Multi-sensor Activity Recognition of An Elderly Person

Saisakul Chernbumroong

Faculty of Science and Technology

Bournemouth University

A thesis submitted for the degree of

Doctor of Philosophy

January 2015

I would like to dedicate this thesis to my mother, Pranorm, and my father, Somchit Chernbumroong.

Acknowledgements

This thesis would have not been possible without the help and support from the kind people around me. Above all, I thank my father and mother for their unconditional love and support at all times, my sister for her advice and care whenever I need to retreat from studies, my brother for looking after our parents when I am far away.

This thesis would not have been possible without the help, support, and patience of my principal supervisors, Prof Hongnian Yu. I am deeply grateful to him for his tremendous support in both academics and personal life. I would like to thank him for encouraging my research and for allowing me to grow as a researcher. I owe my deepest gratitude to my second supervisor, Dr. Shuang Cang, for her guidance and support and for introducing me to the world of forecasting and classification. I would like to thank her drive to success which has motivated me to work hard and aim high.

I would like to express my gratitude to my supervisor, Prof Anthony Atkins, for his valuable support, time, and patience during my studies in Staffordshire University. I would like to thank Dr. Christopher Richardson for his constructive comments and warm encouragement. I thank Prof Bernadette Sharp and Dr. Raian Ali for their insightful comments and suggestions.

I also would like to express my gratitude to College of Arts, Media, and Technology, Chiang Mai University, Thailand, Staffordshire University, Bournemouth University, and EU Erasmus Mundus (eLINK) Projects, for giving me the opportunity to embark on this PhD journey.

I am indebted to all the participants who were involved in the project; without them my work would not have been possible. I would particularly like to thank Mrs Pan Ayumak and Mrs Juntip Tubtimsri, the representatives of Watket Elderly Community, Chiang Mai, Thailand, for their support in the participant recruitment and data collection process. Special thanks to North for helping me with the battery.

Last but not least, I would like to thank all my friends in Thailand, Stafford, and Bournemouth for their support, discussions and encouragement. I thank Rashid Bakirov and Manuel Martin for their illuminating discussions. Special thanks to my dearest friend, Joy, for always being there for me.

Abstract

The rapid increase in the number of ageing population brings major issues to health care including a rise in care cost, high demand in long-term care, burden to caregivers, and insufficient and ineffective care. Activity recognition can be used as the key part of the intelligent systems to allow elderly people to live independently at homes, reduce care cost and burden to the caregivers, provide assurance for the families, and promote better care. However, current activity recognition systems mainly focus on the technical aspect i.e. systems accuracy and neglects the practical aspects such as acceptance, usability, cost and privacy. The practicality of the system is the vital indication whether the system will be adopted. This research aims to develop the activity recognition system which considers both practical and technical aspects using multiple wrist-worn sensors.

An extensive literature review in wearable sensor based activity recognition and its applications in healthcare have been carried out. Novel multi-sensor activity recognition utilising multiple low-cost, non-intrusive, non-visual wearable sensors is proposed. The sensor fusion is performed at feature and classifier levels using the proposed feature selection and classifier combination techniques. The multi-sensor activity recognition data sets have been collected. The first data set contains data from accelerometer collected from seven young adults. The second data set contains data from accelerometer, altimeter, and temperature sensor collected from 12 elderly people in home environment performing 10 activities. The third data set contains sensor data from accelerometer, gyroscope, temperature sensor, altimeter, barometer, and light sensor worn on the users wrist and a heart rate

monitor worn over the users chest. The data set is collected from 12 elderly persons in a real home environment performing 13 activities.

