly applied, which includes features such as average torso velocity, duration of an activity, upper and lower envelope variances [59, 169]. This information, also known as empirical information, can be very useful for finding the intuitive relation between Doppler samples and feature. However, datasets with unknown and uncontrolled conditions require different features such as treating the micro-Doppler samples as time-spectrogram and range-Doppler time points [170, 171, 172]. In this work, we will explore two novel local feature extraction approaches including local DCT-based method and local entropy-based method. The proposed methods will be compared with the existing convolutional-based methods such as CAE and CVAE. In addition, PCA and its variation 2DPCA will also be applied for comparison.

#### 4.2.3.1 The Proposed Local DCT-based Method

The first feature extraction method proposed in this study considers local areas from the original 2D maps  $80 \times 80$  from the Doppler radar data, where 2D DCT is applied [164]. The 2D DCT method has already been described in Section 3.4.1.2.3.

As observed earlier, some of the local areas in the 2D Doppler maps contain richer information opposed to other areas with little or no information. By using this notion, a systematic search algorithm can be applied in order to discover the most optimum patch. More specifically, non-overlapping patches of different square sizes are considered -  $10 \times 10$ ,  $20 \times 20$ ,  $40 \times 40$  including the original  $80 \times 80$  images as illustrated in Figure 4.7. Therefore, 2D DCT is applied to each local patch, which then divides the patch to  $3 \times 3$  sub-patches. The

following step is to select six coefficients in a zig-zag pattern from the top-left corner of each sub-patch. The resulting six features from each of the  $3 \times 3$  sub-patches are then concatenated. This results in only 54 features considering the extremely high dimensionality of  $80 \times 80$  original features. In order to assess the suitability of each strategy, Dunn's index is applied, where a higher value refers to better clustering results.



**Figure 4.7:** Flowchart of the proposed local DCT-based method.

Dunn's index is selected as an unsupervised model selection metric in this work. This is due to the fact that it represents an easily interpretable metric. As shown in equation (4.4), the numerator finds the minimum distance between cluster centroids for all combinations, while the denominator refers to the maximum intra-cluster distance. By using this model selection strategy, the unsupervised nature of the work is preserved.

#### 4.2.3.2 The Proposed Local Entropy-based Method

Considering the fact that the Doppler data are treated as  $80 \times 80$  images, texture analysis methods can be applied in order to interpret the patterns for different activities. Entropy is used in this study, which is a statistical measure based on randomness. It is defined by Shannon's equation [173] given by:

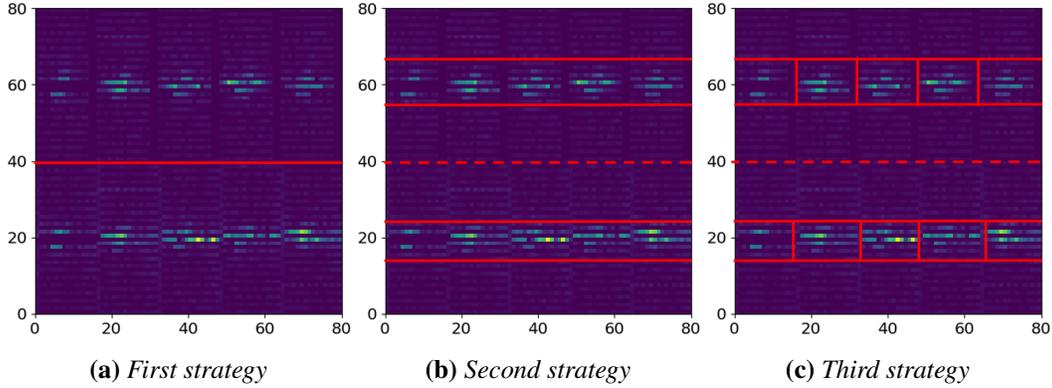
$$H(\rho) = - \sum_{i=1}^b \rho_i \log(\rho_i) \quad (4.6)$$

where  $\rho_i$  is the normalised histogram counts for each image, while  $b$  refers to the total number of histogram bins. More specifically,  $\rho_i$  is computed based on the histogram of each image.

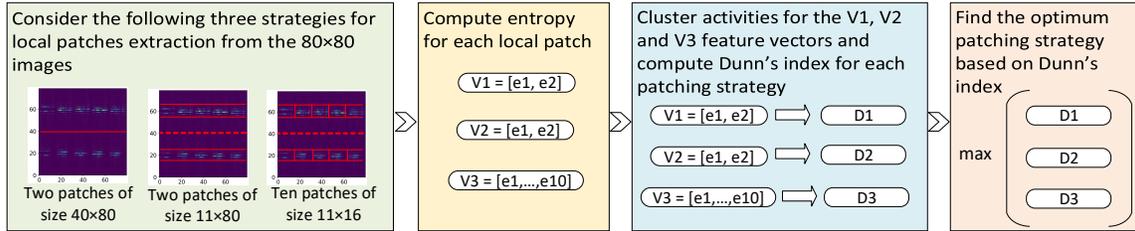
The entropy value changes based on the local variations in the image. If the local patch contains similar information with low variation, then the entropy value will be small. On the other hand, the entropy value increases with the increased variation in the images. Based on the notion used in the previous technique, some of the patterns in the 2D Doppler maps can be removed from the analysis. In addition, a careful selection of local patches can show good discriminative entropy.

Three local patching strategies are considered, which start with larger patches that are narrowed down systematically. The proposed local patches strategies are illustrated in Figure 4.8. Similarly to the previous method, the unsupervised metric Dunn's index is selected for

estimating the goodness of clustering. The steps of the proposed local entropy-based method are outlined in Figure 4.9.



**Figure 4.8:** The three patching strategies for entropy analysis.



**Figure 4.9:** Flowchart of the proposed local entropy-based method.

### 4.2.3.3 Convolution Filter-based Methods

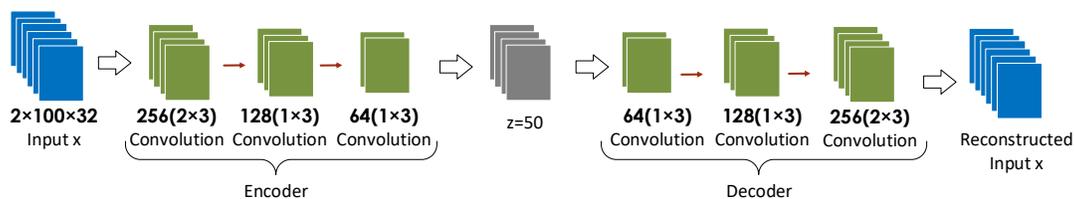
This section outlines the convolutional-based method for feature extraction including CAE and CVAE. More specifically, the drawbacks of CAE are described, which naturally leads to the improved CVAE architecture.

**4.2.3.3.1 Convolutional Autoencoder (CAE)** Autoencoders (AEs) are unsupervised neural networks commonly used for feature extraction. In addition to their applications, popular areas of investigation include data denoising [174], anomaly detection [175], and image generation [176]. The structure of AEs consists of an encoder and a decoder. The input of the encoder is the original image  $x$ , which is then transformed to a latent information  $z$  with lower dimensionality. On the other hand, the role of the decoder is to reconstruct  $x$  given the limited information of  $z$ . A reconstruction loss is only included in the implementation of AEs, which guides the decoder in the correct reconstruction of  $x$ .

The disadvantages of CAE refer to the fact that the architecture fails to obtain any statistical strength between all data samples. More specifically, it learns local parameters of each data point individually. The inclusion of a reconstruction loss only usually leads to mean-

ingless representations of similar data samples. As such, these data samples belonging to the same class may be given very distinct representations in the Euclidean space.

A deep CAE with three hidden layers for both the encoder and decoder is used in this work as shown in Figure 4.10. A three-dimensional input of  $2 \times 100 \times 32$  is used for the Doppler samples. It is important to note that the structure of the encoder is symmetric to the structure of the decoder. The first convolutional layer in the encoder and its corresponding third convolutional layer in the decoder have 256 filters with size of  $2 \times 3$ . In addition, the second convolutional layers in both encoder and decoder have 128 filters with size of  $1 \times 3$ . Finally, the third convolutional layer in the encoder and the first convolutional layer in the decoder both have 64 filters with size of  $1 \times 3$ . A ReLU activation function is used for the layers in both structures. The latent information, which is the output of the encoder  $z$ , contains only 50 features, which is an extreme decrease in the dimensionality. As already described, the decoder aims to reconstruct  $x$  given  $z$ . For the purpose of feature extraction, the encoded features are used. In addition, the robustness of the deep architecture is performed in Section 4.3.3.



**Figure 4.10:** *The CAE architecture.*

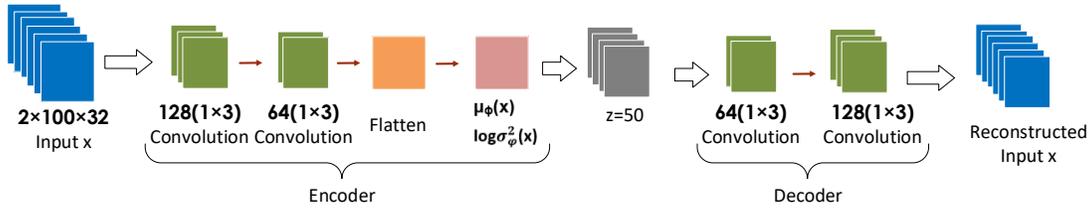
**4.2.3.3.2 Convolutional Variational Autoencoder (CVAE)** CVAEs are generative models defined recently [177], which have found applications in feature extraction [178], data augmentation [179], reinforcement learning [180], and others. To begin with, the encoder is also referred as recognition model or inference model, while the other name for the decoder is a generative model. Similarly to CAE, the input of the encoder  $x$  is transformed to the latent information  $z$ . Then, the tasks for the decoder is to reconstruct  $x$  given  $z$ . In addition to the implementation of CAE, each feature in the latent representation is treated as a single-valued output. To improve this constraint, the CVAE architecture considers a probability distribution for each feature, which is commonly a normal distribution also known as Gaussian distribution. As already specified, the use of only a reconstruction loss in the CAE implementation may produce meaningless content. CVAE includes a regularisation term, which ensures that the probability of the encoded space is close to standard normal distribution, which has  $\mu = 0$  and  $\sigma = 1$ . This is developed with Kullback-Leibler divergence. The implementation of CVAE is given by:

$$l_i(\theta, \phi) = -E_{z \sim q_\theta(z|x_i)}[\log p_\phi(x_i|z)] + KL(q_\theta(z|x_i)||p(z)) \quad (4.7)$$

where the first term describes the reconstruction loss, while the regularisation term is given as the second term. More specifically,  $p_\phi(x_i|z)$  describes the generative probability of the

reconstructed output  $x_i$  given the latent information  $z$ . Then, the distribution of  $z$  given  $x$  is given by  $q_\theta(z|x_i)$ . The parameters of the distributions are given as  $\theta$  and  $\phi$ .

The architecture of the used CVAE model is outlined in Figure 4.11. The three-dimensional Doppler radar data are used as an input of the encoder. Similarly to the CAE architecture, the encoder and the decoder are symmetric. In addition to the similarities of both deep architectures, ReLU activation functions are used for the convolutional layers. In terms of the illustrated CVAE architecture, two convolutional layers are used for both encoder and decoder structures. More specifically, the first convolutional layer in the encoder and its corresponding second convolutional layer in the decoder contain 128 with size of  $1 \times 3$  referring to height and width respectively. Then, the second convolutional layer in the encoder and its corresponding first convolutional layer in the decoder have 64 filters with the size of  $1 \times 3$ . The regularisation term in the CVAE architecture ensures that the distribution of the encoded data is close to standard normal distribution with parameters mean  $\mu_\phi$  and variance  $\log \sigma_\phi^2$ . Instead of outputting the direct latent information, the encoder outputs the mean and variance of the information. The size of the latent representation  $z$  is again 50, which is used by the decoder in order to reconstruct  $x$ .



**Figure 4.11:** The CVAE architecture.

#### 4.2.3.4 Variation-based Projection Techniques

**4.2.3.4.1 Principal Components Analysis (PCA)** 1D-PCA is one of the most popular methods for linear feature extraction [181], which was briefly described in Section 3.3.2.1.1. The method aims to find the directions of the main variations of the data and project these variations to a smaller dimensional space. More specifically, 1D-PCA is aimed at computing the covariance matrix of the centred data. Therefore, eigendecomposition of the covariance matrix is applied, which results in principal components. It is known that the first principal component (PC1) preserves the most variance of the data, which allows a smaller number of PCs to be selected. The overall goal is to select PCs preserving 95% of the data variance. Therefore, the original data is transformed to a smaller space  $D_{n \times m} W_{m \times s} = Z_{n \times s}$ , where  $s$  refers to the number of PCs.

**4.2.3.4.2 2D Principal Components Analysis (2DPCA)** The discussed method 1D-PCA requires the data to be projected to a 1D vector. This can often lead to a high-dimensional vector with a large covariance matrix. In order to preserve the originality of 2D data, 2D-

PCA is proposed [182], which allows for a more accurate calculation of the covariance matrix. Therefore, instead of vectorising the 2D Doppler images, the  $80 \times 80$  dimensionality is used. This also leads to a faster computational time, which is an additional advantage. The covariance matrix is calculated as follows:

$$V = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^T (X_i - \bar{X}) \quad (4.8)$$

where  $n$  refers to the total number of training samples, while  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  is the average training image among all samples. The size of the average training image is also  $80 \times 80$ .

2DPCA usually computes the covariance matrix only for the row or column dimension only. In this work, the  $80 \times 80$  dimensionality is considered, where only the columns are used for the covariance matrix computation. As described above, eigendecomposition is then applied on the covariance matrix. Therefore, a limited number of PCs is selected preserving 95% of the data variance.

## 4.2.4 Clustering Methods

Two clustering methods are used for grouping the data samples. Both of the methods are used by computing distances: K-Means and K-Medoids.

### 4.2.4.1 K-Means

K-Means clustering is probably the most common method for unsupervised learning clustering [183]. An important constraint of the method is the consideration of the number of clusters  $K$  is known a priori. Then, the method aims to group the data samples  $D_1, D_2, \dots, D_n$  based on minimisation of intra-class variability and maximisation of inter-cluster variability. The method is distance-based, while Euclidean distances are used in this study. In addition, a new data sample is assigned to the already specified clusters by computing the Euclidean distance between the sample and the corresponding cluster centroids  $\omega_1, \omega_2, \dots, \omega_k$ . Therefore, the data sample is assigned to the cluster with the smallest distance. K-Means clustering objective function is given by:

$$J = \sum_{i=1}^n \sum_{k=1}^K r_{ik} \|D_i - \omega_k\|^2 \quad (4.9)$$

where  $r_{ik}$  is a binary function, which is 1 if a data sample  $D_i$  is assigned to cluster  $k$  with centroid  $\omega_k$ . It is 0 otherwise:

$$r_{ik} = \begin{cases} 1, & \text{if } k = \arg \min_j \|D_i - \omega_j\|^2 \\ 0, & \text{otherwise} \end{cases} \quad (4.10)$$

The aim of the method is to minimize the objective function  $J$  for  $r_{ik}$  and  $\omega_k$  by iterative optimization. In the first step,  $\omega_k$  is fixed, while the focus is on optimising  $r_{ik}$ . In the second

step, the roles are reversed with the same intuition. The whole process is repeated until there is no further re-assignment of data samples in the defined clusters.

#### 4.2.4.2 K-Medoids

K-Medoids is a clustering method, which is usually concerned with finding dissimilarities instead of calculating distances [83]. Therefore, it has shown improved results for noisy and problematic data in comparison with the distance-based K-Means clustering. However, the method can also be used with distances, which is the case for this study. More specifically, Manhattan distances are used.

Unlike K-Means clustering, K-Medoids considers a data sample  $D_i$  for the cluster centroid. The first step of the method aims to find such data sample in a cluster  $C(i) = k$  that is in minimum distance to the remaining data points in the cluster  $D_{i'}$ . The Manhattan distance is given by  $\|D_i - D_{i'}\|$ , while the minimisation is as follows:

$$i_k^* = \arg \min_{i:C(i)=k} \sum_{C_{i'}=k}^n \|D_i - D_{i'}\| \quad (4.11)$$

In order to find a new centroid or medoid, the output of  $i_k^*$  is used considering  $\omega_k = D_{i_k^*}$ ,  $k = 1, 2, \dots, K$ .

The second step aims to re-assign each data sample  $D_i$  to the closest centroid or medoid by calculating Manhattan distances. Similarly, the centroids are given by  $\omega_1, \omega_2, \dots, \omega_K$ .

$$C(i) = \arg \min_{1 \leq k \leq K} \|D_i - \omega_k\| \quad (4.12)$$

The K-Medoids clustering iterates the two steps until no further re-assignment of data samples  $D_1, D_2, \dots, D_n$  to centroids  $\omega_1, \omega_2, \dots, \omega_k$  is observed.

### 4.2.5 Visualisation Techniques for High-Dimensional Data

The following section outlines three common manifold learning methods for high-dimensional data visualisation by projecting it to 2- or 3-dimensional space. More specifically, these manifold learning techniques are known to preserve the original distances in the high-dimensional space when projected to a much lower dimensionality. Hence, their similarity patterns are preserved, which improves visualisation. As described in Section 4.1, the methods will be used in two scenarios. Therefore, important insights for mapping of individual classes of the dataset can be extracted.

#### 4.2.5.1 t-Distributed Stochastic Neighbour Embedding (t-SNE)

T-SNE is a popular manifold learning method for non-linear dimensionality reduction, which has been proposed recently [184]. The method has quickly gained attention for its superior

ability for visualising high-dimensional data by projecting it to a 2- or 3-dimensional space. In addition, t-SNE aims to preserve the similarity patterns from the high-dimensional space into the reduced space, which is not the case for common feature extraction techniques such as PCA and SVD.

T-SNE begins with analysing the Doppler sequences  $D_1, D_2, \dots, D_n$  in their original high-dimensional spaces. The Euclidean distances between the data samples are computed, which are later converted to probabilities with normal distribution. Similarity score for each data points pair  $D_{ij}$  is computed, which is obtained based on probabilities  $p_{ij}$ . Therefore, the data samples are initially projected randomly to the lower dimensional space.

The aim for t-SNE is to preserve the similarities in the high-dimensional space to the lower dimensionality. In order to achieve this, similarities based on Euclidean distances are computed similarly. Again, they are converted to probabilities with t-Distribution  $q_{ij}$ , which has higher tails. The higher tails avoid dissimilar data samples to be positioned in closed locations in the Euclidean space. The probabilities of data samples in the high-dimensional space  $p_{ij}$ , and the corresponding low-dimensional space  $q_{ij}$  are then optimised with KL divergence. The KL divergence iteratively compares the distances in the high-dimensional and low-dimensional space, which allows the data samples to be re-positioned accordingly.