This research proposes two feature selection methods, Feature Combination (FC) and Maximal Relevancy and Maximal Complementary (MRMC), based on the relationship between feature and classes as well as the relationship between a group of features and classes. The experimental studies show that the proposed techniques can select an optimum set of features from irrelevant, overlapped, and partly overlapped features. The studies also show that FC and MRMC obtain higher classification performances than popular techniques including MRMR, NMIFS, and Clamping. Two classifier combination techniques based on Genetic Algorithm (GA) are proposed. The first technique called GA based Fusion Weight (GAFW), uses GA find the optimum fusion weights. The results indicate that 99% of classifier fusion using GAFW achieves equal or higher accuracy than using only the best classifier. While other fusion weight techniques cannot guarantee accuracy improvement, GAFW is a more suitable method for determining fusion weight regardless which fusion techniques are used. Another algorithm called GA based Combination Model (GACM) is proposed to find the optimal combination between classifier, weight function, and classifier combiners. The algorithm does not only find the model which has the minimum classification error but also select the one that is simpler. Other criteria e.g. select the classifier with low computation can also be easily added to the algorithm. The results show that in general GACM can find the optimum combinations automatically. The comparison against manually selection revealed that there is no statistical significant in the performances.

Applications of the proposed work in home care and decision support system are discussed The results of this research will have a significant impact on the future health care where people can be health monitored from their homes to promote healthy living, detect any changes in behaviour, and improve quality of care.

Publications

[1] Saisakul Chernbumroong, Anthony S Atkins, and Hongnian Yu. Document management system using wireless rfid technology for intelligent healthcare operations. In *IADIS International Conference e-Health 2010*, pages 69–76. IADIS, 2010.

[2] Saisakul Chernbumroong, Anthony S Atkins, and Hongnian Yu. Perception of smart home technologies to assist elderly people. In *The fourth international conference on software, knowledge, information management and applications (SKIMA 2010)*, pages 90–97, 2010.

[3] Saisakul Chernbumroong, Anthony S Atkins, and Hongnian Yu. Activity classification using a single wrist-worn accelerometer. In *Software, Knowledge Information, Industrial Management and Applications (SKIMA), 2011 5th International Conference on*, pages 1–6. IEEE, 2011.

[4] Saisakul Chernbumroong, Anthony S Atkins, and Hongnian Yu. Wrist-worn accelerometer based activity classification using decision tree and neural network for smart health application. *The Mediterranean Journal of Computers and Networks*, 9(1):10–19, Jan 2013. ISSN 1744-2397.

[5] Saisakul Chernbumroong, Shuang Cang, Anthony Atkins, and Hongnian Yu. Elderly activities recognition and classification for applications in assisted living. *Expert Systems with Applications*, 40(5):1662–1674, 2013.

[6] Saisakul Chernbumroong, Anthony S Atkins, and Hongnian Yu. Perception of smart home technologies to assist elderly people. In *The fourth international*

conference on software, knowledge, information management and applications (SKIMA 2010), pages 90–97, 2010.

[7] S. Chernbumroong, S. Cang, and H. Yu. Ga-based classifiers fusion for multisensor activity recognition of elderly people. *Biomedical and Health Informatics, IEEE Journal of*, PP(99):1–1, 2014. ISSN 2168-2194.

[8] Saisakul Chernbumroong, Shuang Cang, and Hongnian Yu. A practical multisensor activity recognition system for home-based care. *Decision Support Systems*, 66(0):61 – 70, 2014. ISSN 0167-9236.

[9] Saisakul Chernbumroong, Shuang Cang, and Hongnian Yu. Maximum relevancy maximum complementary feature selection for multi-sensor activity recognition. *Expert Systems with Applications*, 42(1):573 – 583, 2015. ISSN 0957-4174.