#### 4.2.5.2 MultiDimensional Scaling (MDS)

MDS is one of the older non-linear methods for dimensionality reduction [83, 185, 186]. Unlike other manifold learning techniques, MDS is concerned with dissimilarities instead of similarities. More specifically, the method calculates the Euclidean distances between the data samples  $D_1, D_2, \dots, D_n \in \mathbb{R}^m$ , while  $d_{ij}$  denotes the dissimilarities of each pair of two samples  $D_i$  and  $D_j$ . Therefore, MDS is focused on preserving the distances  $z_1, z_2, \dots, z_n \in \mathbb{R}^k$  in the lower dimensionality  $k$  by ensuring  $k < m$ . The defined stress function is given by:

$$Stress_{(z_1, \dots, z_n)} = \sum_{i \neq j}^n (d_{ij} - \|z_i - z_j\|)^2 \quad (4.13)$$

where the Euclidean distance between the samples in the lower dimension  $z_i$  and  $z_j$  is given by  $\|z_i - z_j\|$ . The MDS method aims to preserve the distances in the low dimensional space, so the focus is on minimising the so-called stress function. This is achieved by using a gradient descent algorithm. Similarly to t-SNE, close samples in the original space are projected in close locations in the lower space, while the same notion is used for dissimilar data samples.

#### 4.2.5.3 Locally Linear Embedding (LLE)

LLE is a non-linear method for dimensionality reduction, which uses geometric principles [187]. The LLE method assumes that the data samples are positioned on an underlying manifold represented by vectors. More specifically, each vector is denoted by  $\vec{X}_i$  with index

$i$  in the high dimension  $m$ , while it has a number of so-called neighbours. Considering the fact that each vector is a member of a neighbourhood, the similar data samples are expected to lie on a close locally linear patch of the manifold. There exist  $p$  neighbours in each neighbourhood. In addition, each vector can be reconstructed by its neighbours, which is given by:

$$\epsilon(W) = \sum_{i=1}^n \left\| \vec{X}_i - \sum_{j=1}^p W_{ij} \vec{X}_j \right\|^2 \quad (4.14)$$

where  $W_{ij}$  refers to weights defined for the reconstruction of  $\vec{X}_i$  based on  $\vec{X}_j$ . More specifically, the weights show the contribution of a data point for reconstructing another data point in the same neighbourhood. The number of samples is given by  $n$ , while  $p$  denotes the number of neighbours. There exist two main constraints for the weights function. It is 0, if the two data points do not belong to the same neighbourhood. This ensures that each data point is reconstructed only by its neighbours. To make the weights invariant to rotation, and scaling, the weight matrix's rows should be one or  $\sum_{j=1}^p W_{ij} = 1$ .

By having the weights  $W_{ij}$  fixed, the original high-dimensional data points  $\vec{X}_i$  can be projected to a lower dimension, where the points are denoted by  $\vec{Y}_i$ . This is achieved by minimising an embedded cost function defined as:

$$\Phi(Y) = \sum_{i=1}^n \left\| \vec{Y}_i - \sum_{j=1}^p W_{ij} \vec{Y}_j \right\|^2 \quad (4.15)$$

### 4.3 Evaluation and Results

This section presents the clustering results obtained using K-Means and K-Medoids based on the projected data. As described, the projected data are achieved using a number of feature extraction methods, which are compared. These include the proposed local DCT-based method and local entropy-based method. The proposed techniques are then compared with existing convolutional-based CAE and CVAE, and variational-based PCA and 2DPCA.

The two proposed local methods and 2DPCA take the 2D inputs of  $80 \times 80$ . The CAE and CVAE architectures consider the  $2 \times 100 \times 32$  inputs, while for 1D-PCA, the data are vectorised. As already specified, the selected eigen vectors for 1D-PCA and 2DPCA preserve 95% of the data variance.

To strengthen the analysis and avoid over-fitting, leave-one-subject-out CV (LOOCV) is applied. More specifically, the data from three subjects are used for training, while the remaining data from one subject are used for clustering. This is repeated for all four combinations, where the average and standard deviation of the clustering are outlined. Hence,

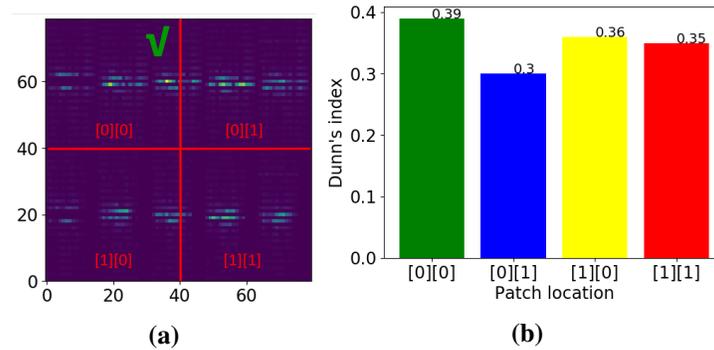
$Z_{tr} = 150 \times m$  defines the training matrix, while the testing matrix is given by  $Z_{ts} = 50 \times m$ , where  $m$  denotes the number of features, which are different for each feature extraction method.

The three manifold learning methods t-SNE, MDS, and LLE are applied in two scenarios for the sake of visualisation results. In the first scenario, the methods transform the original raw data with 6400 feature to a 2-dimensional space. Then, the projected data are visualised. In the second scenario, the CVAE encoded features are projected to a 2-dimensional space. Similarly, the projected data are then visualised.

The true labels are available in this study. However, they are only used for model evaluation and mapping the samples in the visualisation graphs. Considering the fact that the predicted labels by clustering do not follow the original order, a mapping algorithm is applied. The predicted labels are then matched to the corresponding actual true labels by finding the best-matching pairs of clusters labels and true labels.

### 4.3.1 Local DCT-based Analysis Results

As already described, the proposed local DCT-based method extracts non-overlapping square-sized patches from the original  $80 \times 80$  images, while the original large patch is also used. To preserve the unsupervised scenario of the study, Dunn's index is used as an unsupervised metric for evaluating the goodness of the clustering results. The highest values for Dunn's index correspond to the  $40 \times 40$  patches, which are four non-overlapping patches. These are illustrated in Figure 4.12a. By applying 2D DCT to each patch, it is then divided to  $3 \times 3$  sub-patches. The aim is to extract six coefficients in a zig-zag order from the top-left corner of each sub-patch, where the most information is preserved. Therefore, K-Means is used in combination with Dunn's index to find the most optimum local patch. As observed in Figure 4.12b, the top-left  $40 \times 40$  local patch is selected as the most optimum patch.



**Figure 4.12:** (a) Illustration of the  $40 \times 40$  local patches in the Doppler spectrogram. The optimum patch for feature extraction is ticked in green color. (b) The corresponding Dunn's index for the four  $40 \times 40$  local patches.

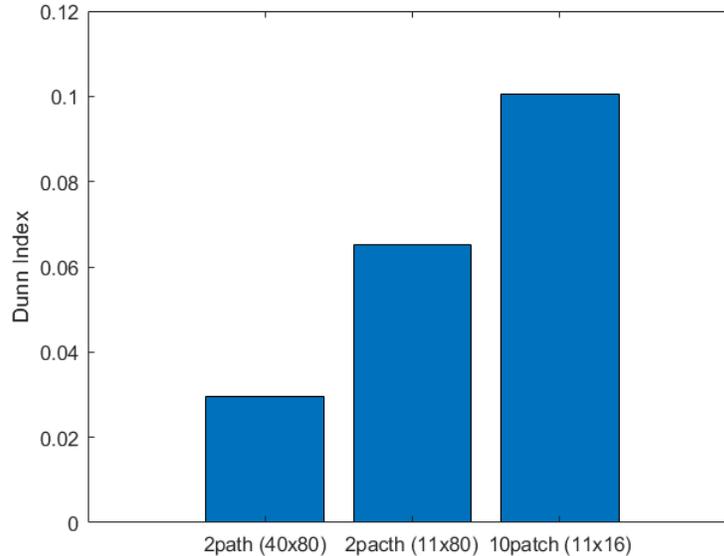
To strengthen the findings, 2D DCT is applied on the original  $80 \times 80$  image by preserving exactly 54 features as it was the case with the local patching strategy. 4-subjects LOOCV is applied with average and standard deviation as shown in Table 4.1. The results reveal that the proposed method improves the clustering performance by more than 10%.

**Table 4.1:** Average and standard deviations of testing accuracies based on K-Means and K-Medoids using DCT features from the raw data  $80 \times 80$  single patch and those from the selected  $40 \times 40$  local patch features over 4 rounds of LOOCV.

	DCT Raw Data	DCT-Based Method
K-Means	63.5%±8.64	75%±5.74
K-Medoids	62%±9.89	77%±4.58

### 4.3.2 Local Entropy Analysis Results

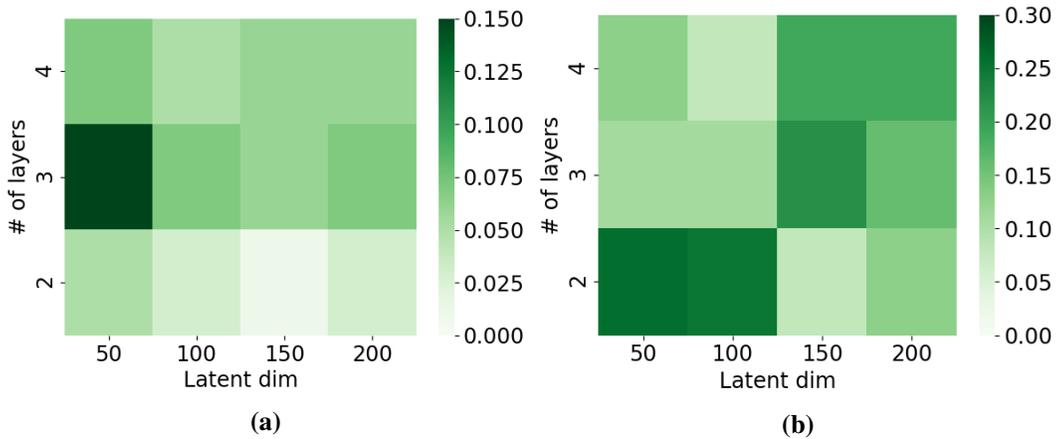
The entropy analysis considers three strategies as described earlier in Figure 4.8. The first strategy and the second strategy result both in two-dimensional features. On the other hand, the third strategy results in 10-dimensional features. Similarly to the local DCT-based method, Dunn's index is used for unsupervised clustering evaluation. As observed in Figure 4.13, the third strategy obtains the highest results. Unsurprisingly, the winning strategy considers the most features. The other two strategies do not represent the variations of the data as well as the third strategy. The clustering results of the local entropy-based method will be shown in Table 4.2 and Table 4.3 for training and testing respectively.



**Figure 4.13:** Illustration of the Dunn's index for the three patching strategies used for feature extraction based on local entropy analysis.

### 4.3.3 CAE and CVAE Robustness Evaluation

In order to evaluate the robustness of the two convolutional-based architecture CAE and CVAE, two criteria are addressed including the number hidden layers, and the number of extracted features  $z$  as latent dimension. Two, three, and four convolutional layers for the both the encoder and decoder are considered. The consideration of the number of projected features is in line with the total number of data samples, which is 200. Therefore, 50, 100, 150, and 200 features are used for evaluation. Dunn's index is applied similarly considering the fact that the study is unsupervised. Figure 4.14a and Figure 4.14b show the Dunn's index for the CAE and CVAE architectures respectively.



**Figure 4.14:** (a) CAE and (b) CVAE robustness evaluation using the Dunn's index based on average testing results.

Based on these results, three convolutional layers with 50 features are selected for the CAE architecture. In terms of the CVAE architecture, again 50 features are selected, while two convolutional layers are retained.

### 4.3.4 Comparison of the Average Training and Testing Accuracies for K-Means and K-Medoids Using all Feature Extraction Techniques

To evaluate the different feature extraction strategies, K-Means and K-Medoids clustering are applied. The 4-subjects LOOCV pattern is applied, where the average with standard deviation are reported in Table 4.2 and Table 4.3 for training and testing respectively.

**Table 4.2:** Average training accuracies with standard deviations of K-Means and K-Medoids using the five feature extraction methods over 4-subjects LOOCV.

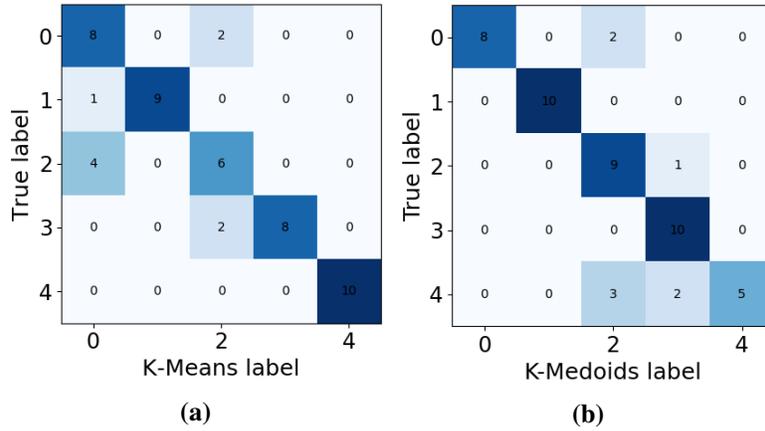
	DCT-Based Method	Entropy-Based Method	CVAE	CAE	PCA	2DPCA
K-Means	80%±4.41	69.5%±3.23	83.75%±3.11	52%±4.74	64.25%±7.32	67%±10.29
K-Medoids	77.5%±8.87	69.5%±4.05	79.75%±3.11	54.5%±4.55	47%±4.41	64.5%±5.93

**Table 4.3:** Average testing accuracies with standard deviations of K-Means and K-Medoids using the five feature extraction methods over 4-subjects LOOCV.

	DCT-Based Method	Entropy-Based Method	CVAE	CAE	PCA	2DPCA
K-Means	75%±5.74	72%±7.11	84%±5.09	66%±8.6	58.5%±1.65	57.5%±2.59
K-Medoids	77%±4.58	72%±8.48	82%±4.89	65%±5.74	57%±2.23	64%±8.12

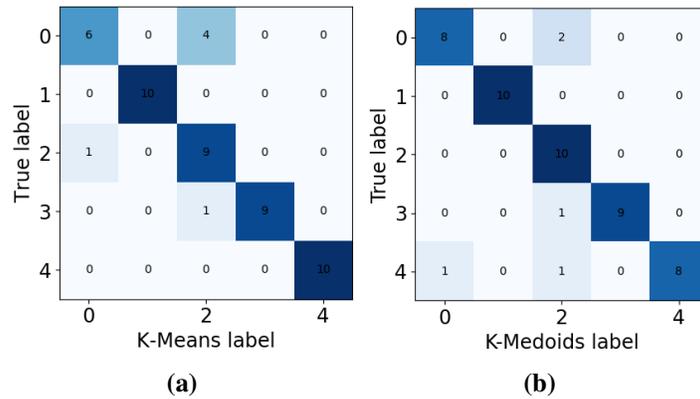
The reported results show that the superior architecture for feature extraction is the convolution filter-based CVAE. In addition, the local DCT-based method shows reliable results for an unsupervised architecture. More specifically, the method has been validated with Dunn’s index for the scope of the study. The other proposed technique for feature extraction, local entropy-based method, provides an improvement to CAE, PCA, and 2DPCA. Considering the variation-based PCA and 2DPCA, the latter shows a minor improvement to the former only for K-Medoids. To conclude the findings, as CVAE is a deep NN architecture, the training time is significantly more expensive compared to the simpler local DCT-based method. This will be evaluated later in Chapter 6, where the advantages and disadvantages of CVAE and the proposed methods will be discussed.

To discover the presence of any problematic classes, confusion matrices are used, which are supervised techniques. The proposed local DCT-based method along with CVAE are used with K-Means and K-Medoids. Figure 4.15a and Figure 4.15b reveal the confusion matrices for the local DCT-based method for K-Means and K-Medoids respectively.



**Figure 4.15:** Confusion matrices for DCT features using (a) K-Means, and (b) K-Medoids for different subjects.

Then, Figure 4.16a and Figure 4.16b show the confusion matrices for CVAE using K-Means and K-Medoids respectively.

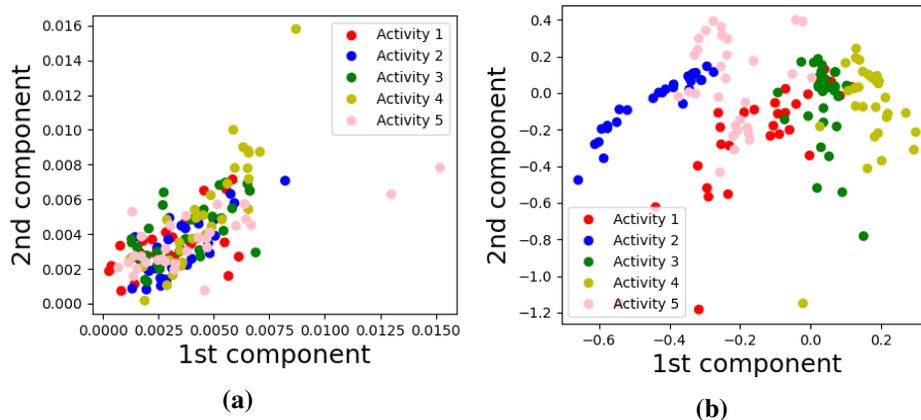


**Figure 4.16:** Confusion matrices for CVAE encoded features using (a) K-Means and (b) K-Medoids for different subjects.

The confusion matrices for both feature extraction methods show that walking (1) and jumping (3) are problematic as they are often mis-classified. In addition, the fifth class standing is also frequently confused with other classes for K-Medoids. To address the issue with misclassification, more data can be collected from diverse subjects. Even in cases with richer data captured by diverse subjects, the unsupervised scenario of the project still can influence the accuracy of the clustering results in the absence of labels.

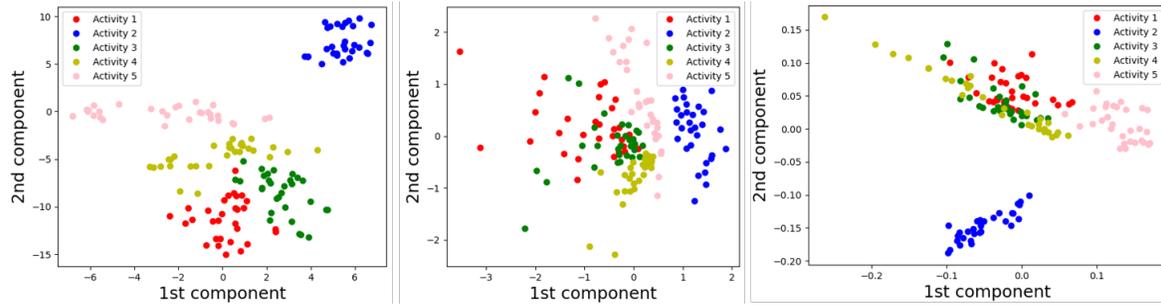
### 4.3.5 Visualisation Results

The first two components of the extremely high-dimensional data  $D_{n \times 6400}$  are initially visualised, where the labels are used for mapping the samples and distinguishing the activities. As observed in Figure 4.17a, there is a mismatch between the samples of the five classes. In addition, the first two components of the CVAE encoded features  $D_{n \times 50}$  are also visualised in Figure 4.17b. The mismatch between the samples has been improved, but still, some of the samples are overlapping. This is aimed to be improved in the following two scenarios.



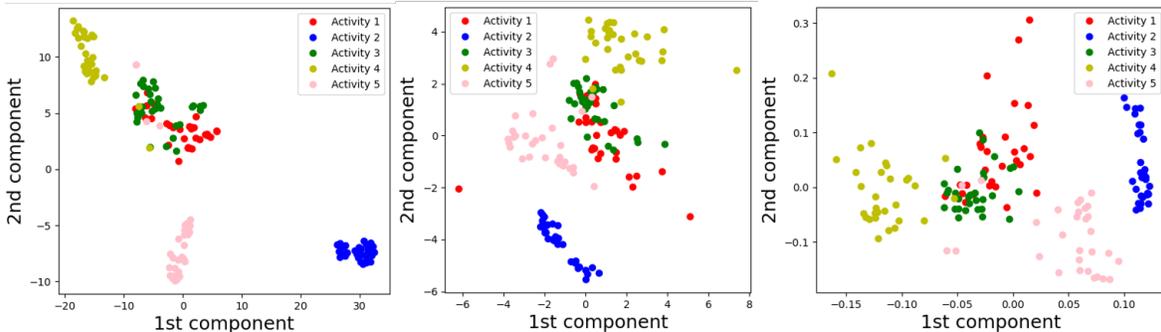
**Figure 4.17:** Visualisation of the first two components of (a) the raw data, and (b) the first components of the CVAE encoded data.

The first scenario of visualisation considers the original data  $D_{n \times m}$ , where  $m = 6400$ . The extremely high-dimensionality is transformed to a 2-dimensional space using the manifold learning method t-SNE, MDS, and LLE. As observed in Figure 4.18, t-SNE proves its superiority by providing the best separability between the classes. In the cases of MDS, and LLE, over-lapping between data samples from different classes are present.



**Figure 4.18:** Visualisation of the transformed raw data  $D_{n \times 6400}$  using t-SNE, MDS, and LLE.

The second scenario considers the CVAE encoded data. Since CVAE is known to map close samples in close locations, a comparison can be achieved. Similarly, t-SNE, MDS, and LLE are used for projecting the CVAE encoded data to a 2-dimensional space. As observed in Figure 4.19, t-SNE is still superior, while the remaining two methods have been improved. More specifically, the intra-cluster distances have been decreased, which is desired. However, no accurate clustering can be expected using the resulting 2-dimensional features from t-SNE. K-Means and K-Medoids provide 42% and 45% average testing accuracies respectively. It can be concluded that these methods are good for visualisation, but cannot be used for clustering.



**Figure 4.19:** Visualisation of the transformed encoded data  $D_{n \times 50}$  using t-SNE, MDS, and LLE.

## 4.4 Chapter Conclusion

This chapter has explored the novel idea of using unsupervised learning for Doppler radar based human activity recognition. Two main contributions were proposed: 1) a unsupervised local DCT-based feature extraction method for 2D Doppler maps; 2) a unsupervised local entropy-based feature extraction method for 2D Doppler maps. Additionally, the chapter included a novel application of CVAE for feature extraction of Doppler radar data. Application of a wide range of methods for estimating the number of clusters  $K$  was included. Three visualisation methods were compared for illustration purposes of raw data and CVAE encoded data.

All three methods provided reasonable recognition performances in the sense of unsupervised learning. More specifically, the methods outperformed state-of-the-art techniques such as CAE, PCA, and 2DPCA. Despite the superiority of the CVAE architecture for feature extraction, the computational time for training the model is more expensive in comparison with the two proposed methods.

The following chapter investigates the user acceptance of (1) sensors, and (2) robots among older adults living in a care home. Sensors are investigated from the perspective of improved health and well-being. On the other hand, robots are evaluated in terms of their physical appearance, characteristics, and functionalities. In addition, a comparison between the user acceptance of these two types of innovative technologies is achieved.

# Chapter 5

## User Acceptance of Sensors and Robots

### 5.1 Introduction

This chapter will address the problem of health monitoring of aged population using sensing technologies and the relevant robotic systems from a different perspective. It is mainly focused on the deployment of sensors and robots in such cases. For this aim, the user acceptance including preferences and needs of older adults for using such systems is studied. This is different from the previous two chapters, where different technical aspects of using non-intrusive sensing technologies for human activity monitoring have been studied.

The most recent research identifies negative preconceptions as one of the leading barriers for positive user acceptance [133, 134]. Despite an increasing push for such technologies in the context of elderly care [131, 132], relatively little attention has been paid to understanding factors influencing acceptance by the end-users, in this case an increasingly older, frail population living in care homes.

This study also covers an extension to the popular TAM [188]. TAM is an information technology framework, which studies the acceptance of any technology by people in their workplaces. In its originality, the TAM framework considers that people have the likelihood to accept any technology purely based on its (1) perceived usefulness, and (2) perceived ease-of-use. Considering the complex user acceptance of older adults, these two factors can be seen as limited.

The majority of studies exploring the user acceptance of sensors and robots with a physical presence among older populations are based on evaluations of a specific agent following interaction [30, 31, 32, 33, 34, 35, 39, 189, 190]. Hence, these studies can be regarded as limited considering the fact that they evaluate a specific agent only. As such, exploration of a diverse range of sensor systems and robots with a physical presence is not present. In addition, comparison between the user acceptance of sensors and robots among the older population is not evident in the literature. As described, the most influential acceptance model in the literature TAM only covers two proposed factors. As such, these factors can

be viewed as limited and further exploration is needed. Current studies also lack perceived health improvement analysis based on the two types of technology. In addition, age-based groups among the extremely older population are not distinguished, and therefore, the studies do not include the analysis of responses.

The following study will investigate the views and attitudes of older adults living in Long-Term Care (LTC) facility in the UK towards non-intrusive sensor systems and robots with a physical presence. More specifically, a questionnaire format is employed in the LTC settings, which covers a wide range of commercially available robots with a physical presence. In terms of sensor systems, a non-intrusive LRIR sensor is illustrated and presented as an example, which is identical to the used system in Chapter 3. The survey aims to investigate the perceived health improvement based on the two types of agents including both physical health, and the less explored among this population, mental health. Considering only robots with a physical presence, specific factors are explored such as physical appearance, activities, and characteristics. In regards to robot's characteristics, the described TAM framework is enriched to cover more aspects that may influence the user acceptance.

## **5.2 Methodology**

### **5.2.1 Study Aims**

The current study is focused on evaluating the views and attitudes of older adults unaffected by prior experience with a specific agent. As such, the study ensured that the selected participants lacked any previous interaction with the researched intelligent agents. This important factor was incorporated to reduce any potential bias, which can influence the study's results. Specific categories related to user acceptance of both discussed agents were identified as important: general acceptance of a non-intrusive monitoring technology and robots with a physical presence, robot's physical appearance distinguished in three categories, robot's roles including social role and physical role, perceived improvement of health.

### **5.2.2 Survey Design**

A questionnaire was designed to address the aforementioned factors (see Appendix A). The questionnaire contained illustrative examples to aid understanding. A specific example included a non-intrusive wall-mounted sensor, which can detect subject's activities including falls. In addition, commercially available robots based on different physical appearance such as humanoid, zoomorphic, and mechanoid were illustrated. The questionnaire was written in English language, as this was the native language of all participants. The majority of questions adopted the 5-point Likert scale answers (definitely yes, probably yes, might or might not, probably no, definitely no) or the 7-point Likert scale answers (strongly agree, agree, somewhat agree, neither agree nor disagree, somewhat disagree, disagree, strongly disagree), where the participants were asked to show a level of agreement. The final ques-

tion was an open-ended question, which allowed the participants to leave notes for requested robot's activities, not included in the previous questions.

The study strictly followed the ethical standards for working with vulnerable people. Ethical approval for the study was granted by the Health and Life Sciences Research Ethics Committee, University of Coventry (Project P65532).

### 5.2.3 Participants and Data Collection

The study was employed in a 84-bed residential care home owned by WCS Care in Kenilworth, England (<https://www.wcs-care.co.uk/our-homes/kenilworth-castle-brook>). The requirement for inclusion in the study was the absence of neurological diseases such as dementia, epilepsy, Parkinson's disease, and MND, which were discussed in detail in Chapter 1. A total of 24 healthy individuals were identified by the care home manager for participation in the study. All of the listed potential participants were approached in their living areas for informed consent. The final number of residents who agreed to participate was 21. Despite the fact that the study sample is limited, the current literature includes similar small to medium studies with interesting findings, which generated discussion [25, 26, 27, 28, 29, 30, 31, 32, 34, 39]. Anonymity of the participants was ensured by collecting only age (instead of date of birth) and gender statistics.

The questionnaire was administered among the participants who signed an informed consent over the period of September 2019 - October 2019 at WCS Castle Brook, Kenilworth. All participants were approached in their preferred living area including communal areas or private rooms. Considering the fact that a number of the participants had visual impairment, the questionnaire was employed in the format of a one-to-one interview, where the participant had the opportunity to ask the interviewer questions or repetitions of questions. In addition, participants could write their answers with a pen/pencil or verbally, where in the latter the interviewer wrote their answers. Overall, the interviews had different duration based on the participant's ability to understand the questions and provide answers. On average, each interview lasted approximately 1 hour.

### 5.2.4 Analysis of Responses

All participants were analysed using descriptive statistics such as age and gender. Based on the two statistics, the participants were distinguished in two groups. The age of the categories of participants designated in groups was summarised based on mean and standard deviation (SD). Considering the fact that the male group was limited in terms of participants, no specific comparison could be achieved. As such, the age-based groups included younger participants (aged 81-89), and older participants (aged 90-99). The statistics used for comparison were one-way ANOVA, Spearman's rank correlation coefficient, and Kendall's rank correlation coefficient.

In order to strengthen the findings of ANOVA, a correlation analysis with Spearman's coefficient was also undertaken. The Spearman's rank correlation coefficient can quantify any linear or nonlinear statistical dependence between the rankings of two variables [191]. Then, the relationship between two variables is described based on a monotonic function. It actually computes the Pearson correlation between the rank values of two variables. However, since the rank of the variables are used, instead of the linear characteristics of Pearson's correlation, Spearman's correlation is capable of assessing monotonic relationships. It varies between -1 and +1. So that, when observations have similar rank it is high (absolute value 1 when identical) and it is low when observations have dissimilar rank. The Spearman's coefficient is given by:

$$r_s = p_{rg_x, rg_y} = \frac{cov(rg_x, rg_y)}{\sigma_{rg_x} \sigma_{rg_y}} \quad (5.1)$$

where  $p$  denotes the Pearson's correlation coefficient, but it is applied to ranks in the case of Spearman's rank correlation coefficient. The rank variables are given by  $rg_x$  and  $rg_y$  respectively. The covariance of the rank variables is given by  $cov(rg_x, rg_y)$ , while  $\sigma_{rg_x}$  and  $\sigma_{rg_y}$  denote the standard deviations of the rank variables respectively.

Kendall's rank correlation coefficient is a statistical method, which measures the ordinal relationship between two groups [192]. Similarly to Spearman's method, it measures rank correlation, where coefficient values are in the same range (-1, 1). The interpretation of the values is similar: an absolute value closer to 1 indicates strong correlation. On the other hand, absolute values closer to 0 show weak or negligible correlation. The formulation of Kendall's method is as follows:

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\binom{n}{2}} \quad (5.2)$$

where concordant pairs of values  $(x_i, y_i)$  and  $(x_j, y_j)$  are only if  $i < j$ . The opposite implies they are discordant. Otherwise, the pairs are discordant. In addition,  $\binom{n}{2} = \frac{n(n-1)}{2}$  denotes binomial coefficient.

Open-text comments expanding on responses to closed questions and text replying to open questions were analyzed using thematic content analysis. The results of these analyses will be presented in the following sections.

### 5.3 Study Findings

The demographics for each participant are presented in Table 5.1. The gender breakdown reflects the fact that residents in UK care homes are mostly women [193]. All male residents in the care home were approached but the remaining residents had a chronic condition, which prevented them from participating. Two groups are distinguished in the Table: younger group (aged 81-89: P1,...,P6) and older group (aged 90-99: P7,...,P21).

**Table 5.1:** All participants including their gender and age.

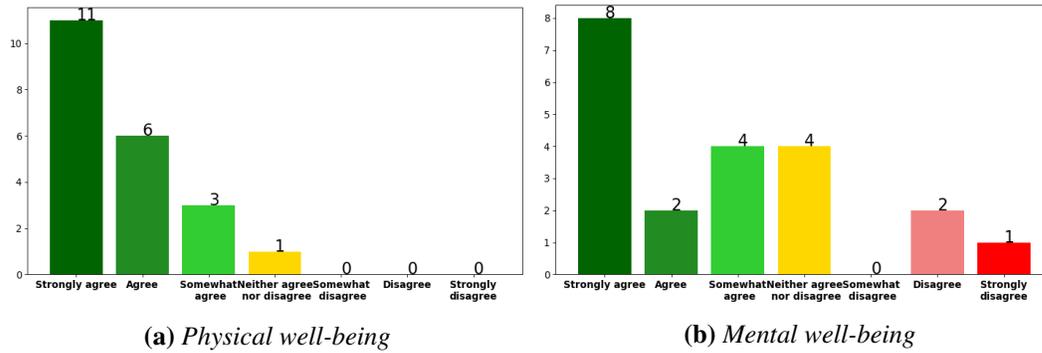
Participants	Gender		Age
	Male	Female	
P1		✓	84
P2		✓	82
P3		✓	81
P4		✓	88
P5		✓	89
P6		✓	84
P7	✓		91
P8		✓	90
P9	✓		94
P10		✓	90
P11	✓		94
P12		✓	93
P13		✓	99
P14		✓	97
P15		✓	93
P16		✓	92
P17		✓	93
P18		✓	91
P19		✓	92
P20		✓	91
P21		✓	91

### 5.3.1 User Acceptance of Sensors

Initially, non-intrusive sensor systems were analysed by asking the participants whether such technology can improve their (1) physical health and well-being, and (2) mental health and well-being. In terms of physical health, a rapid detection of a fall was given as an example. Considering mental health, residents were asked whether they would feel more stable and secure by having such technology in their room. The responses for physical health and mental health are given in Figure 5.1a and Figure 5.1b respectively.

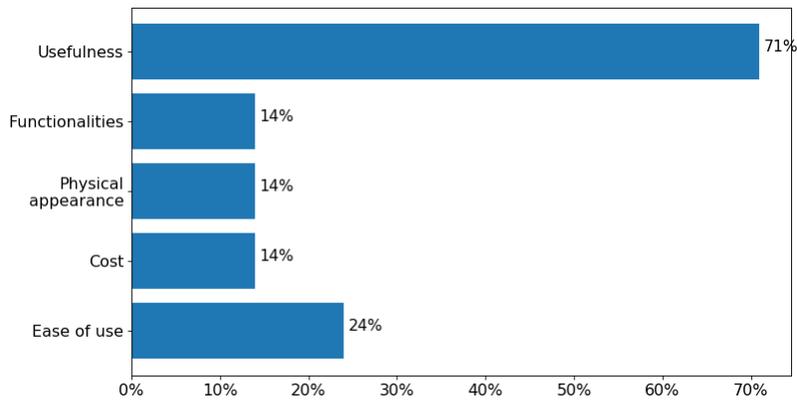
### 5.3.2 Factors Affecting the User Acceptance of Robots

The classical TAM framework considers only two factors for potential user acceptance in its originality: perceived ease-of-use and perceived usefulness. Considering the fact that these two factors can be regarded as limited, additional characteristics are included for user



**Figure 5.1:** Health and well-being improvement based on sensing technology.

acceptance of robots. These include functionalities, physical appearance, and cost. The participants had the opportunity to select more than one option. Figure 5.2 shows the relative importance of the five factors, which can influence the user acceptance of robots among older adults. As observed, usefulness is seen as the most important factor, while ease-of-use is the second most favoured option. However, the newly included factors, functionalities, physical appearance, and cost, still find sympathy among some of the residents.

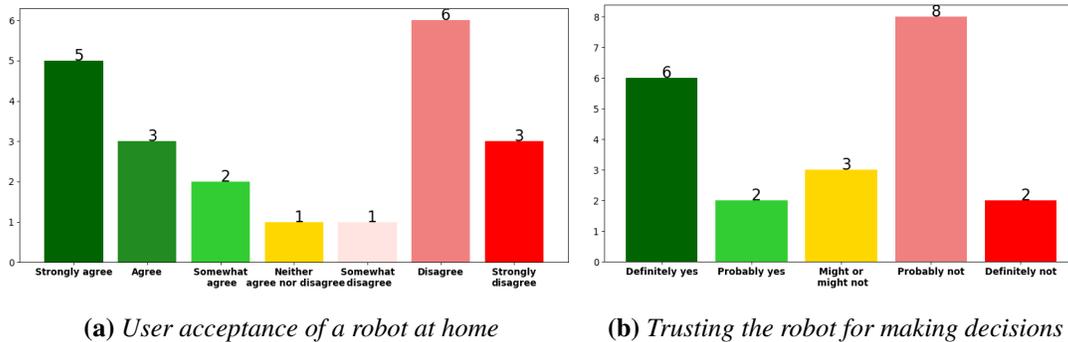


**Figure 5.2:** Relative importance of five robot characteristics

In terms of user acceptance of robots, residents were asked to show their willingness to have robots in their current home (considering the fact that the LTC facility is now their own home). More specifically, the residents had the opportunity to show how far they agree with the sentence “I am likely to use robots in my home”. Figure 5.3a shows the outcome of the question. As observed, the responses are very mixed considering the fact that the residents have previously only interacted with Amazon Alexa. Only 5 residents strongly agreed and 3 agreed, while 6 disagreed and 3 strongly disagreed. This can be attributed to the fact that the robots are often faced with stigma and preconceptions by older populations as discussed in Section 2.6.

The user acceptance of robots was explored further, the residents were asked whether they would trust the robot for making decision for them. The decisions were closely related

to the residents' health. Figure 5.3b reveals breakdown of responses, which is very similar to the previous question.