Contents

List of Figures

List of Tables

Nomenclature

Roman Symbols

Com_{f_i} Complementary score of feature i

Rel_{f_i} Relevancy score of feature i

ADL Activity of Daily Living

AGREE Agreement of the classifiers

ANN Artificial Neural Network

ANOVA Analysis of Variable

AR Activity recognition

AUC Area under ROC curve

BDM Bayesian decision making

BI best individual classifier

CFS Correlation-based Feature Selection

COM Combination of Clamping, MRMR and NMIFS

COMP Comparison with the highest confidence

DMSFE Discounted mean square forecast error

DS Dempster-Shafer

DSS Decision support system

DT Decision Tree

DWT Discrete Wavelet Transform

EKG, ECG Electrocardiogram

EMG Electromyogram

FC Feature Combination

FFT Fast Fourier Transform

FHR Fetal heart rate

FN False Negative

FP False Positive

GA Genetic Algorithm

GACM Genetic Algorithm based Combination Model

GAFW Genetic Algorithm based Fusion Weight

GPS Global Positioning System

HMM Hidden Markov Model

I-ADL Instrumental Activities of Daily Living

k-NN k-Nearest Neighbour

KM Knowledge management

LDA Linear Discriminant Analysis

LR Logistic Regression

LSM Least-squares method

MEMS Micro Electro Mechanic Systems

MI Mutual information

MLP Multi-Layer Perceptron

MM Model management

MRMC Maximal Relevance Maximal Complementary

MRMR Maximal Relevant Minimal Redundant

MSL Multiple Sensor Location

NB Naïve Bayes

NMI Normalized mutual information

NMIFS Normalized Mutual Information Feature Selection

NN Neural Networ

NNFS Neural Network Feature Selection

PCA Principal Component Analysis

RBA Rule-based algorithm

RBF Radial Basis Function network

RC Relevance-complementary score

RFID Radio Frequency Identification

RMS root-mean-square

ROC Receiver Operating Characteristic

RSSI received signal strength indication

SA Simple average

SMA Signal Magnitude Area

SSL Single Sensor Location

STD Standard deviation

SVM Support Vector Machine

TN True Negative

TP True Positive

TPS Thin Plate Spline

UI User interface management

UK United Kingdom

USA United States of America

VACO Variance-covariance

WACC Weighted accuracy

WMA Weighted Moving Average

Chapter 1

Introduction

This chapter presents the background and motivation of the research. The challenges on wearable sensor based activity recognition are identified and discussed. This is followed by the research questions and hypotheses, and contributions and novelty arises from the study. Finally, the structure of the thesis is presented.

1.1 Background and motivation

Over the past decades, there has been a significant increase in the number of people aged 65 years and over worldwide. Population ageing phenomenon is enduring and expected to continue (Figure 1.1). This is the result of the demographic transition from high to low levels of fertility and mortality [45]. In 2010, the percentage of the ageing population globally is 7.58% and is expected to rise to 16.25% by 2050. It is estimated that the population of older persons is rising by 2.6% each year which is considering faster comparing to 1.34% of the population as a whole. By 2050 nearly 1.5 billion people will age 65 years and over which are more than double of the elderly population in 2010.

The rapid growth of the ageing population has an impact on humans life in many aspects. In economic area, issues such as economic growth, taxation, pensions, labour market, etc. will be affected. In social area, population ageing will have an impact on family composition and arrangement, housing and migration. In political area, voting pattern and representation will be influenced by the

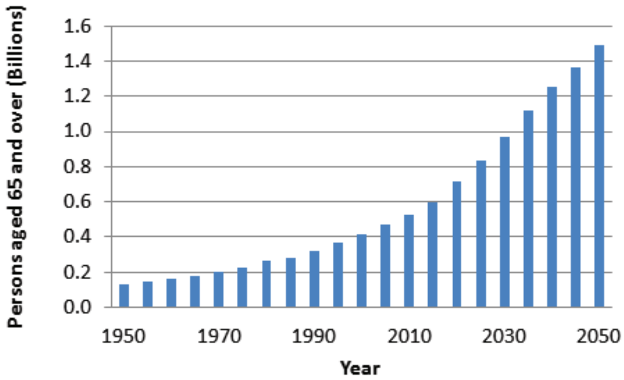

Figure 1.1: Population of persons aged 65 and over worldwide from year 1950 2050 [45].

change of the demographic [44]. Importantly, population ageing will have major impacts on health and health care as the health of older persons normally become deteriorate with increasing age. Long-term care will be more demanded. Issues such as increasing in health care expenditure, burden to carers and insufficient and ineffective care are more likely to arise.