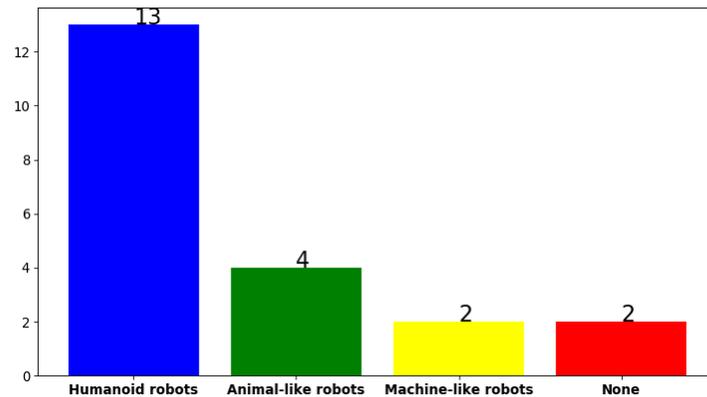
(a) *User acceptance of a robot at home*      (b) *Trusting the robot for making decisions*  
**Figure 5.3:** *User acceptance of a robot at home and trusting the robot for making decisions*

### 5.3.3 User Acceptance of Robots Based on Appearance

To explore the preferred robotic physical appearance, three main types of robots were included in the question - mechanical humanoid, zoomorphic (animal-like), and mechanoid (machine-like). As discussed in Section 2.6.2.1, Mori's uncanny valley is an important concept in field of robotics. More specifically, the uncanny valley shows a situation when a robot resembles humans to a greater extent by possessing nails, skin, and hair. This can cause repulsion and confusion and it is highly undesirable especially among the older populations. Therefore, mechanical humanoid robots are included as examples in this study - NAO [194], Pepper [195], and Romeo [196]. In terms of zoomorphic robots, three robots were presented - AIBO (dog-like), MiRo (donkey-like) and Paro (seal-like). Machine-like robots were also included although they are widely used in manufacturing. However, their functionalities can still be applied in care homes for environments navigation and serving meals. The given examples were Baxter and OSARO. It is important to note that none of the given examples were familiar to the residents. Figure 5.4 demonstrates the given responses. Mechanical humanoid robots were selected by the vast majority of residents (13), while zoomorphic robot were slightly ahead (4) from machine-like robots (2). Considering machine-like robots, one of the residents (a female) referred to Baxter as "It is awful! It looks like a spider". Two of the residents (a male and a female) stated no interest in robots and their physical appearance.

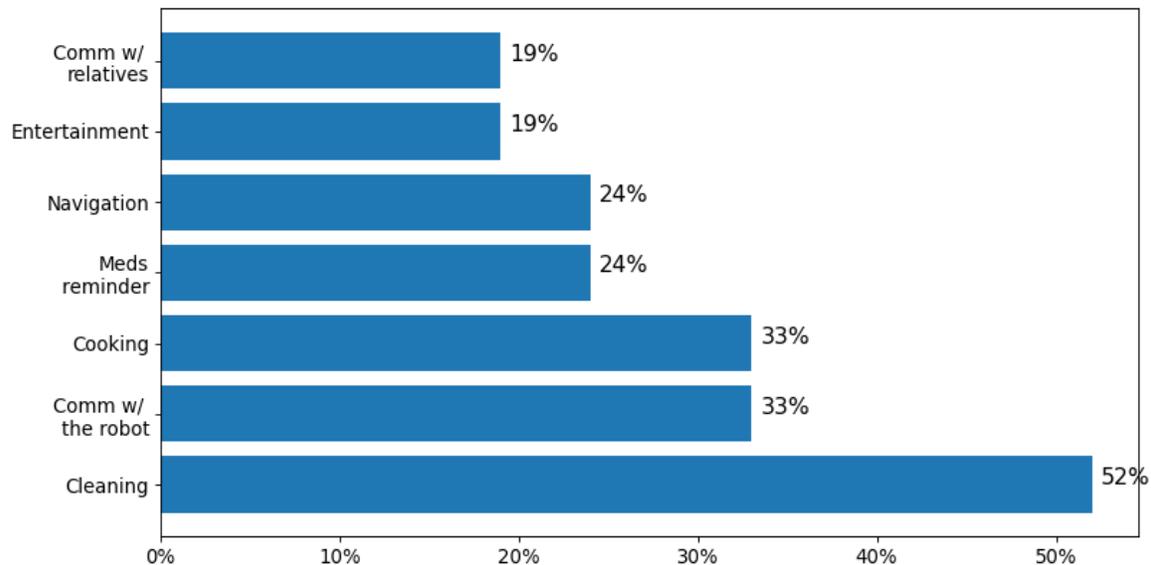
### 5.3.4 User Acceptance of Robots Based on Role and Functions

The user acceptance of robots is further explored by considering their activities based on physical or social role. More specifically, the proposed activities included: (A1) cooking, (A2) cleaning, (A3) medicines reminder, (A4) environments navigation, (A5) entertainment such as singing and dancing, (A6) communication with the robot, and (A7) communication provision with relatives. The social role includes activities A3, A5, A6, and A7, while the physical role was related to activities A1, A2, and A4. Figure 5.5 shows the results for each



**Figure 5.4:** *User acceptance of robots based on physical appearance*

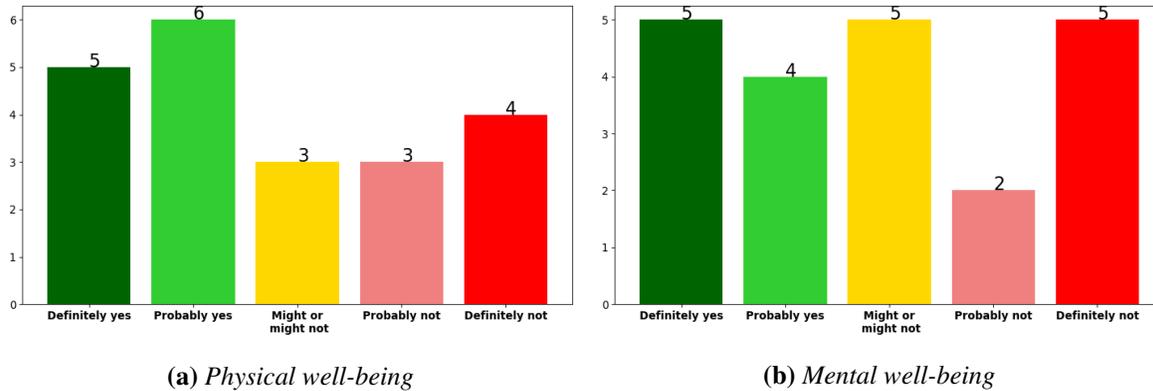
activity considering the fact that the residents had the opportunity to select more than one option. As observed, the most preferred activity was cleaning with more than half of the residents regarding it as important. Perhaps surprisingly, communication with the robot was the second most preferred activity. This result can be attributed to the fact that loneliness is a major concern among older populations. Despite the fact that the residents were living in a care home with peers, many of them preferred spending their time in their own rooms. Communication provision with relatives and entertainment were the least preferred activities. In terms of communication provision, many of the residents had their own mobile phone, laptop or personal computer. Considering entertainment, the residents expressed feelings that the robot cannot entertain them since they preferred old-fashioned activities such as knitting and crafting.



**Figure 5.5:** *User acceptance of robots based on functions*

The final two questions in the survey were linked to the previous question of preferred

activities. More specifically, the residents were asked whether these activities can improve their (1) physical health and well-being, and (2) mental health and well-being. Figure 5.6a and Figure 5.6b reveal the results for both types of health respectively.



**Figure 5.6:** Health and well-being improvement based on robot's activities

### 5.3.5 User Acceptance of Robots Based on Cost

As discussed in Section 2.6.1, cost was regarded as an important constraint for home residents. However, considering the current study, the residents were living in a care home, and expressed feelings that as long as they do not pay, they consider the price unimportant. To conclude, cost was regarded as important as physical appearance and functionalities.

### 5.3.6 Open-Ended Question

An open-ended question was included in the questionnaire by allowing the participants to list any additional activities they wish to be performed by the robot. The majority of the residents (14) decided to leave the question blank while the remaining seven responses were mostly focused on cleaning activities, including a smart vacuum cleaner. This further stresses on the fact that the activity cleaning was the most preferred activity among the residents. In addition, one residents who experienced navigation problems left the comment “*Help me walk*”.

### 5.3.7 Comparison Between User Acceptance of Sensors and Robots

Considering the fact that the user acceptance of both (1) sensors, and (2) robots is evaluated in this study, a comparison between the two intelligent agents can be achieved. Both agents were evaluated in terms of health improvement considering physical health and mental health. In terms of sensor systems, Figure 5.1a showed a strong acceptance based on physical health. However, in terms of mental health, Figure 5.1b showed mixed results.

The user acceptance of robots was very mixed considering both physical health shown in Figure 5.6a and mental health in Figure 5.6b. The reasons can be attributed to the fact that the work “robot” can engender feelings of unease and fear as described in a previous study [197]. Further discussion and conclusion of the findings are provided in Chapter 6.

### 5.3.8 Participants’ Responses Based on Age

To explore any statistically significant differences between the responses of the two age-based groups, ANOVA was used initially. The null hypothesis showing no differences was accepted for the majority of answers ( $p\_value > 0.05$ ) as shown in Table 5.2. However, it was rejected for Question 7 ( $p = 0.001$ ) exploring the preferred robotic activities, Question 8 ( $p = 0.01$ ) and Question 9 ( $p = 0.01$ ) exploring whether robots can enhance physical health and mental health respectively based on these activities.

**Table 5.2:** One-way ANOVA statistics for the two age-based groups described in Table 5.1

Question	f-score	p-value
Q1	1.16	0.3
Q2	2.5	0.13
Q3	3.08	0.11
Q4	4.41	0.057
Q5	4.37	0.06
Q6	0.47	0.52
Q7	16.5	0.001**
Q8	9.52	0.01**
Q9	9.52	0.01**

To strengthen the findings extracted from the ANOVA, two additional statistical methods are applied - Spearman’s rank correlation coefficient and Kendall’s rank correlation coefficient. Considering the fact that the p-values derived from both methods are in the range of (-1, 1), a closer value to -1 or 1 shows a negative or positive correlation respectively. On the other hand, closer values to 0 show weak correlation. As observed in Table 5.3, both methods discover weak correlation regarding Question 7, Question 8, and Question 9. This closely matches the findings derived from ANOVA. In addition, the values for Q6 are none due to a constant input, where one of the groups achieved the same distribution of responses.

Considering Question 7, both age groups demonstrated the same preferences for activities (A1) cooking, (A2) cleaning, and (A7) communication provision with relatives. However, physical health-related activities such as (A3) medicines reminders, and (A4) environments navigation were more preferred by the younger group (aged 81-89). In addition, (A5) entertainment found stronger sympathies among the older group (aged 90-99), but still, not very strongly.

**Table 5.3:** Spearman's rank correlation coefficient and Kendall's rank correlation coefficient for the two age-based groups

	Spearman's correlation		Kendall's tau	
	correlation	p-value	tau	p-value
Q1	0.93	0.002	0.88	0.01
Q2	0.65	0.11	0.57	0.1
Q3	0.54	0.34	0.47	0.28
Q4	0.42	0.34	0.37	0.3
Q5	0.91	0.02	0.88	0.04
Q6	None	None	None	None
Q7	0.16**	0.72	0.11**	0.74
Q8	0.18**	0.76	0.17**	0.7
Q9	0.18**	0.76	0.17**	0.7

In terms of enhancing physical well-being (Q8), very similar results were achieved in overall agreement and disagreement with robots activities. More specifically, a larger group of younger adults selected “definitely agree” compared to the older adults who have chosen “probably agree”. Also, the latter selected more “definitely disagree” compared to the former that preferred to “probably disagree”.

Considering the perceived enhancement of mental well-being (Q9), a slightly larger group of the younger adults have selected “might or might not” compared with the remaining answers, where an equal distribution is observed. In terms of the older group, four have selected “definitely agree” along with other four who have selected “definitely disagree”.

## 5.4 Chapter Conclusion

This chapter has provided the much needed evidence for user acceptance of non-intrusive sensing technologies and robots among the extremely older, frail population living in LTC facilities. Sensors appeared to be more acceptable among older adults compared to robots. In addition, the approach presented an enriched TAM framework for the acceptance of both sensors and robots, also seen from the perspective of health improvement. Both physical health and mental health were considered for evaluation with interesting findings showing that older adults had reservations in regard to their mental health and well-being. Mechanical humanoid robots were selected by most of the residents compared to zoomorphic and mechanoid robots. Cleaning, as a physical activity, was determined as the most preferred robotic activity by the participants.

The novelty of the work should be seen from the perspective of very limited studies examining the user acceptance of sensors and robots among the population living in LTC. More specifically, the current literature lacks such diverse evaluation of both (1) sensors, and (2)

robots including health improvement, extended TAM framework, trust, physical appearance, and functions of the robot.

This work provides interesting findings regarding the attitudes and thoughts of older adults in regards to the two types of technology. Shortcomings of the study include the limited number of older adults in this experiment, which can be improved in future studies. In addition, the participant's prior experience of information technology and robots can be considered for future work. The following chapter discusses the findings of this chapter as well as Chapter 3, and Chapter 4. Strengths and limitations are presented as well as areas for future work.

# Chapter 6

## Discussion

### 6.1 Introduction

This chapter is aimed at discussing the findings of the three proposed sets of contributions: (1) LRIR data analysis, (2) micro-Doppler radar data analysis, and (3) user acceptance of sensors and robots among older people. Strengths and limitations of the proposed methods and frameworks will be considered as well as areas for future study.

The next section of this chapter begins with IR data discussion considering the presented holistic understanding of such solutions for human activity recognition as described in Chapter 3. More specifically, limitations of the extrapolation method along with the most optimum recognition models will be evaluated by attributing factors for their success.

### 6.2 Discussions on the Analysis of IR Data

#### 6.2.1 Limitations of the Extrapolation Method

Considering the fact that the durations of sensor recordings are not identical due to activities differences, two methods for frames equalisation were proposed in Section 3.4.1.1.1. In terms of the extrapolation method (described in Algorithm 2) aimed at removing frames from recordings larger than 40 frames, specific limitations were observed. More precisely, the method was shown to favour earlier frames for deletion in some cases. This can be regarded as biased. As discussed earlier in Section 3.3.3.2, the sparse methods for dimensionality reduction showed that central frames contributed to classification more significantly compared to the two extremes. Hence, it is expected that the essence of activities is distributed in the central frames. In future work, sampling techniques for extrapolation method including Gaussian sampling, uniform sampling, smooth fractionator, and variation analysis can be researched for finding the optimum strategy.

## 6.2.2 Optimum Model for Activity Recognition

The obtained results from the analysis of *Coventry-2018* data in Section 3.4.2 demonstrated that SVD was the most successful feature extraction method for activity detection using LR, RF, k-NN, and SVM classifiers. That means the most discriminating features lie along the directions with highest variance in orthogonal subspace. SVD uses the whole training samples of all classes to find the eigenvectors. That is 75% of the total 450 number of samples (of the Coventry 15-class problem)  $.75 \times 450 = 337$ . However, considering the other used feature extraction technique, Fisher's method requires computing the within-class covariances. Since there are only 30 samples in each of the 15 classes, then  $0.75 \times 30 = 22$  training samples are used for computation of the within-class covariance matrix  $W$ . That is very low compared to the large number of variables 2560. This probably puts SVD in a better rank condition referring to the achieved accuracies. Another successful architecture was the CNN-LSTM model, where features were automatically extracted from the CNN layers, while the LSTM layers learned the temporal variations of the sequences. While SVD showed success with different classifiers, the particular combination of SVD+LR achieved 100% average testing accuracy. K-NN showed lower accuracies due to the fact that it computes Euclidean distances, but in some cases two or more classes can be very close, which explains the poorer performance. LR assumes a hypothesis for binomial distribution of the data. In addition, it tries to find the parameters of the model based on generative solution. Finally, LR finds a boundary between classes based on the likelihood maximisation. To conclude, these constraints could be the potential reason for its superior performance.

## 6.2.3 The Effect of Changes in Sensing Layout and Model Generalization

GLCM showed some level of success for model generalisation, when a model trained on small layout data was used to classify samples of large layout. That is due to the nature of GLCM in quantifying image texture patterns. In regards to SVD, the main directions of variations found in orthogonal subspaces might not match between the small and large layout. In addition, due to the changes in spectral variations of 2D arrays in the small and large layouts, a previously trained model on the compressed DCT coefficients of small layout does not work on large layout data. However, a model trained on the texture patterns captured by GLCM for small layout, identifies the same patterns in large layout data and shows generalisation capability.

## 6.3 Discussions on the Analysis of Doppler Radar Data

### 6.3.1 Analysis of Local Patching Strategy

The proposed method for local patching combined with DCT presents a simplistic and intuitive strategy, while providing accurate clustering results. More specifically, the method's

parameters were selected using the unsupervised Dunn’s index, which provided effective results for the unsupervised framework of the study. The method can be further applied to supervised projects, where the classification accuracy can be used for providing even more accurate results in the selection of the model’s parameters.

### 6.3.2 The Effect of the Clustering Strategy

The clustering results of extracted features with CVAE, DCT-based method, and entropy-based method, were implemented with K-Means and K-Medoids. Both methods can be implemented as distance-based: Euclidean distance was used with K-Means, while K-Medoids was incorporated with Manhattan distance. The results showed that K-Medoids provides better clustering for the local DCT-based features, while CVAE encoded features revealed better incorporation with K-Means. For the case of the local entropy-based extracted features, both clustering techniques showed similar results, which is attributed to the much lower number of extracted features.

### 6.3.3 Optimum Models for Doppler Radar Data

The clustering results in Section 4.3.4 showed that the most superior feature extraction architecture was CVAE. On the other hand, the two proposed local patching strategies also provided accurate clustering results, especially the local DCT-based method. While accuracy is only one factor, it is important to consider computational complexity between the three methods. Table 6.1 reveals the total computational time over  $4 \times 150 = 600$  samples of the four folds. As observed, the deep CVAE architecture exhibits significantly more expensive computational time in comparison with the two proposed methods.

**Table 6.1:** Total computational time over 600 samples of the four folds for the two proposed feature extraction methods as well as CVAE.

	DCT-Based Method	Entropy-Based Method	CVAE
Time	0.58 s	1.1074 s	698.39 s

While CVAE is a deep NN architecture, the features derived from the two proposed methods are extracted from local patches. The considered local patches in both proposed techniques are smaller than the original 2D maps. Considering the CVAE architecture, the three-dimensional  $2 \times 100 \times 32$  cube was used, where a number of different convolutional filter weights were learnt. Unsurprisingly, this leads to a significantly more complex model. The two proposed methods will be able to learn faster on new data in comparison with CVAE. It can be concluded that CVAE, and the two proposed methods both have advantages and disadvantages. Hence, they can be applied to other unsupervised architectures or even supervised architectures, where the accuracy can be used for model parameters selection.

### 6.3.4 Fusion of Different Methods

A potential idea is to fuse the features derived from different feature extraction techniques in order to improve the clustering results. However, this would result in a higher number of features, which is undesirable given the context of a small dataset in this study. Fusion of the DCT-based features and entropy-based features was implemented, which did not result in improved clustering results. This result can be attributed to the fact that the nature of the features extracted from the two methods is extremely different. In these terms, normalisation of the features can also be incorporated, which can also produce weak clustering results.

### 6.3.5 Analysis of Higher Standard Deviation

As described, LOOCV was implemented for the clustering results in order to strengthen the findings in the study. This, on the other hand, led to higher standard deviation in some case. The higher standard deviation can be attributed to the fact that the subjects who volunteered for capturing the dataset, differed in gender and body size. For future work, a higher number of subjects can be applied for capturing datasets.

### 6.3.6 High-Dimensional Data Projection and Visualisation

Manifold learning methods can be useful for visualising high-dimensional dataset by projecting it a 2- or 3-dimensional space. The three used methods, t-SNE, MDS, and LLE, are known to preserve the similarity patterns of the high-dimensional data when projecting it to a much smaller dimension. The idea to use these manifold learning methods in this study is also linked to extracting interesting insights for the different classes in the dataset. T-SNE, MDS, and LLE were first applied to the raw data by projecting it to 2-dimensional feature space. The visualisation results revealed high similarities and overlapping between data samples belonging to different classes. To improve this, the CVAE encoded features were used as an input of the methods for a similar projection. The subsequent results showed improvements for the three manifold learning techniques. More specifically, the intra-cluster distance was shown to be improved considering the fact that data samples from the same class were more closely projected. On the other hand, classes such as walking (1) and jumping (3) had overlapping classes, which was also evident in the confusion matrices. The running (2) class was the most accurately clustered activity, and the visualisation results proved this result.

### 6.3.7 Results Comparison with Supervised Strategies

The same micro-Doppler radar dataset was used in a supervised study [60]. Considering the dataset's high-dimensionality, feature extraction was achieved using off-the-shelf techniques. More specifically, SVD, PCA, and physical features were applied in the previous study despite the close similarity of the first two methods. SVM was used for classification, while different sizes of the training set were proposed for comparison. The testing accuracies

reveal more than 80% performance for different sizes of the training/testing sets. The features in our work based on local DCT and CVAE were used in a supervised fashion, where similar results were observed over 4-subject LOOCV.

## 6.4 User Acceptance Analysis

The user acceptance of sensors and robots study provided findings on the views, attitudes, and perceptions among the older population. In terms of sensors, such systems are now widely researched and introduced in care home [16, 18, 21] for the purpose of human activity recognition. Such systems are often non-intrusive (excluding wearable technologies) and users do not have a direct contact with them. Considering robots, the applications are usually more complex, but still find applications in care homes [132].

### 6.4.1 User Acceptance of Non-Intrusive Sensing Technology

The older participants in our study showed strong acceptance of non-intrusive monitoring technology. Many of the residents commented that as long as the technology is beneficial, they are supportive. Sensor systems were further favoured based on their non-intrusive and indirect contact with the residents. Researchers have discovered that older adults valued the “felt need” and “product quality” for assistive technology [198], which is highly consistent with the current study. As discussed previously, the residents had reservations in terms of their mental health enhancement based on sensor systems. This can be attributed to the fact that older adults wish to portray themselves as independent and capable instead of frail and vulnerable [199]. In addition, the mental health topic may have been a taboo topic in the past considering the age of the participants [200].

### 6.4.2 Older People’s Acceptance of Robots

The classical TAM framework including perceived ease-of-use and usefulness only confirmed in our study that these two factors were the most favoured. However, additional characteristics such as physical appearance, role and functions, and cost also found sympathy among some of the residents.

In terms of robot’s physical appearance, our research discovered that mechanical humanoid appearance was the most preferred among the residents. The selected humanoid examples for inclusion in the questionnaire resembled humans, but not to a great extent, so they could not “fall” in Mori’s uncanny valley. In addition, animal-like robots were the second most preferred category followed by machine-like robots. The same preference has been observed in an earlier study [30].

Considering the preferred robotic role and activities, physical role with cleaning found the most sympathy among the residents. This can be attributed to the fact that many older

adults who move to LTC facilities have physical problems affecting their mobility and ability to perform daily tasks including cleaning and cooking. Communication with the robot was a highly favoured activity, which is explained by the major loneliness affecting older adults including residents in care homes [201].

By analysing the two age-based groups, three statistical methods discovered stronger preferences by the younger group (aged 81 to 89) towards more health-related activities such as medicine reminder and environments navigation. This was also confirmed by the responses of the subsequent related question for perceived physical health improvement. Entertainment was only preferred by some of the residents in the older group, but not as significantly.

In conclusion, the robot's cost was regarded as less important than usefulness and ease-of-use, but as important as physical appearance and functionalities. The literature review covering the importance of cost among care home residents is vastly limited, however, our finding corresponds to the findings derived from private households [202].

### **6.4.3 Acceptance and Trust of Robot in Own Home**

Studies have discovered that trust is an important factor for user acceptance of robots [203] and home monitoring systems [204]. Residents in this research were largely cautious in their responses to the questions of trust and accepting a robot in their home (i.e. the care home where they were living), with the most common response being to not accept a robot at home and probably not to trust it to make decisions. This contradicts with previous research [149] where 21 older adults ( $\text{Age}=80.25\pm 7.19$ ) were identified as generally willing to accept robots at home. However, it is worth noting that the mean age ( $90.42\pm 4.46$ ) in the current research is higher, which potentially affects the acceptance of robots.

### **6.4.4 Comparison Between User Acceptance of Sensors and Robots**

The findings of the study have shown that non-intrusive sensor systems are more accepted compared to robots with a physical presence. This corresponds to previous research, where interaction-free sensor systems were preferred by the participants [25]. In addition, physical robots with cleaning capabilities were preferred by the majority of residents. On the other hand, social robots were mostly dismissed as the findings are again in line with interaction-free agents.

It is important to consider the concept of dignity for both physical and social roles. There exists an emerging concept of “weak” and “vulnerable” robots, which aims to address the issue of losing dignity in the care home sector [205, 206, 207]. To respond to this, “vulnerable” robots aim to retain the self-esteem and dignity of residents. For example, the most preferred activity in our study, cleaning, can reduce the human contact of the residents [208]. In this situation, “vulnerable” robots can provoke the user's participation by moving objects

[209]. In addition, this can allow the residents to feel that they have a useful role in the society, which is crucial for their self-esteem [210]. The human carers can then focus on more social and meaningful activities while being “replaced” by the robot [207].

#### **6.4.5 Comparison Between Age Groups**

The two age-based groups including 81-89 and 90-99 year old showed correlation in their responses for the majority of questions. However, a weak correlation was discovered for the preferred robotic activities along with the following perceived physical health and mental health improvement based on these activities. More specifically, the younger group preferred more health-related activities as well as showed higher perceived physical health improvement based on these activities. Considering mental health improvement, the younger group was uncertain by selecting “might or might not” more strongly compared to the older group.

#### **6.4.6 Strengths and Limitations of Study**

In terms of strengths, the study achieved comparison in the user acceptance of (1) sensors, and (2) robots, which has not been evaluated in previous studies. Additionally, user acceptance factors beyond the ones incorporated in the classical TAM framework were included (physical appearance, cost, functionalities), which played an important role. The likely impact on the physical health and mental health was analysed including comparison between the two types of health. Acceptance and trust of robots with a physical presence was also researched along with preferred physical appearance, and role and functions of the robot. Considering limitations, these can be viewed by the relatively small study sample, which included 21 residents living in one LTC facility. In addition, the number of male residents was extremely small, which did not allow for a similar comparison between the responses as with the two age-based groups. This is due to the fact that the number of healthy (excluding neurological diseases) individuals in care homes is very small. Still, the study showed interesting findings derived from the extremely old population, which can be confirmed in further studied incorporating a greater number of participants.

### **6.5 Overall Discussions on the Use of Non-Intrusive Sensing Technologies for Health Monitoring of Aged Population**

The overall aim of this thesis was linked to the possibility of the use of non-intrusive sensing strategies for monitoring healthcare of aged populations. Both LRIR data and Doppler radar data were analysed with the application of different AI methods for them. In addition, a number of methods for pre-processing and feature extraction were proposed specifically for these two types of data.

Considering the fact that both IR sensor and Doppler radar sensor will potentially have application in healthcare and eldercare, the privacy of the care home residents is a major concern. Both methods have proven to be non-intrusive since subjects identification from LRIR data and Doppler spectrogram is avoided. The field of view of the utilised IR sensor in this study is  $60^\circ$ , while the claimed coverage is 7 metres. While the coverage range of the Doppler radar is much higher, both devices are ideal for living area dimensions and settings. In regards to the economical challenge, IR sensors are considered relatively inexpensive. A concern would be the need for more IR sensors as their coverage range is not substantial. As such, the overall economical challenge would be increased. The Doppler radars have shown to be considerably more expensive, but on the other side offer higher coverage range.

In regards to complexity, the Doppler radar's refresh rate is slower in comparison with the IR sensors. As discussed in Chapter 3, the IR sensor can have a refresh rate of 10 FPS. This advantage places the IR sensor in a better position compared with the more sensitive Doppler radar. IR sensors are sensitive in terms of heat sources, which can pose an issue if positioned in the sensor's field of view. Considering Doppler radars, the devices have shown sensitivity toward small fractions of the body or even the heart beat of a person. This ability can be seen as an advantage towards the detection of more sensitive activities. The more specific activities involving smaller fractions of the human body can be extremely difficult for detection with IR sensors. The currently existing datasets along with the proposed dataset cover only non-specific activities such as standing, sitting, walking, etc.

In terms of accuracy, which is a crucial factor, IR sensors have shown to have better recognition. As outlined in this thesis, the accuracy can be increased even to 100% using the correct combination of AI strategies for pre-processing, feature extraction, and classification. While unsupervised strategies were used for Doppler radar data in this thesis, it is expected that their results would not be comparable with supervised strategies. However, the same Doppler radar dataset has been used with supervised strategies by fellow researchers [60]. The achieved results were above 80% for different sizes of the testing set. Considering this finding, it can be concluded that IR sensors provide better recognition in terms of classification accuracies.

Unsupervised learning is a very important field of machine learning. Using unsupervised strategies for IR sensor data and Doppler radar data can have benefits for improving the model by purely providing new data captured by the devices. Hence, the burden of manually labelling the data can be eased.

## 6.6 Chapter Conclusion

This chapter provided discussion of the main findings for all three proposed areas of contributions. Initially, extrapolation limitations were presented. The most optimum classification modelling strategy for the LRIR data was considered with results interpretations. In addition, the modelling strategy with GLCM+LR was described for its model generalisation abilities.

In terms of micro-Doppler radar data, the two proposed methods for unsupervised feature extraction were compared with the state-of-the-art techniques. Both advantages and disadvantages were provided to strengthen the analysis. In addition, data visualisation with manifold learning methods in the two scenarios was analysed. Potential fusion of methods was discussed as well as the higher standard deviation in the average results. Finally, comparison of the proposed architectures with supervised strategies was made.

The proposed user acceptance framework was analysed in regards to comparison between the acceptance of sensors and robots. In addition, the extended factors in the TAM framework were discussed for their role in the acceptance of new technologies. The two-age based groups' responses were analysed in cases where they differed statistically significantly. Finally, strengths and limitations of the work were provided.

The next chapter will conclude the thesis by providing the final outcomes and remarks of the findings. More specifically, the research questions will be discussed as well as thesis limitations and areas for future work.

# Chapter 7

## Conclusion

This thesis has addressed the development of machine learning and data analysis approaches for the efficient classification of human activity data derived from non-intrusive sensors such as IR sensors and micro-Doppler radar sensors. It has also explored the important factors influencing the user acceptance for using non-intrusive sensing technology and robots for health monitoring of aged population.

The current literature for human activity recognition with IR sensors is very limited. The majority of the studies provided an introductory analysis, which was not researched deeper. Considering IR data, a number of studies have proposed analyses [16, 17, 19, 20, 21]. These studies lacked holistic evaluation of LRIR data including optimum feature extraction and classification selection, noise analysis, sensor position, multiple layouts, model generalisation, and multi-subject activities. In this thesis, two LRIR datasets were employed with diverse capabilities in Chapter 3 in order to address the aforementioned shortcomings of the previous studies. In addition, a number of feature extraction and classification combinations were used to discover the most optimum model for this data. More precisely, novel methods for interpolation, extrapolation, and periodic noise reduction were proposed. All proposed methods were verified, which proved their success for application to LRIR data.

Micro-Doppler radar for human activity recognition is used predominantly in supervised applications. The current literature identified only two studies concerned with unsupervised learning based human activity recognition with such radars [23, 24]. The first study [23] only used existing feature extraction and clustering techniques for the purpose. Regarding the second study [24], a novel feature extraction architecture was proposed, but still the accuracy was intermediate. In this thesis, two unsupervised feature extraction methods for 2D Doppler radar images were proposed in Chapter 4. Additionally, a holistic evaluation and understanding of Doppler radar data was achieved based on unsupervised strategies.

Besides the technical work, a general user acceptance framework for sensors and robots among the older population was proposed in Chapter 5. The current literature mostly included user acceptance of sensors and robots following short to long term interaction [30, 31, 32, 33, 34, 35, 39, 189, 190]. In addition, user acceptance comparison between sensors

and robots was not observed in the literature. In this thesis, a number of additional characteristics were included. These included health improvement, extended TAM framework, trust, and robot's physical appearance and functionalities.

The thesis was designed to address the shortcomings of previous literature. In the following section, the answers of the research questions will be given.

## 7.1 Research Questions and Thesis Contributions

The main goal of this research was to provide answers to the following research questions:

1. *“Is it possible to develop novel approaches for machine learning and data analysis to correctly classify human activities derived from non-intrusive remotely sensed data?”*

In this thesis, the research question is addressed based on the following research work:

- Acquisition and preparation of human activity data using two types of non-intrusive sensing technologies: (a) LRIR sensors (*Coventry-2018* dataset was generated and *Infra-ADL2018* dataset was used); (b) the previously acquired dataset based on Doppler radar sensing systems.
- Pre-processing of data such as frame equalisation for LRIR data using interpolation and extrapolation, and denoising of 2D spatio-temporal maps of LRIR data by developing a new method.
- Interpretation of 2D Doppler maps for human activity recognition by developing two novel methods for unsupervised feature extraction.
- Comparing different supervised and unsupervised feature extraction and classification methods as well as deep NN techniques for LRIR data and Doppler radar data.

2. *“What factors influence the user acceptance of non-intrusive sensing technology and robots with a physical presence by older adults?”*

In this thesis, the research question is addressed based on a proposed user acceptance framework for non-intrusive sensing technologies and robots with a physical presence among aged populations. The results of the analysis, covering the acceptance of 21 older adults, revealed the original factors incorporated in TAM framework as the most influential for user acceptance of robots. Other factors included in the analysis such as cost, functionalities, and physical appearance played a less important role for the user acceptance. Physical health improvement was highly valued by the participants in the study, while the factor of mental health improvement was met with mixed feelings for both sensors and robots. In addition, mechanical humanoid robots were valued by the majority of the participants compared with mechanical zoomorphic robots and machine-like robots. Finally, the participants preferred physical robots with cleaning as the most valued robotic activity.

## 7.2 Human Activity Recognition Using IR Sensors

The thesis firstly explored the use of IR sensors for human activity recognition. The main advantages of using IR sensors are their non-intrusiveness, cost-effectiveness, and ease-of-use. Additionally, the IR sensors are capable of providing extremely low resolution images, which correspond to cheap processing time. As a whole, a limited number of studies were identified from the literature for the purpose of human activity recognition with potential applications in healthcare and eldercare. In addition, the studies lacked deeper insights in regards to the aforementioned constraints.

A holistic analysis was undertaken, which initially included pre-processing. More precisely, two novel methods for interpolation and extrapolation were proposed for frames equalisation. The frames equalisation issue was discussed only in a single study [21] moderately with no further details. As such, in our analysis the two proposed methods were compared with random methods for adding / deleting frames with improved recognition capabilities. However, the extrapolation method used for deleting frames favours earlier frames in the sensors recordings in some cases. As such, it can be further improved by using a more sophisticated technique, using some sampling strategies that follow standard distributions, such as Gaussian or uniform or based on analysing the frames variations. Then, periodicity in spatio-temporal maps of IR data was addressed, which was formed by vectorisation of the low resolution images over 40 frames of each activity. This periodic noise negatively influenced the human activity recognition accuracy. A novel supervised periodic noise reduction technique was proposed to address the periodicity issue, which exhibited up to 13.28% recognition improvement in the conducted tests.

Secondly, the feature extraction and classification were considered for analysis. While previous studies were based on a single method for feature extraction and classification, this thesis explored a number of methods for both purposes to discover the most optimum model. Following a number of tests with machine learning and deep learning approaches, the SVD+LR model was discovered as the most successful by obtaining an average testing accuracy of 100%. The deep learning architecture in the face of CNN-LSTM also obtained a high accuracy recognition rate, but the model training can be considered computationally expensive compared to the traditional machine learning architectures. The achieved robust accuracies over the datasets with a limited number of activities and subjects increase the expectations for having very high accuracy in real settings with more subjects and activities.

The *Coventry-2018* dataset included two layouts: (1) small layout, and (2) large layout. Tests were conducted to discover the models' sensitivity towards distance change. It was then observed that all models performed better when the small layout was used compared to the large layout. This is due to the fact that the Grid-EYE sensors have a sensitivity range of 5 to 7 metres and perform poorly when the distance increases.

Optimum sensor position and layout are crucially important considering the increasing cost of the system with an increasing number of sensors. As such, the goal is to use fewer

resources. By using the *Coventry-2018* dataset, it was discovered that Sensor-1, which was a side sensor, had the most optimum performance. Similarly, in *Infra-ADL2018* dataset it discovered that Sensor-2, again a side sensor, provided the most optimum performance. By having this knowledge, unnecessary sensors can be removed from the scenario, which would only require the most successful sensors or a set of sensors.

Finally, model generalisation was researched in two perspectives. Firstly, model generalisation in terms of layout was undertaken in two scenarios. In the first scenario, the model was trained on the small layout and tested on the large layout, and vice versa. All models were deployed for this purpose and it was discovered that GLCM+LR was the most optimum modelling strategy. This can be extremely useful in situations where sensor displacement might happen in practical settings. In addition, model generalisation including mixing small layout data and large layout data was achieved. For this purpose, CNN-LSTM was discovered as the most optimum architecture. Secondly, model generalisation in terms of unseen subjects for the *Infra-ADL2018* was conducted as this dataset was composed of nine subjects for single-subject activities and eight subject combinations for double-subject activities. Results revealed stable performances, while some activities for specific subjects or subject pairs were not recognised in any of the trials. Therefore, analysis was undertaken to solve the issue. It was discovered that the patterns of these activities greatly differed from the patterns of the same activities performed by the other subjects or subject pairs. This reveals the need for a larger number of subjects with diverse age, body behavior and shape to generalise the models further for real settings. In addition, it is required to arrange different experimental settings by making them more realistic. These include the number and type of furniture in the room and different heating sources such as kettles, heaters, and others. As discussed in Chapter 3, a higher number of sensors might be required to capture information in cases where one or more sensors become occluded as well as covering more space in the room.

### 7.3 Human Activity Recognition Using Doppler Radars

Micro-Doppler radars have been used for human activity recognition with significant results and findings. However, the current literature is focused mainly on supervised learning applications. On the other hand, unsupervised learning is of crucial importance in cases of unlabelled or poorly labelled projects considering the fact that labelling is time-consuming and expensive. There were only two existing papers that addressed unsupervised learning for Doppler radar based human activity recognition. However, they suffered from intermediate accuracies and lack of holistic understanding of such data for the aim of unsupervised learning.