Research studies show that the expenditure on long-term care provision in Germany, Italy, Spain, the United Kingdom (UK) and United States of America (USA) is projected to increase significantly [42, 43]. In USA, it is projected that between 2010 and 2040, the median share of household income spent on health care will increase from 10% to 19%. A steady rise in health care cost threatens to bankrupt Medicare and strain the federal budget and may potentially swarming out other government priorities [42]. Similarly, as depicted in Figure 1.2, the health care expenditure in the UK in 2009 is £119.81 billion and is expected to rise to £138 billion in 2015 [41]. It is estimated that an average cost of a four-year stay in a care home is going to double from £112,312 to £223,476 in the next 20 years [40].

The number of older people admitted to the hospital also rose at much faster rate over the last decade. A rapid increase in the number of older persons indicates

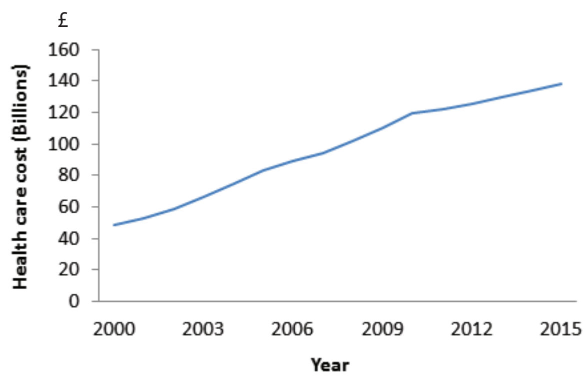

Figure 1.2: Health care expenditure in the UK during 2000 and 2015.

added burden to the carers as older persons require more intensive cares. Under general practice, the ratio of nursing home staff to patient is set to one nursing home staff to a maximum of 10 patients and the average care hour per patient per day is 3.4 hours [39]. As the number of ageing population increases, the nursing staff to patient ratio will be affected and set to increase. Also, with increasing loads for the carers, the provided care may be insufficient and inefficient or below the standard.

Due to the effects arisen from the increase in ageing population, a new model of care which supports preventive care should be encouraged. This will help prevent acute illness or delay entry into institutional care e.g. nursing homes and hospitals. The quality of life for people remaining in their own homes is generally better than for those who are institutionalised. Furthermore, the cost of care for a patient at home is also lower than the cost for institutional care [78]. Activity recognition (AR) can be used as the key part of the intelligent systems to allow elderly people to live independently at homes, reduce care cost and burden to the carers, provide a mean of assurance for family members, and promote a better care.

AR can be used to monitor elderly people from their own homes allowing them to remain at home as long as possible. It can help promote healthy living as well as detect early sign of deterioration so that earlier treatment and care can be given. Also, AR in home can be used for monitoring patient care, judging independence of elderly people, detecting changes in behaviour over time and human-computer interfaces can motivate healthy behaviour [163].

Prior works in AR are usually performed through visual sensing [86, 87, 88, 89, 160]. However, this is not practical for elderly care application due to privacy violation resulting from the use of cameras. Due to this reason, a non-visual based i.e. sensor-based AR approach is more suitable. There are two main approaches used in sensor-based AR for assisted living applications i.e. object-based and wearable sensor-based. In the object-based approach [79, 80, 81, 82, 83, 166, 173, 174], the activity is inferred from the object the user used, and changes in the environment. This approach can provide a detailed activity detection, however it suffers limitations in term of practicality such as feasibility, cost, and acceptance. For example, the object-based approach requires a vast number of sensors to be deployed in home, a sensor needs to be changed over times for some objects, specialised sensor and retrofitting may be required, in the system which uses RFID, a user needs to wear RFID glove which may not be easily accepted by the elderly people. In wearable sensor-based [85, 101, 108, 117, 135, 145], activity is determined from the sensors worn over a user's body. In some prior studies, the sensors needed to be worn over several parts of the body which may not be suitable for elderly people in term of usability and acceptance. Some studies only use sensors at a single location such as chest and waist. This reduces the issue of sensors interrupting with daily activities. However, not every location is practical to use in reality, also some locations may have higher usability and acceptance than others. The disadvantages of the single location approach are that the classification accuracy for the system which uses a single location is normally lower comparing to the system which uses sensors at multiple locations and the activities studies are still limited, often these are postures and transition activities.

Existing works in sensor-based AR for assisted living application often focuses on the technical aspect i.e. systems accuracy. However, in order to develop an activity recognition system which will be used in reality, practical issues such as acceptance, usability, and cost need to be realised. The main goal of this research is to develop the AR which takes both technical and practical aspects into account so that the system can be used in reality. Nowadays, small and inexpensive sensors are available universally due to the advance in sensor technology. In this research, small, low-cost, non-intrusive, non-stigmatise wrist-worn sensors will be used to provide a richer set of data for determining an activity of a human while reducing the number of sensors required for AR. This will allow the system to be affordable for general population and increase its acceptance and usability which is important to the elderly persons where sensors should not obstruct their daily activities and cause stigmatisation.

A summary of the major motivations to conduct this research are as follows:

- The rapid increase in the number of ageing population brings major issues to health care including a rise in care cost, high demand in long-term care, burden to caregivers, and insufficient and ineffective care. The development of AR can be used as the key part of the intelligent systems to allow elderly people to live independently at homes, reduce care cost and burden to the caregivers, provide ensuring for the families, and promote better care.

- AR will provide an instrumental tool to support preventive and home-based care. This will have a major impact on the future health care where the aim is to promote preventive care. People can be health monitored from their homes to promote healthy living as well as to be able to detect any changes so that earlier treatment and care can be given.

- At the present, sensor technologies have been advanced and are available prevalently at a lower cost. This research investigates several low-cost sensors for AR. The development of a low cost AR system will allow general population to be able to afford the technology and use to improve their life.

- The current AR systems mainly concerns the technical aspect i.e. systems accuracy and neglects the practical aspects such as acceptance and usability. The practicality of the system is the key factor which indicates whether the system will be used in reality or not. This research aims to develop the AR system which considers both practical and technical aspects so that the acceptance and usability are increased allowing the system to be used in reality.

In this research, multiple sensors are used to detect activity of an elderly person. Several challenges need to be overcome in order to successfully develop the AR method.

1. Sensor fusion

   As this research will employ several sensors, the challenges in sensor fusion are arisen. Sensor fusion can be performed at two levels: feature and classifier level [182]. At the feature level, features are calculated from different sensors and used in a classification algorithm. In wearable sensor-based AR, majority of sensor fusion are performed at feature level as it is easy to implement [79, 84, 101, 107, 117, 124, 129, 139, 152, 155, 177]. Sensor fusion at feature level is suitable when a sensor cannot be used for classification on its own or provide low classification rate. Fusing sensor at the feature level creates data-rich information for the classifier. However, sensor fusion at feature level may be difficult to perform for noncommensurate data i.e. data that are not comparable [73]. Different sensors may generate data in different form and size. For example, data obtained from camera is an image which represents in pixel, while data from accelerometer is an acceleration respective to the axis. Also, sensor may have different sampling rates or is deployed on different platforms, therefore make the sensor fusion more complicated. Another issue of feature level fusion is that it may generate a large feature space. This can lead to a common problem known as the curse of dimensionality. Also a large feature space may contain irrelevant and redundant features which directly impact the classification performance, and computation cost.