In Chapter 4, two unsupervised feature extraction methods for 2D micro-Doppler images were proposed. Both methods were based on local feature extraction based on different filtering methods for feature quantification. More precisely, the first method was based on applying 2D DCT on local patches of the image. The second method was based on entropy

analysis of local patches extracted from certain image patterns. Both methods were verified with the unsupervised metric based on Dunn's index.

In addition to the proposed two methods for unsupervised feature extraction, deep CVAE architecture was considered for the same purpose. The CVAE architecture exhibited improved recognition capabilities compared to the two proposed methods. However, the training time for the CVAE architecture was much higher compared with the two proposed methods. As such, despite the slightly lower performance of the proposed methods, they represent an interesting finding by allowing faster updates of the previously trained models on new data.

Besides the aforementioned three methods for unsupervised feature extraction, state-of-the-art techniques were considered including CAE, PCA, and 2DPCA,. The two proposed methods and CVAE performed noticeably better compared to these methods. In addition, a number of heuristic techniques were used for identifying the correct number of clusters  $K$ . These included Elbow method, Silhouette coefficient, Davies-Bouldin index, Dunn score, and Calinski-Harabasz index.

In addition to these contributions, visual learning enhancement with manifold learning methods was considered. More precisely, the extremely high-dimensional data was transformed to a 2-dimensional space. This was achieved by firstly, transforming the raw data to a 2-dimensional space, and by transforming the CVAE encoded data to a 2-dimensional space. The visualisation results revealed better class separability when the CVAE encoded data was used.

## 7.4 User Acceptance of Sensors and Robots By Older Adults

In Chapter 5, the social aspect of the work was outlined. The acceptance of sensors and robots is a crucial factor since current research identified the social stigma and negative preconceptions as major challenges. In this work, the user acceptance was observed from a number of perspectives, which were not covered in previous literature.

A user acceptance questionnaire was administered in care home settings by assessing the willingness of 21 older adults to accept sensors and robots. The TAM framework, which is a model of extreme importance in the literature, was enriched to cover beyond perceived usefulness and ease-of-use. In addition, health improvement, considering both physical health and mental health, were assessed. Trust and physical appearance with preferred role and capabilities for robots were included as well.

The findings revealed the high user acceptance of remote sensors for monitoring activities. The user acceptance in this sense was seen from both physical health and mental health improvement. Results showed that the physical health improvement was praised by the participants, while they had some reservations in terms of mental health improvement. On the

other hand, the acceptance of robots uncovered mixed feelings and preconceptions. Mechanical humanoid physical appearance was mostly preferred by the residents, while cleaning was identified as the most important activity to be conducted by a robot. Finally, perceived usefulness and perceived ease-of-use were identified as the most important robotic characteristics, while physical appearance, cost and functionalities also played an important role.

## 7.5 Research Limitations

One of the main limitations of this research was the fact that datasets were built by using younger participants, whose patterns of gestures and activities may not resemble the ones of the older population. This is due to the difficulties of recruiting older participants as it was observed in the user acceptance survey in Chapter 5. The number of participants for the datasets can be regarded as limited, which will be part of future work. In addition, the current activities included standing, sitting, walking, but more precise activities indicating subjects' healthy or unhealthy mode and behavior such as drinking, immobility, etc were missing from the analysis. While these activities are more difficult for recognition, they represent a bigger challenge.

In addition to dataset design, the currently used datasets in this thesis were collected in university and laboratory settings. Hence, they may not be very representative in real world scenarios such as care homes. The care home settings may include a number of furniture, other subjects, or different sources of heat. These constraints represent a great challenge, which can lead to the development of novel methods.

An aspect, which was not considered in this work, was the real time application of the analysis of human activity recognition. This analysis would lead to stronger conclusions for the suitability of the proposed frameworks.

The user acceptance questionnaire had a limitation corresponding to the limited number of participants. More precisely, the questionnaire was administered only in a single care home, where 21 older adults agreed to participate. Hence, the expressed views and expectations towards the research technologies may not be representative of the whole older population. In addition, only a limited number of male participants was recruited, which did not allow for gender-based analysis of the responses.

## 7.6 Future Directions

Human activity recognition is an important field of computer vision, which can have application in healthcare, eldercare, security, gaming, and others. Future work will consider the application of the current findings in real world scenarios in order to test their suitability.

The datasets can be expanded to include more activities such as the aforementioned activities and beyond. In addition, older adults can be recruited for participation in dataset creation, which would be more realistic for applications in real world scenarios. Even more realistically, a dataset can be derived from the daily activities of older adults in their normal settings. Hence, the performance of the activities will be considerably more realistic, which will improve the potential applications.

The questionnaire can be deployed in a number of care homes to assess a more representative view of the researched questions. More male participants, which were a few in this experiment, can be recruited. Hence, gender-based analysis of the responses can be further employed to discover the presence or absence of significant differences in the responses.

Sensor fusion, which is an emerging topic in machine learning and signal processing, can be incorporated considering that IR sensors and micro-Doppler radars were used in this thesis individually. By having two or more sensors, the fusion can ideally provide better results for some activities compared to the results from the individual sensors.

Inspired by the CVAE architecture that was used for unsupervised analysis, it is possible to develop a new NN architecture for unsupervised clustering of image data. Such NN model, is expected to have both automated unsupervised feature extraction as well as unsupervised clustering steps embedded in its structure using an appropriate objective function and training strategy.

The supervised periodic noise reduction technique can be applied to other projects, which suffer from the presence of similar vertical and horizontal periodic noise. The other projects might not be limited to LRIR sensing data, but can also include camera images. Hence, the proposed method can be evaluated and verified on data derived from different sources.

The local DCT-based method and local entropy-based method were both used for 2D images derived from micro-Doppler radars for unsupervised learning. The robustness of both methods was evaluated with the unsupervised metric Dunn's index. The methods can also be applied in other applications including supervised learning. Then, classification accuracy can be used as a supervised model evaluation metric.

## 7.7 Chapter Conclusion

This thesis proposed a number of machine learning and data analysis architectures for human activity recognition using non-intrusive sensors with potential applications in healthcare and eldercare. The proposed methods were verified and compared with other methods in order to stress on their effectiveness. Additionally, these contributions are important in the field of human activity recognition for eldercare, but can be applied for a number of other applications.

The thesis has explored the periodicity in the 2D spatio-temporal maps derived from LRIR data, which posed a significant problem for the data recognition. As such, the proposed novel method for periodic noise reduction represents an important finding in the field.

The findings of this thesis bring us closer to the application of human activity recognition using non-intrusive sensors in real world scenarios. To the best of the author's knowledge, this work presented the first holistic understanding and evaluation of LRIR data for human activity recognition, and unsupervised Doppler radar based human activity recognition. As such, the discoveries and analyses outlined in this thesis bring us ahead in the path of effectively addressing the problem of an ageing population.

# Bibliography

- [1] United Nations. World Population Ageing 2020, 2020.
- [2] United Nations. World Population Ageing 2015, 2015.
- [3] D. Laar, J. Edwards, and S. Easton. The Work-Related Quality of Life (WRQoL) scale for healthcare workers. *Journal of advanced nursing*, 60:325–33, 12 2007. doi: 10.1111/j.1365-2648.2007.04409.x.
- [4] Office for National Statistics. Living longer: how our population is changing and why it matters, 2018.
- [5] Mayo Clinic. Dementia - symptoms and causes, 2021. URL <https://www.mayoclinic.org/diseases-conditions/dementia/symptoms-causes/syc-20352013>.
- [6] Alzheimer’s Society. Dementia UK report, 2014. URL <https://www.alzheimers.org.uk/about-us/policy-and-influencing/dementia-uk-report>.
- [7] Social care institute for excellence. Early signs and diagnosis of dementia, 2020. URL <https://www.scie.org.uk/dementia/symptoms/diagnosis/early-signs-of-dementia.asp>.
- [8] Epilepsy Action. Epilepsy facts and terminology, 2018. URL <https://www.epilepsy.org.uk/press/facts>.
- [9] Mayo Foundation for Medical Education and Research (MFMER). Epilepsy - symptoms and causes, 2021. URL <https://www.mayoclinic.org/diseases-conditions/epilepsy/symptoms-causes/syc-20350093>.
- [10] Chegg Inc. Learn about epilepsy, 2022. URL <https://www.chegg.com/learn/biology/introduction-to-biology/epilepsy-in-introduction-to-biology>.
- [11] NHS. Parkinson’s disease - NHS, 2019. URL <https://www.nhs.uk/conditions/parkinsons-disease/>.

- [12] National Institute on Aging. Parkinson's disease, 2017. URL <https://www.nia.nih.gov/health/parkinsons-disease>.
- [13] Parkinson's UK. The incidence and prevalence of Parkinson's in the UK, 2017.
- [14] VeryWell Health. 10+ Facts About Parkinson's Disease, 2021. URL <https://www.verywellhealth.com/facts-about-parkinsons-disease-5200700>.
- [15] Brain Research UK. Motor neurone disease, 2021. URL <https://www.brainresearchuk.org.uk/neurological-conditions/motor-neurone-disease>.
- [16] S. Mashiyama, J. Hong, and T. Ohtsuki. A fall detection system using low resolution infrared array sensor. In *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, pages 2109–2113, 2014. doi: 10.1109/PIMRC.2014.7136520.
- [17] S. Mashiyama, J. Hong, and T. Ohtsuki. Activity recognition using low resolution infrared array sensor. In *2015 IEEE International Conference on Communications (ICC)*, pages 495–500, 2015. doi: 10.1109/ICC.2015.7248370.
- [18] T. Hosono, T. Takahashi, D. Deguchi, I. Ide, H. Murase, T. Aizawa, and M. Kawade. Human tracking using a far-infrared sensor array and a thermo-spatial sensitive histogram. In *Computer Vision - ACCV 2014 Workshops*, pages 262–274, 11 2014. ISBN 978-3-319-16630-8. doi: 10.1007/978-3-319-16631-5\_20.
- [19] C. Basu and A. Rowe. Tracking motion and proxemics using thermal-sensor array, 2015.
- [20] A. Trofimova, A. Masciadri, F. Veronese, and F. Salice. Indoor human detection based on thermal array sensor data and adaptive background estimation. *Journal of Computer and Communications*, 5:16–28, 2017.
- [21] L. Tao, T. Volonakis, B. Tan, Y. Jing, K. Chetty, and M. Smith. Home activity monitoring using low resolution infrared sensor, 2018.
- [22] L. Tao, T. Volonakis, B. Tan, Z. Zhang, Y. Jing, and M. Smith. 3D convolutional neural network for home monitoring using low resolution thermal-sensor array. In *3rd IET International Conference on Technologies for Active and Assisted Living (TechAAL 2019)*, pages 1–6, 2019. doi: 10.1049/cp.2019.0100.
- [23] W. Li, Y. Xu, B. Tan, and R. Piechocki. Passive wireless sensing for unsupervised human activity recognition in healthcare. In *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 1528–1533, 2017. doi: 10.1109/IWCMC.2017.7986511.

- [24] W. Li, B. Tan, Y. Xu, and R. Piechocki. Log-likelihood clustering-enabled passive RF sensing for residential activity recognition. *IEEE Sensors Journal*, 18(13):5413–5421, 2018. doi: 10.1109/JSEN.2018.2834739.
- [25] R. Steele, A. Lo, C. Secombe, and Y. Kuen Wong. Elderly persons’ perception and acceptance of using wireless sensor networks to assist healthcare. *International Journal of Medical Informatics*, 78(12):788–801, 2009.
- [26] M. Pol, F. van Nes, M. van Hartingveldt, B. Buurman, S. de Rooij, and B. Kröse. Older people’s perspectives regarding the use of sensor monitoring in their home. *The Gerontologist*, 56(3):485–493, 2016.
- [27] T. Iio, M. Shiomi, K. Kamei, C. Sharma, and N. Hagita. Social acceptance by senior citizens and caregivers of a fall detection system using range sensors in a nursing home. *The Gerontologist*, 30(3):190–205, 2016.
- [28] A. Y. Wu and C. Munteanu. Understanding older users’ acceptance of wearable interfaces for sensor-based fall risk assessment. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.
- [29] C. Cohen, T. Kampel, and H. Verloo. Acceptability of an intelligent wireless sensor system for the rapid detection of health issues: findings among home-dwelling older adults and their informal caregivers. *Patient Prefer Adherence*, 10:1687–1695, 2016.
- [30] M. Pino, M. Boulay, F. Jouen, and A.S. Rigaud. “Are we ready for robots that care for us?” Attitudes and opinions of older adults toward socially assistive robots. *Frontiers in aging neuroscience*, 7(141), 2015.
- [31] Y. H. Wu, J. Wrobel, M. Cornuet, H. Kerhervé, S. Damnée, and A. S. Rigaud. Acceptance of an assistive robot in older adults: a mixed-method study of human-robot interaction over a 1-month period in the living lab setting. *Clinical interventions in aging*, 9:801–811, 2014.
- [32] L.M. Beuscher, J. Fan, N. Sarkar, M. S. Dietrich, P. A. Newhouse, K. F. Miller, and L. C. Mion. Socially assistive robots: Measuring older adults’ perceptions. *Journal of gerontological nursing*, 43(12):35–43, 2017.
- [33] M. Heerink, B. Kröse, B. Wielinga, and V. Evers. Measuring the influence of social abilities on acceptance of an interface robot and a screen agent by elderly users. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, pages 25–36, 2009.
- [34] S. Bedaf, P. Marti, F. Amirabdollahian, and L. de Witte. A multi-perspective evaluation of a service robot for seniors: the voice of different stakeholders. *Assistive Technology*, 13(6):592–599, 2018.

- [35] R. Khosla, K. Nguyen, and M.-T. Chu. Human robot engagement and acceptability in residential aged care. *International Journal of Human-Computer Interaction*, 33(6): 510–522, 2017.
- [36] Wing-Yue Geoffrey Louie, Derek McColl, and Goldie Nejat. Acceptance and attitudes toward a human-like socially assistive robot by older adults. *Assistive Technology*, 26: 140 – 150, 2014.
- [37] M. Heerink, B. Kröse, V. Evers, and B. Wielinga. Studying the acceptance of a robotic agent by elderly users. *International Journal of Assistive Robotics and Mechatronics*, 7:33–43, 2006.
- [38] M. Heerink, B. Kröse, V. Evers, and B. Wielinga. Assessing acceptance of assistive social agent technology by older adults: the Almere model. *Int J of Soc Robotics*, 2: 361–375, 2010.
- [39] M. Shiomi, T. Iio, K. Kamei, C. Sharma, and N. Hagita. Effectiveness of social behaviors for autonomous wheelchair robot to support elderly people in Japan. *PLoS ONE* 10, 5, 2015.
- [40] S. Park, J. Park, M. Al-masni, M. Al-antari, Md. Uddin, and T. Kim. A depth camera-based human activity recognition via deep learning recurrent neural network for health and social care services. *Procedia Computer Science*, 100, 2016.
- [41] I. Pavlidis R. Murthy and P. Tsiamyrtzis. Touchless monitoring of breathing function. In *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 1196–1199, 2004. doi: 10.1109/IEMBS.2004.1403382.
- [42] H. Elphick K. S. Tan, R. Saatchi and D. Burke. Real-time vision based respiration monitoring system. In *2010 7th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP 2010)*, pages 770–774, 2010. doi: 10.1109/CSNDSP16145.2010.5580316.
- [43] M. Goffredo, I. Bouchrika, J. Carter, and M. Nixon. Performance analysis for automated gait extraction and recognition in multi-camera surveillance. *Multimed Tools Appl*, 50, 2010.
- [44] F. Ahmad, A. Najam, and Z. Ahmed. Image-based face detection and recognition: “State of the art”, 2013.
- [45] E. Heinz, K. Kunze, M. Gruber, D. Bannach, and P. Lukowicz. Using wearable sensors for real-time recognition tasks in games of martial arts - An initial experiment. In *IEEE Symposium on Computational Intelligence and Games 2006*, 2006.

- [46] G. Nugraha A. Dwiyanoro and C. Deokjai. A simple hierarchical activity recognition system using a gravity sensor and accelerometer on a smartphone. *International Journal of Technology*, 7:831, 07 2016. doi: 10.14716/ijtech.v7i5.3460.
- [47] S. Tedesco, D. Alfieri, E. Perez-Valero, D.-S. Komaris, L. Jordan, M. Belcastro, J. Barton, L. Hennessy, and B. O’Flynn. A wearable system for the estimation of performance-related metrics during running and jumping tasks. *Applied Sciences*, 11(11), 2021. ISSN 2076-3417. doi: 10.3390/app11115258. URL <https://www.mdpi.com/2076-3417/11/11/5258>.
- [48] T. H. Jo, J. K. Ma, and S. H. Cha. Elderly perception on the Internet of Things-based integrated smart-home system. *Sensors*, 21(4), 2021.
- [49] Y. Zheng G. Rong and M. Sawan. Energy solutions for wearable sensors: A review. *Sensors*, 21(11), 2021. ISSN 1424-8220. doi: 10.3390/s21113806. URL <https://www.mdpi.com/1424-8220/21/11/3806>.
- [50] B. Fu, J. Karolus, T. Grosse-Puppenthal, J. Hermann, and A. Kuijper. Opportunities for activity recognition using ultrasound Doppler sensing on unmodified mobile phones. In *iWOAR ’15: Proceedings of the 2nd international Workshop on Sensor-based Activity Recognition and Interaction*, 2015.
- [51] J. Sim, Y. Lee, and O. Kwon. Acoustic sensor based recognition of human activity in everyday life for smart home services. *International Journal of Distributed Sensor Networks*, 2015:1–11, 09 2015. doi: 10.1155/2015/679123.
- [52] F. Alías and R. Alsina-Pagès. Review of wireless acoustic sensor networks for environmental noise monitoring in smart cities. *Journal of Sensors*, 2019, 05 2019. doi: 10.1155/2019/7634860.
- [53] B. Ugur Toreyin, E. Birey Soyer, Ibrahim Onaran, and A. Enis Cetin. Falling person detection using multi-sensor signal processing. In *2007 IEEE 15th Signal Processing and Communications Applications*, pages 1–4, 2007. doi: 10.1109/SIU.2007.4298709.
- [54] A. W. van Herwaarden, D. C. van Duyn, B. W. van Oudheusden, and P. M. Sarro. Integrated thermopile sensors. *Sensors Actuators*, A21-A23:621–630, 1989.
- [55] Panasonic. Infrared Array Sensor Grid-EYE, 2020. URL <https://industrial.panasonic.com/cdbs/www-data/pdf/ADI8000/ADI8000C66.pdf>.
- [56] W. Li, R. J. Piechocki, K. Woodbridge, C. Tang, and K. Chetty. Passive WiFi radar for human sensing using a stand-alone access point. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3):1986–1998, 2021. doi: 10.1109/TGRS.2020.3006387.