   On the contrary, fusion at the classifier level, features from each sensor are calculated and used in an individual classifier. The result from each classifier is then combined to give the final result. A few studies in wearable sensor-based AR employed this approach [146, 175, 182]. For example, two microphones and two accelerometers worn on wrists and arms are used [182]. The data fusion is performed at classifier level. The sound features are generated from microphone and used in Linear Discriminant Analysis (LDA) for classification. The features generated from accelerometers are used in Hidden Markov Model (HMM) classifier. Each classifier generates class rankings which are combined to give a final prediction. The result from each classifier can be combined using three approaches: hard, soft, and semi-soft fusion. In the hard fusion approach, the decision is compared against each other. In case of the disagreement, the result is discarded. In the soft fusion approach, the combination is based on class probabilities. Stochastic approaches such as Dempster-Shafer (DS) or statistical techniques such as product, maximum, minimum, and sum are used. In the semi-soft approach, the class probabilities are converted into rank before combination. Sensor fusion at classifier level is convenient for noncommensurate data [73]. However, if a sensor fails to detect the signal, the full benefit of sensor fusion will not be achieved. The combination of the decisions can also be difficult and complex.

2. Large feature space

   A multi-sensor activity classification system normally contains a large number of input features generated from different sensors. Using high dimension feature space increases the activity recognition models complexity and computational cost. Also, a large number of features can deteriorate the classification performance as irrelevant or redundant features might overfit the classification model or even confuse the learning algorithm [72]. Therefore, it is necessary to perform feature selection which helps to select the optimum set of features. The aim of feature selection is to identify the smallest subset of input features which explains the output classes. Feature selection can help reduce the size of feature space which leads to a reduction in computational cost and complexity in the classification system. In a large feature space that contains irrelevant and redundant features, feature selection can be used to identify a relevant feature set leads to an improvement in classification performances.

   In wearable sensor-based AR which has a large feature space, feature selection or feature reduction is performed. For example, Boosting technique is used to select features [107]. Features are selected by analysing its principal component [79, 139]. In some studies, manual analysis of features using bar chart, visualisation, or Receiver Operating Characteristic (ROC)is also carried out to select the features [124, 155]. However, these approaches have some limitations. Firstly, the manual analysis is not suitable for a large number of features. Secondly, these techniques such as Boosting, Principal Component Analysis (PCA), and Clamping only concern the relationship between the feature and the classes. They do not concern the relationship between features which may result in the selection of redundant features. In other popular feature selection techniques such as Maximal Relevant Minimal Redundant (MRMR) and Normalized Mutual Information Feature Selection (NMIFS), the relationship between features is considered. However, they only consider the one-to-one relationship i.e. feature to feature, but do not consider the relationship between a group of features to the classes. In the wrapper approach such as forward selection, backward selection, etc., only a relationship between a group of features and the classes is considered.

3. Classifier combination

   In wearable sensor-based AR, the sensor fusion at classifier level usually performed in a way that one sensor is associated with one classifier and the final result is obtained from the combination of the decision. This is, however, difficult when applying to sensors that are not useful on their own. In this research we propose to firstly fuse the sensor data at feature level then use the selected features in multiple classifiers. Classifier combination can improve the performance of activity recognition when different classifiers are superior in different classes. The main challenge is how these classifiers should be combined. Using the hard fusion approach, information regarding the posterior probability or confidence probability of classes is lost. Soft fusion using DS uses a high computational cost and counterintuitive result may occur if high conflict between evidences exists. Statistical techniques use lower computation cost, however it cannot guarantee to improve the classification performances in every case. Also, the combination model generated from the statistical technique cannot be applied to different data sets. For example, a combination model of classifier 1 and classifier 2 using product combination function on one data set may not be suitable for the other data sets.

1.2 Aims and objectives

The main aim of this research is to develop a novel method for multi-sensor based activity recognition of Activity of Daily Living (ADL) of an elderly for an intelligent assisted living system. Although there are many works in this area, there still exist unsolved problems. Especially in elderly care applications, in which factors such as cost, usability, acceptance, and privacy need to be taken into account for practical usage. The main problem this research deals with is how to recognise activities of daily living of an elderly person using non-intrusive, inexpensive wearable sensors. In this work, seven types of sensors are investigated: accelerometer, temperature sensor, altimeter, heart rate sensor, gyroscope, barometer, and light sensor. The research is focused on the study of activities that are commonly occurred in independent living situation i.e. basic ADL i.e. brushing teeth, feeding, walking, using stairs, and sleeping and instrumental ADL i.e. sweeping floor, washing dishes, ironing, watching television, scrubbing, wiping, reading, and exercising. The main questions that will be addressed in this research are:

- How to detect the interested activities of an elderly person using multiple wearable sensors worn on wrist?
- Does using multiple sensor improve classification accuracy? Does the heart rate sensor help increase the classification accuracy of the wrist-worn sensor based AR?
- How to select the features using the relationship between feature and classes as well as the relationship between a group of features and classes?
- How to combine classifiers by utilising class probabilities and are generalise enough to be apply in other data set?

The following objectives are set in order to help achieve the above aim:

1. To carry out literature reviews in wearable sensor based AR and its application in assisted living and to identify research gap (Chapter 2).

2. To design and develop hardware for sensor data collection (Chapter 3).
3. To collect sensor data in a real home setting (Chapter 3).
4. To carry out a feasibility study on using wrist worn sensor to detect activities and to identify features and techniques for data pre-processing and segmentation for multi-sensor based AR (Chapter 4).
5. To investigate and evaluate techniques for feature selection and to propose novel feature selection techniques for multi-sensor based AR (Chapter 4).
6. To investigate techniques for activity classification and to evaluate classification results generated from different techniques (Chapter 5).
7. To investigate and evaluate techniques for classifier fusion and to propose a novel classifier fusion technique based on Genetic Algorithm (GA) (Chapter 5).
8. To investigate the contributions of sensors for AR (Chapter 4).
9. To discuss the application of the proposed multi-sensor AR in assisted living (Chapter 5).

1.3 Contributions and novelty

The main contributions of the research are as follow:

1. The extensive literature review has been conducted on wearable sensor-based activity recognition and its application in assisted living. Classification regarding the approaches in AR and sensor fusion in wearable sensor-based AR has developed based on the analysis of literatures. The limitations regarding the use of wearable sensor-based AR in assisted living have been identified.
2. The multi-sensor AR data sets have been collected. This contribution is significant and valuable for other sensor-based AR works. Three data sets are collected from wearable sensors. The first data set contains data from

accelerometer collected from seven young adults performing five activities including walking, standing, sitting, running, and lying down. The second data set contains data from accelerometer, altimeter, and temperature sensor collected from 12 elderly people in home environment performing 10 activities including brushing teeth, dressing, sweeping floor, feeding, walking, walking upstairs, walking downstairs, sleeping, watching TV, and washing dishes. The third data set contains sensor data from accelerometer, gyroscope, temperature sensor, altimeter, barometer, and light sensor worn on the users wrist and a heart rate monitor worn over the users chest. The data set is collected from 12 elderly persons in a real home environment performing 13 activities of daily living including brushing teeth, exercising, feeding, ironing, reading, scrubbing, sleeping, using stairs, sweeping, walking, washing dishes, watching TV and wiping. This contribution is significant as the process in collecting activity data is time consuming and difficult for some activities. Especially in supervised learning where labelled data and experienced annotators are required. The data sets will provide valuable resources for other sensor-based AR works and machine learning society. Another contribution from the data collection is the design and development of multi-sensor instrument which is used to collect data. A part of sensors are developed using Microsoft Gadgeteer microcontroller board and sensors. The software is implemented using Matlab and C#.NET to collect the sensor data.

3. Two feature selection methods are proposed and evaluated. One of the research questions is how to select the features using the relationship between feature and classes as well as the relationship between a group of features and classes. The first feature selection method called Feature Combination (FC) is based on Clamping and forward selection. It emphasises on the performances of a combination of features rather than single feature. Experimental studies are conducted using two multi-sensor AR data sets. The results show that the proposed feature selection method can achieve higher classification accuracy comparing to Clamping, MRMR, and NMIFS methods. The second feature selection method called Maximal Relevance

Maximal Complementary (MRMC). It is based on the criteria of maximum relevance and maximum complementary of the feature. The method employed Multi-Layer Perceptron for the calculation of the relevance and complementary score. The experiments are carried out against Clamping, MRMR, and NMIFS using two well-defined problem and four benchmark classification data sets including iris, breast cancer, cardiotocography, and chess which are obtained from UCI Machine Learning Repository and one multi-sensor activity data set. The results show that MRMC is able to select relevant features in a very noisy data set containing irrelevant, highly redundant, and partly redundant features.