- [57] X. Li, D. Qiao, Y. Li, and H. Dai. A novel through-wall respiration detection algorithm using UWB radar. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1013–1016, 2013. doi: 10.1109/EMBC.2013.6609675.
- [58] M. Alizadeh, G. Shaker, J. Almeida, P.P. Morita, and S. Safavi-Naeini. Remote monitoring of human vital signs using mm-Wave FMCW radar. *IEEE Access*, 7:54958–54968, 2019. doi: 10.1109/ACCESS.2019.2912956.
- [59] Y. Kim and H. Ling. Human activity classification based on Micro-Doppler signatures using a Support Vector Machine. *IEEE Transactions on Geoscience and Remote Sensing*, 47(5):1328–1337, 2009. doi: 10.1109/TGRS.2009.2012849.
- [60] W. Li, B. Tan, and R. Piechocki. Passive radar for opportunistic monitoring in E-Health applications. *IEEE Journal of Translational Engineering in Health and Medicine*, 6:1–10, 2018. doi: 10.1109/JTEHM.2018.2791609.
- [61] F. Wang, M. Skubic, M. Rantz, and P. E. Cuddihy. Quantitative gait measurement with pulse-Doppler radar for passive in-home gait assessment. *IEEE transactions on bio-medical engineering*, 61(9):2434–2443, 2014.
- [62] G. Diraco, A. Leone, and P. Siciliano. A radar-based smart sensor for unobtrusive elderly monitoring in ambient assisted living applications. *Biosensors*, 7(4):55, 2017.
- [63] P. P. Markopoulos and F. Ahmad. Indoor human motion classification by L1-norm subspaces of micro-Doppler signatures. In *2017 IEEE Radar Conference (Radar-Conf)*, pages 1807–1810, 2017. doi: 10.1109/RADAR.2017.7944501.
- [64] A.-K. Seifert, L. Schäfer, M.G. Amin, and A.M. Zoubir. Subspace classification of human gait using radar Micro-Doppler signatures. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 311–315, 2018. doi: 10.23919/EUSIPCO.2018.8553592.
- [65] B. Erol and M.G. Amin. Radar data cube processing for human activity recognition using multisubspace learning. *IEEE Transactions on Aerospace and Electronic Systems*, 55(6):3617–3628, 2019. doi: 10.1109/TAES.2019.2910980.
- [66] Mu Jia, Shaoxuan Li, Julien Le Kernec, Shufan Yang, Francesco Fioranelli, and Olivier Romain. Human activity classification with radar signal processing and machine learning. In *2020 International Conference on UK-China Emerging Technologies (UCET)*, pages 1–5, 2020. doi: 10.1109/UCET51115.2020.9205461.
- [67] Y. Kim and T. Moon. Human detection and activity classification based on Micro-Doppler signatures using Deep Convolutional Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, 13(1):8–12, 2016. doi: 10.1109/LGRS.2015.2491329.

- [68] J. Zhu, H. Chen, and W. Ye. A hybrid CNN–LSTM network for the classification of human activities based on Micro-Doppler radar. *IEEE Access*, 8:24713–24720, 2020. doi: 10.1109/ACCESS.2020.2971064.
- [69] M. Wang, Y. D. Zhang, and G. Cui. Human motion recognition exploiting radar with stacked recurrent neural network. *Digital Signal Processing*, 87:125–131, 2019. ISSN 1051-2004. doi: <https://doi.org/10.1016/j.dsp.2019.01.013>. URL <https://www.sciencedirect.com/science/article/pii/S1051200418308571>.
- [70] B. Jokanovic, M. Amin, and F. Ahmad. Radar fall motion detection using deep learning. In *2016 IEEE Radar Conference (RadarConf)*, pages 1–6, 2016. doi: 10.1109/RADAR.2016.7485147.
- [71] M. S. Seyfioğlu, A. M. Özbayoğlu, and S. Z. Gürbüz. Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities. *IEEE Transactions on Aerospace and Electronic Systems*, 54(4):1709–1723, 2018. doi: 10.1109/TAES.2018.2799758.
- [72] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 0387310738.
- [73] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., 2001.
- [74] E. Barshan, A. Ghodsi, Z. Azimifar, and M. Z. Jahromi. Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds. *Pattern Recognition*, 44(7):1357–1371, 2011. ISSN 0031-3203. doi: <https://doi.org/10.1016/j.patcog.2010.12.015>. URL <https://www.sciencedirect.com/science/article/pii/S0031320310005819>.
- [75] A. Y. Chervonenkis. *Empirical Inference (Early history of support vector machines)*. Springer, 2013. ISBN 0387310738.
- [76] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152, 1992.
- [77] E. Fix and L. Joseph. Discriminatory analysis. nonparametric discrimination: Consistency properties, 1951.
- [78] Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. *Neural Comput.*, 9(7):1545–1588, October 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.7.1545. URL <https://doi.org/10.1162/neco.1997.9.7.1545>.

- [79] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.
- [80] A. Géron. *Neural networks and deep learning*. O’Reilly Media, Inc., 2018.
- [81] P. J. Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD dissertation, Harvard University, 1974.
- [82] R. Collobert and S. Bengio. Links between Perceptrons, MLPs and SVMs. In *Proceedings of the Twenty-First International Conference on Machine Learning*, page 23, New York, NY, USA, 2004. Association for Computing Machinery.
- [83] T. Hastie, R. Tibshirani, and J. Friedman. “*Unsupervised learning*” in *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., 2001.
- [84] V. Melnykov and R. Maitre. Finite mixture models and model-based clustering. *Statist. Surv.*, 4:80–116, 2010.
- [85] S. Dasgupta. A cost function for similarity-based hierarchical clustering, 2015.
- [86] V. Cohen-addad, V. Kanade, F. Mallmann-trenn, and C. Mathieu. Hierarchical clustering: Objective functions and algorithms. *J. ACM*, 66(4), June 2019. ISSN 0004-5411. doi: 10.1145/3321386. URL <https://doi.org/10.1145/3321386>.
- [87] D. Reynolds. *Gaussian Mixture Models*. Springer US, Boston, MA, 2015. ISBN 978-1-4899-7488-4. doi: 10.1007/978-1-4899-7488-4\_196. URL [https://doi.org/10.1007/978-1-4899-7488-4\\_196](https://doi.org/10.1007/978-1-4899-7488-4_196).
- [88] J. Li and H. Zha. Two-way Poisson mixture models for simultaneous document classification and word clustering. *Computational Statistics & Data Analysis*, 50(1):163–180, 2006.
- [89] A. Banerjee, I. Dhillon, J. Ghosh, and S. Sra. Clustering on the unit hypersphere using von Mises-Fisher distributions. *Journal of Machine Learning Research*, 6:1345–1382, 2005.
- [90] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1):32–40, 1975. doi: 10.1109/TIT.1975.1055330.
- [91] R. P. W. Duin, A. L. N. Fred, M. Loog, and E. Pekalska. Mode seeking clustering by KNN and mean shift evaluated. In G. Gimel’farb, E. Hancock, A. Imiya, A. Kuijper, M. Kudo, S. Omachi, T. Windeatt, and K. Yamada, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, pages 51–59, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-34166-3.
- [92] H. Sasaki, T. Kanamori, A. Hyvarinen, G. Niu, and M. Sugiyama. Mode-seeking clustering and density ridge estimation via direct estimation of density-derivative-ratios. *Journal of Machine Learning Research*, 18:1–47, 2018.

- [93] F. Zhu, L. Shao, J. Xie, and Y. Fang. From handcrafted to learned representations for human action recognition. *Image Vision Comput.*, 55(P2):42–52, November 2016. ISSN 0262-8856. doi: 10.1016/j.imavis.2016.06.007. URL <https://doi.org/10.1016/j.imavis.2016.06.007>.
- [94] G. W. Taylor, R. Fergus, Y. LeCun, and C. Bregler. Convolutional learning of spatio-temporal features. In *Proceedings of the 11th European Conference on Computer Vision: Part VI*, pages 140–153, Berlin, Heidelberg, 2010. Springer-Verlag.
- [95] S. Ji, W. Xu, M. Yang, and K. Yu. 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1): 221–231, 2013. doi: 10.1109/TPAMI.2012.59.
- [96] T. Kohonen. *Self-Organizing Maps*. Springer Series in Information Sciences. Springer New York Inc., New York, NY, USA, 1995.
- [97] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 11:3371–3408, December 2010. ISSN 1532-4435.
- [98] D. P. Kingma and M. Welling. Auto-encoding variational bayes, 2014.
- [99] K. J. Cios and G. William Moore. Uniqueness of medical data mining. *Artif. Intell. Med.*, 26(1–2):1–24, September 2002. ISSN 0933-3657. doi: 10.1016/S0933-3657(02)00049-0. URL [https://doi.org/10.1016/S0933-3657\(02\)00049-0](https://doi.org/10.1016/S0933-3657(02)00049-0).
- [100] Erica Briscoe and Jacob Feldman. Conceptual complexity and the bias/variance tradeoff. *Cognition*, 118(1):2–16, 2011. ISSN 0010-0277. doi: <https://doi.org/10.1016/j.cognition.2010.10.004>. URL <https://www.sciencedirect.com/science/article/pii/S0010027710002295>.
- [101] S. Fortmann-Roe. Understanding the bias-variance tradeoff, 2012. URL <http://scott.fortmann-roe.com/docs/BiasVariance.html>.
- [102] T. Hastie, R. Tibshirani, and J. Friedman. “*Model Assessment and Selection*” in *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., 2001.
- [103] G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978. doi: 10.1214/aos/1176344136. URL <https://doi.org/10.1214/aos/1176344136>.
- [104] H. Akaike. *Information Theory and an Extension of the Maximum Likelihood Principle*. Springer New York, New York, NY, 1973.

- [105] C. D. Brown and H. T. Davis. Receiver operating characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, 80(1):24–38, 2006. ISSN 0169-7439. doi: <https://doi.org/10.1016/j.chemolab.2005.05.004>. URL <https://www.sciencedirect.com/science/article/pii/S0169743905000766>.
- [106] R. Trevethan. Sensitivity, specificity, and predictive values: Foundations, plia-bilities, and pitfalls in research and practice. *Frontiers in Public Health*, 5:307, 2017. ISSN 2296-2565. doi: [10.3389/fpubh.2017.00307](https://doi.org/10.3389/fpubh.2017.00307). URL <https://www.frontiersin.org/article/10.3389/fpubh.2017.00307>.
- [107] D. L. Streiner and J. Cairney. What’s under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry*, 52(2):121–128, 2007. doi: [10.1177/070674370705200210](https://doi.org/10.1177/070674370705200210). URL <https://doi.org/10.1177/070674370705200210>. PMID: 17375868.
- [108] J. Davis and M. Goadrich. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 233–240, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [109] M. Rehmsmeier T. Saito. The Precision-Recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*, 10(3), 2015. doi: <https://doi.org/10.1371/journal.pone.0118432>.
- [110] Y. Sasaki. The truth of the F-measure. *Teach Tutor Mater*, 01 2007.
- [111] C. Yuan and H. Yang. Research on K-Value selection method of K-Means clustering algorithm. *J*, 2(2):226–235, 2019. ISSN 2571-8800. doi: [10.3390/j2020016](https://doi.org/10.3390/j2020016). URL <https://www.mdpi.com/2571-8800/2/2/16>.
- [112] T. Thinsungnoen, N. Kaoungku, P. Durongdumronchai, K. Kerdprasop, and N. Kerdprasop. The clustering validity with Silhouette and Sum of Squared Errors. In *Proceedings of the 3rd International Conference on Industrial Application Engineering 2015*, pages 44–51, 01 2015. doi: [10.12792/liciae2015.012](https://doi.org/10.12792/liciae2015.012).
- [113] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20: 53–65, 1987. ISSN 0377-0427. doi: [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7). URL <https://www.sciencedirect.com/science/article/pii/S0377042787901257>.
- [114] D. L. Davies and D. W. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227, 1979. doi: [10.1109/TPAMI.1979.4766909](https://doi.org/10.1109/TPAMI.1979.4766909).

- [115] J. C. Dunn. Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, 4(1):95–104, 1974. doi: 10.1080/01969727408546059. URL <https://doi.org/10.1080/01969727408546059>.
- [116] T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-Simulation and Computation*, 3(1):1–27, 1974.
- [117] A Rosenberg and J. Hirschberg. V-measure: A conditional entropy-based external cluster evaluation measure. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 410–420, Prague, Czech Republic, June 2007. Association for Computational Linguistics. URL <https://aclanthology.org/D07-1043>.
- [118] S. Peek, E. Wouters, J. van Hoof, K. Luijkx, H. Boeije, and H. Vrijhoef. Factors influencing acceptance of technology for aging in place: A systematic review. *International Journal of Medical Informatics*, 84(4), 2014.
- [119] R. Al-Shaqi, M. Mourshed, and Y. Rezgui. Progress in ambient assisted systems for independent living by the elderly. *SpringerPlus*, 5(624), 2016.
- [120] A. Puri, B. Kim, O. Nguyen, P. Stolee, J. Tung, and J Lee. User acceptance of wrist-worn activity trackers among community-dwelling older adults: Mixed method study. *JMIR mHealth and uHealth*, 5(11), 2017.
- [121] Yang Che-Chang and Yeh-Liang Hsu. A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors*, 10, 08 2010. doi: 10.3390/s100807772.
- [122] Z. Wang, Z. Yang, and T. Dong. A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time. *Sensors*, 17(2), 2017.
- [123] Kiran Maini Gerhardsson and Thorbjörn Laike. User acceptance of a personalised home lighting system based on wearable technology. *Applied Ergonomics*, 96:103480, 2021. ISSN 0003-6870. doi: <https://doi.org/10.1016/j.apergo.2021.103480>. URL <https://www.sciencedirect.com/science/article/pii/S0003687021001277>.
- [124] D. Dimitrov. Medical Internet of Things and big data in healthcare. *Healthcare informatics research*, 22(3), 2016.
- [125] Robert Steele, Chris Secombe, and Wayne Brookes. Using wireless sensor networks for aged care: The patient’s perspective. In *2006 Pervasive Health Conference and Workshops*, pages 1–10, 2006. doi: 10.1109/PCTHEALTH.2006.361661.

- [126] A. Hall, C. Brown Wilson, E. Stanmore, and C. Todd. Implementing monitoring technologies in care homes for people with dementia: A qualitative exploration using normalization process theory. *International Journal of Nursing Studies*, 72, 2017.
- [127] Y. Gu, A. Lo, and I. Niemegeers. A survey of indoor positioning systems for wireless personal networks. *IEEE Communications Surveys & Tutorials*, 11(1), 2009.
- [128] T. Thayaparan, L. Stankovic, and I. Djurovic. Micro-Doppler-based target detection and feature extraction in indoor and outdoor environments. *Journal of the Franklin Institute*, 345(6), 2008.
- [129] M. Baig, S. Afifi, H. GholamHosseini, and F. Mirza. A systematic review of wearable sensors and IoT-Based monitoring applications for older adults - a focus on ageing population and independent living. *Journal of medical systems*, 43(8), 2019.
- [130] S. Vishwakarma, W. Li, C. Tang, K. Woodbridge, R. Adve, and K. Chetty. Simhumalator: An open source wifi based passive radar human simulator for activity recognition, 2021.
- [131] T. G. Stavropoulos, A. Papastergiou, L. Mpaltadoros, S. Nikolopoulos, and I. Kompatsiaris. IoT wearable sensors and devices in elderly care: A literature review. *Sensors (Basel, Switzerland)*, 20(10), 2020.
- [132] J. Abdi, A. Al-Hindawi, T. Ng, and M. P. Vizcaychipi. Scoping review on the use of socially assistive robot technology in elderly care. *BMJ Open*, 8(2), 2018. ISSN 2044-6055. doi: 10.1136/bmjopen-2017-018815. URL <https://bmjopen.bmj.com/content/8/2/e018815>.
- [133] C. Papadopoulos, R. Koulouglioti, and Lazzarino S. Ali. Enablers and barriers to the implementation of socially assistive humanoid robots in health and social care: a systematic review. *BMJ Open*, 10, 2020.
- [134] L. Hung, C. Liu, E. Woldum, A. Au-Yeung, A. Berndt, C. Wallsworth, N. Horne, M. Gregorio, J. Mann, and H. Chaudhury. The benefits of and barriers to using a social robot PARO in care settings: A scoping review. *BMC Geriatr*, 19(232), 2019.
- [135] E. J. Callahan, D. M. Hilty, and T. S. Nesbitt. Patient satisfaction with telemedicine consultation in primary care: Comparison of ratings of medical and mental health applications. *Telemedicine and e-Health*, 4(4):363–369, 1998. ISSN 1530-5627.
- [136] M. A. Loane, S. E. Bloomer, R. Corbett, D. J. Eedy, H. E. Gore, C. Mathews, K. Steele, and R. Wootton. Patient satisfaction with realtime teledermatology in Northern Ireland. *Journal of Telemedicine and Telecare*, 4(1):36–40, 1998. doi: 10.1258/1357633981931254. URL <https://doi.org/10.1258/1357633981931254>. PMID: 9640708.

- [137] A. Allen and J. Hayes. Patient satisfaction with teleoncology: A pilot study. *Telemedicine journal : the official journal of the American Telemedicine Association*, 1(1):41–6, 1995.
- [138] F. Mair and P. Whitten. Systematic review of studies of patient satisfaction with telemedicine. *BMJ (Clinical research ed.)*, 320:1517–20, 07 2000. doi: 10.1136/bmj.320.7248.1517.
- [139] D. A. Norman. *The design of everyday things*. Basic Books, [New York], 2002. ISBN 0465067107 9780465067107.
- [140] N. Ezer, A. D. Fisk, and W. A. Rogers. Attitudinal and intentional acceptance of domestic robots by younger and older adults. In *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, pages 39–48, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. ISBN 978-3-642-02710-9.
- [141] Eurobarometer 382. Public attitudes towards robots. Brussels: European Commission, 2012. URL <https://data.europa.eu/data/datasets/s1044.77.1.ebs382?locale=en>.
- [142] M. Anderson A. Smith. Americans and automation in everyday life, 2017. URL <https://www.pewresearch.org/internet/2017/10/04/automation-in-everyday-life/>.
- [143] M. M.A. de Graaf and S. Ben Allouch. Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, 61(12):1476–1486, 2013. ISSN 0921-8890. doi: <https://doi.org/10.1016/j.robot.2013.07.007>. URL <https://www.sciencedirect.com/science/article/pii/S0921889013001334>.
- [144] Y.-H. Wu, C. Fassert, and A.-S. Rigaud. Designing robots for the elderly: Appearance issue and beyond. *Archives of Gerontology and Geriatrics*, 54(1):121–126, 2012. ISSN 0167-4943. doi: <https://doi.org/10.1016/j.archger.2011.02.003>. URL <https://www.sciencedirect.com/science/article/pii/S0167494311000288>.
- [145] R. Johansson-Pajala, K. Thommes, J. Hoppe, O. Tuisku, L. Hennala, S. Pekkarinen, H. Melkas, and C. Gustafsson. Care robot orientation: What, who and how? Potential users’ perceptions. *Int J of Soc Robotics*, 12:1103–1117, 2020.
- [146] M. Mori. The uncanny valley. *Energy*, 7(4):33–35, 1970.
- [147] B. Duffy. Anthropomorphism and the social robot. *Robotics Auton. Syst.*, 42:177–190, 2003.

- [148] N. Epley, A. Waytz, and J. Cacioppo. On seeing human: A three-factor theory of anthropomorphism. *Psychological review*, 114:864–86, 2007.
- [149] C.-A. Smarr, A. Prakash, J. Beer, T. Mitzner, C. Kemp, and W. Rogers. Older adults' preferences for and acceptance of robot assistance for everyday living tasks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56, 09 2012. doi: 10.1177/1071181312561009.
- [150] L. I. Smith. A tutorial on principal components analysis. Technical report, Cornell University, USA, February 26 2002. URL [http://www.cs.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf).
- [151] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15:2006, 2004.
- [152] L. Clemmensen, T. Hastie, D. Witten, and B. Ersbøll. Sparse discriminant analysis. *Technometrics*, 53(4):406–413, 2011. doi: 10.1198/TECH.2011.08118. URL <https://doi.org/10.1198/TECH.2011.08118>.
- [153] S. Busuttill. Support vector machines. In *1st Computer Science Annual Workshop (CSAW'03)*, 2003.
- [154] P. Cunningham and S. J. Delany. *k-Nearest Neighbour classifiers: 2nd Edition (with Python examples)*, 2020.
- [155] H.C. Reddy, I.H. Khoo, and P.K. Rajan. 2-D symmetry: theory and filter design applications. *IEEE Circuits and Systems Magazine*, 3(3):4–33, 2003. doi: 10.1109/MCAS.2003.1263396.
- [156] S. Ketenci and A. Gangal. Design of Gaussian star filter for reduction of periodic noise and quasi-periodic noise in gray level images. In *2012 International Symposium on Innovations in Intelligent Systems and Applications*, pages 1–5, 2012. doi: 10.1109/INISTA.2012.6246937.
- [157] V. P. Yadav, G. Singh, Md. I. Anwar, and A. Khosla. Periodic noise removal using local thresholding. In *2016 Conference on Advances in Signal Processing (CASP)*, pages 114–117, 2016. doi: 10.1109/CASP.2016.7746148.
- [158] F. Sur and M. Grédiac. Automated removal of quasiperiodic noise using frequency domain statistics. *Journal of Electronic Imaging*, 24, 2015.
- [159] E. Weisstein. Singular Value Decomposition, 1999. URL <https://mathworld.wolfram.com/>.
- [160] S. Sharifzadeh, A. Ghodsi, L. H. Clemmensen, and B. K. Ersbøll. Sparse supervised principal component analysis (SSPCA) for dimension reduction and variable selection. *Engineering Applications of Artificial Intelligence*, 65:168–177,

2017. ISSN 0952-1976. doi: <https://doi.org/10.1016/j.engappai.2017.07.004>. URL <https://www.sciencedirect.com/science/article/pii/S0952197617301525>.
- [161] S. Sharifzadeh, J. L. Skytte, L. H. Clemmensen, and B. K. Ersbøll. DCT-based characterization of milk products using diffuse reflectance images. In *2013 18th International Conference on Digital Signal Processing (DSP)*, pages 1–6, 2013. doi: 10.1109/ICDSP.2013.6622669.
- [162] S. Sharifzadeh, I. Biro, N. Lohse, and P. Kinnell. Abnormality detection strategies for surface inspection using robot mounted laser scanners. *Mechatronics*, 51:59–74, 2018. ISSN 0957-4158. doi: <https://doi.org/10.1016/j.mechatronics.2018.03.001>. URL <https://www.sciencedirect.com/science/article/pii/S0957415818300370>.
- [163] T. Hastie, R. Tibshirani, and J. Friedman. “*Linear Methods for Classification*” in *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., 2001.
- [164] MATLAB & Simulink. Discrete Cosine Transform - MATLAB, 2019. URL <https://www.mathworks.com/help/images>.
- [165] MATLAB & Simulink. Texture Analysis Using the Gray-Level Co-Occurrence Matrix (GLCM) - MATLAB, 2019. URL <https://uk.mathworks.com/help/images/texture-analysis-using-the-gray-level-co-occurrence-matrix-glcm.html>.
- [166] J. Donahue, L. A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description, 2016.
- [167] R. Thorndike. Who belongs in the family? *Psychometrika*, 18(4):267–276, December 1953. doi: 10.1007/BF02289263. URL <https://ideas.repec.org/a/spr/psycho/v18y1953i4p267-276.html>.
- [168] C. Yuan and H. Yang. Research on K-Value selection method of K-Means clustering algorithm. *J*, 2(2):226–235, 2019. ISSN 2571-8800. doi: 10.3390/j2020016. URL <https://www.mdpi.com/2571-8800/2/2/16>.
- [169] C. Karabacak, S. Z. Gurbuz, A. C. Gurbuz, M. B. Guldogan, G. Hendeby, and F. Gustafsson. Knowledge exploitation for human Micro-Doppler classification. *IEEE Geoscience and Remote Sensing Letters*, 12(10):2125–2129, 2015. doi: 10.1109/LGRS.2015.2452311.
- [170] Y. Kim and T. Moon. Human detection and activity classification based on Micro-Doppler signatures using Deep Convolutional Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, 13(1):8–12, 2016. doi: 10.1109/LGRS.2015.2491329.

- [171] H. Du, T. Jin, Y. Song, Y. Dai, and M. Li. A three-dimensional deep learning framework for human behavior analysis using range-Doppler time points. *IEEE Geoscience and Remote Sensing Letters*, 17(4):611–615, 2020. doi: 10.1109/LGRS.2019.2930636.
- [172] Md. R. Islam Minto, B. Tan, S. Sharifzadeh, T. Riihonen, and M. Valkama. Shallow neural networks for mmWave radar based recognition of vulnerable road users. In *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pages 1–6, 2020. doi: 10.1109/CSNDSP49049.2020.9249537.
- [173] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL, 1949. ISBN 978-0-252-72548-7.
- [174] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 11:3371–3408, December 2010. ISSN 1532-4435.
- [175] A. Borghesi, A. Bartolini, M. Lombardi, M. Milano, and L. Benini. Anomaly detection using autoencoders in high performance computing systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:9428–9433, Jul 2019. ISSN 2159-5399. doi: 10.1609/aaai.v33i01.33019428. URL <http://dx.doi.org/10.1609/aaai.v33i01.33019428>.
- [176] W. Xu, S. Keshmiri, and G. Wang. Adversarially approximated autoencoder for image generation and manipulation, 2019.
- [177] Diederik P. Kingma, Danilo J. Rezende, Shakir Mohamed, and Max Welling. Semi-supervised learning with deep generative models, 2014.
- [178] C. Dong, T. Xue, and C. Wang. The feature representation ability of variational autoencoder. In *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, pages 680–684, 2018. doi: 10.1109/DSC.2018.00108.
- [179] H. Nishizaki. Data augmentation and feature extraction using variational autoencoder for acoustic modeling. *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1222–1227, 2017.
- [180] P.-A. Andersen, M. Goodwin, and O.-C. Granmo. The dreaming variational autoencoder for reinforcement learning environments. In M. Bramer and M. Petridis, editors, *Artificial Intelligence XXXV*, pages 143–155, Cham, 2018. Springer International Publishing. ISBN 978-3-030-04191-5.
- [181] I. T. Jolliffe. *Principal Component Analysis*. Springer Verlag, 1986.

- [182] J. Yang, D. Zhang, A.F. Frangi, and J. Yang. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):131–137, 2004. doi: 10.1109/TPAMI.2004.1261097.
- [183] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 0387310738.
- [184] L. J. P. van der Maaten and G. E. Hinton. Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, 2008.
- [185] J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.
- [186] J. B. Kruskal. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29(2):115–129, 1964.
- [187] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000. ISSN 0036-8075. doi: 10.1126/science.290.5500.2323. URL <https://science.sciencemag.org/content/290/5500/2323>.
- [188] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.*, 13:319–340, 1989.
- [189] W. Louie, D. McColl, and G. Nejat. Acceptance and attitudes toward a human-like socially assistive robot by older adults. *Assistive Technology*, 26:140 – 150, 2014.
- [190] L. Sinnema and M. Alimardani. The attitude of elderly and young adults towards a humanoid robot as a facilitator for social interaction. In *Social Robotics*, pages 24–33, Cham, 2019. Springer International Publishing. ISBN 978-3-030-35888-4.
- [191] C. Spearman. The proof and measurement of association between two things. *Am J Psychol*, 15:72–101, 1904. doi: 10.1136/bmj.320.7248.1517.
- [192] M. Kendall. A new measure of rank correlation. *Biometrika*, 30:1–2, 1938. doi: 10.1093/biomet/30.1-2.81.
- [193] M. McCann, M. Donnelly, and D. O’Reilly. Gender differences in care home admission risk: partner’s age explains the higher risk for women. *Age and Ageing*, 41(3):416–419, 05 2012. ISSN 0002-0729. doi: 10.1093/ageing/afs022. URL <https://doi.org/10.1093/ageing/afs022>.
- [194] Softbank Robotics. NAO the humanoid and programmable robot, 2020. URL <https://www.softbankrobotics.com/emea/en/nao>.

- [195] Softbank Robotics. Pepper the humanoid and programmable robot, 2020. URL <https://www.softbankrobotics.com/emea/en/pepper>.
- [196] E. Guizzo. France Developing Advanced Humanoid Robot Romeo, 2010. URL <https://spectrum.ieee.org/france-developing-advanced-humanoid-robot-romeo>.
- [197] R.-M. Johansson-Pajala and C. Gustafsson. Significant challenges when introducing care robots in Swedish elder care. *Disability and Rehabilitation: Assistive Technology*, pages 1–11, 2020. doi: 10.1080/17483107.2020.1773549. URL <https://doi.org/10.1080/17483107.2020.1773549>. PMID: 32538206.
- [198] C. McCreddie and A. Tinker. The acceptability of assistive technology to older people. *Ageing and Society*, 25(1):91–110, 2005. doi: 10.1017/S0144686X0400248X.
- [199] A. J. Astell, C. McGrath, and E. Dove. That’s for old so and so’s!: Does identity influence older adults’ technology adoption decisions? *Ageing and Society*, 40(7): 1550–1576, 2020. doi: 10.1017/S0144686X19000230.
- [200] W. Rössler. The stigma of mental disorders. *EMBO reports*, 17(9):1250–1253, 2016. doi: <https://doi.org/10.15252/embr.201643041>. URL <https://www.embopress.org/doi/abs/10.15252/embr.201643041>.
- [201] J. Peak. All for one and one for all? Can communities in care homes support ageing in place? In *Possibilities of DDRI in Senior People’s Care (2019)*, 2019.
- [202] S. A. Brown and V. Venkatesh. Model of adoption of technology in households: A baseline model test and extension incorporating household life cycle. *MIS Quarterly*, 29(3):399–426, 2005. ISSN 02767783. URL <http://www.jstor.org/stable/25148690>.
- [203] R. Stuck and W. Rogers. Older adults’ perceptions of supporting factors of trust in a robot care provider. *Journal of Robotics*, 2018:1–11, 04 2018. doi: 10.1155/2018/6519713.
- [204] M. Lie and K. Brittain. Technology and trust: Older people’s perspectives of a home monitoring system. *Ageing and Society*, 75:1–25, 06 2015. doi: 10.1017/S0144686X15000501.
- [205] A. Laitinen, M. Niemelä, and J. Pirhonen. Demands of dignity in robotic care in advance: Recognizing vulnerability, agency, and subjectivity in robot-based, robot-assisted, and teleoperated elderly care. *Techné: Research in Philosophy and Technology*, 23, 2019. doi: 10.5840/techne20191127108.

- [206] M. L. Traeger, S. Strohkorb Sebo, M. Jung, B. Scassellati, and N. A. Christakis. Vulnerable robots positively shape human conversational dynamics in a human–robot team. *Proceedings of the National Academy of Sciences*, 117(12):6370–6375, 2020. ISSN 0027-8424. doi: 10.1073/pnas.1910402117. URL <https://www.pnas.org/content/117/12/6370>.
- [207] J. Pirhonen J. Parviainen. Vulnerable bodies in human–robot interactions: Embodiment as ethical issue in robot care for the elderly. *Transformations Journal*, 29: 104–115, 01 2017.
- [208] A. Nwosu, B. Sturgeon, T. McGlinchey, C. Goodwin, A. Behera, S. Mason, S. Stanley, and T. Payne. Robotic technology for palliative and supportive care: Strengths, weaknesses, opportunities and threats. *Palliative Medicine*, 33(8):1106–1113, 2019. doi: 10.1177/0269216319857628.
- [209] C.-A. Smarr, T. Mitzner, J. M. Beer, A. Prakash, T. L. Chen, C. Kemp, and W. Rogers. Domestic robots for older adults: Attitudes, preferences, and potential. *International Journal of Social Robotics*, 6:229–247, 2014.
- [210] A. Laitinen and J. Pirhonen. Ten forms of recognition and misrecognition in long-term care for older people. *SATS*, 20, 2019. doi: 10.1515/sats-2016-0017.

# Appendices

# Appendix A

## User Acceptance Questionnaire

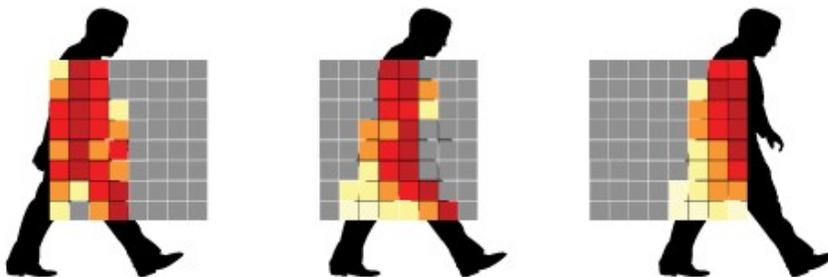
Age:

Gender: Male / Female

**Q1 How far do you agree that non-intrusive sensing technology can improve your physical well-being?**

Examples include a device attached to the wall which will call for help if you fall on the ground.

Thus, we can detect falls immediately and call a member of staff for help. [211]



- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

**Q2 How far do you agree that non-intrusive sensing technology can improve your mental well-being?**

For example, it can remind you when your favourite TV show starts.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

**Q3 What characteristics of the robot do you consider the most important?**

You can select more than one characteristic.

- Ease of use
- Cost
- Physical appearance
- Functionalities
- Usefulness

**Q4 How much do you agree with the statement – “I am likely to use robots in my home”?**

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

**Q5 Can you trust the robot for making decisions?**

- Definitely yes
- Probably yes
- Might or might not
- Probably not
- Definitely not

**Q6 What robots do you prefer based on their physical appearance?**

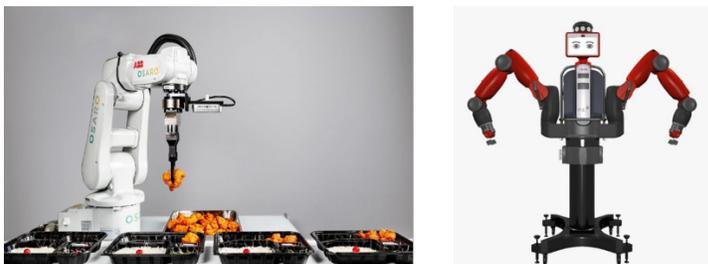
- Humanoid robots [212]-[214]



- Animal-like robots [215]-[217]



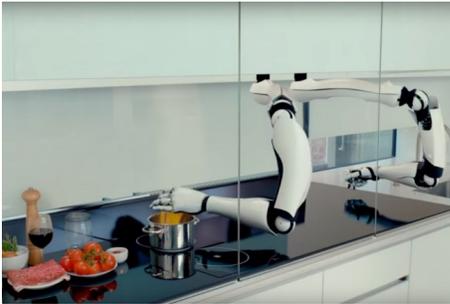
- Machine-like robots [218, 219]



**Q7 What activities do you prefer to be performed by the robot?**

You can select more than one activity.

- Cooking [220]



- Cleaning [221]



- Medicines reminder [222]



- Environments navigation [223]



- Entertainment such as dancing and singing [224]



- Communication with the robot [225]



- Communications provision with relatives [223]



**Q8 Based on the activities in the pictures, do you consider that the robot can help you have a better physical well-being?**

- Definitely yes
- Probably yes
- Might or might not
- Probably not
- Definitely not

**Q9 Based on the activities in the pictures, do you consider that the robot can help you have a better mood and mental well-being?**

- Definitely yes
- Probably yes
- Might or might not
- Probably not
- Definitely not

**Q10 Please state any additional activities that you wish to be performed by the robot.**

---

---

---

## Appendix A References

- [211] Design Online. Panasonic’s New Grid-EYE Sensor, 2012. URL <https://www.designworldonline.com/panasonics-new-grid-eye-sensor/>.
- [212] SoftBank Robotics. Nao the humanoid and programmable robot, 2021. URL <https://www.softbankrobotics.com/emea/en/nao>.
- [213] SoftBank Robotics America. Meet Pepper: The Robot Built for People, 2021. URL <https://us.softbankrobotics.com/pepper>.
- [214] Robot Xperience. Romeo humanoid robot, 2021. URL <https://robots.nu/en/robot/Romeo>.
- [215] Sony Group Corporation. Aibo, 2021. URL <https://us.aibo.com/>.
- [216] Consequential Robotics. Consequential Robotics, 2020. URL <http://consequentialrobotics.com/>.
- [217] IEEE Robots. Paro, 2021. URL <https://robots.ieee.org/robots/paro/>.
- [218] OSARI Inc. Osaro, 2020. URL <https://osaro.com/>.
- [219] CobotsGuide. Rethink Robotics: Baxter, 2021. URL <https://cobotsguide.com/2016/06/rethink-robotics-baxter/>.
- [220] Moley Robotics. The world’s first robotic kitchen, 2021. URL <https://moley.com/>.
- [221] SmartReviews. Shark ION 720 Robot Vacuum with Remote — Review, 2018. URL <https://www.smartreview.com/shark-ion-720-robot-vacuum-with-remote-review>.
- [222] IEEE Robots. Mabu - ROBOTS: Your Guide to the World of Robotics, 2018. URL <https://robots.ieee.org/robots/mabu/>.
- [223] EU Robotics. Robots that may help you in your silver age, 2016. URL <https://www.eu-robotics.net/sparc/newsroom/press/robots-that-may-help-you-in-your-silver-age.html?changelang=2>.

- [224] QUARTZ. Dancing robots have broken a Guinness World Record, 2016.  
URL <https://qz.com/753305/dancing-robots-have-broken-a-guinness-world-record/>.
- [225] Tectales. Are culturally competent robots the future in elderly care?, 2018.  
URL <https://tectales.com/bionics-robotics/are-culturally-competent-robots-the-future-in-elderly-care.html>.