4. Three classification algorithms including Multi-Layer Perceptron (MLP), Radial Basis Function network (RBF), and Support Vector Machine (SVM) are investigated for multi-sensor activity recognition. An analysis of the performances of each algorithm for different activities is carried out.

5. In this research, seven classifier fusion methods including majority voting, product, summation, minimum, maximum, ranking, and weight average, and six fusion weight functions including simple average, variance-covariance, discounted mean square forecast error, unit weight, and weighted accuracy are investigated. Also, the use of GA to determine classifier fusion weight is studied. GA was employed to determine fusion weight [92, 93, 103], however the following factors were not included. Firstly, GA performance was not compared with all possible classifier combinations. Therefore it is not possible to conclude that GA could improve classifier combination accuracy as all possible combinations have not been tested. Secondly, fitness functions such as a function which reflects on the classifier combination function such as summation, minimum, maximum, product, ranking, and weighted average have not yet been investigated. Finally, their results are often compared with the mean accuracy of a set of classifiers rather than the best classifier. This may give misleading results as the mean accuracy is always equal or less than the accuracy of the best individual classifier. Therefore, this research extends previous studies in using GA for fusion weight by proposing a new technique called Genetic Algorithm based Fu-

sion Weight (GAFW). The results indicated that 99% of classifier fusion using GAFW achieves equal or higher accuracy than using only the best classifier. While other fusion weight techniques cannot guarantee accuracy improvement, GAFW is a more suitable method for determining fusion weight regardless which fusion techniques are used.

6. The classifier fusion based on GA is proposed to select optimum model of classifier, weight functions and classifier combination function called Genetic Algorithm based Combination Model (GACM). This technique is based on previous study [103] which used GA to find a combination model between features, classifiers, and classifier combiners. However, using the previous approach [103], the selected classifiers maybe not optimised for the selected features. Also, it is not clear from the study that the obtained model is the optimum comparing to manual selection. In addition, based on experiments it is found that using weight function improve classification accuracy. Therefore, a combination model between classifier, weight function, and combiners is proposed. The algorithm does not only find the model which has the minimum classification error but also select the one that is simpler i.e. use less number of classifier. The proposed technique can be extended so that other constraints maybe added such as use less classifier with high computational cost, complex weight function, etc. The results indicate that in general GACM can find the optimum combinations automatically. The comparison against manually selection revealed that there is no statistical significant in the performances. In addition, GACM allows other criteria for model selection to be added e.g. a simpler model is preferred.

7. A novel multi-sensor based AR is proposed. The AR utilises multiple low-cost, non-intrusive, non-visual wearable sensors. The sensor fusion is performed at two levels i.e. feature and classifier level using the proposed feature selection and classifier combination techniques. Using the sensors on wrist increases the acceptance and usability of the system. Also, the cost of the selected sensors is low which make it affordable for general population. The results of the study also indicate that high classification accuracy

can be achieved. The proposed multi-sensor based AR will provide an instrumental tool to support preventive and home-based care. This will have a significant impact on the future health care where people can be health monitored from their homes to promote healthy living, detect any changes in behaviour, and improve quality of care.

1.4 Organisation of thesis

This thesis contains six chapters and two appendixes. Chapter 2 presents the literature reviews in wearable sensor-based AR and its application in assisted living. It covers the existing works and details in the field including approaches, sensor types, sensor location, and sensor fusion as well as technical detail such as pre-processing, segmentation, feature extraction, feature selection and reduction, classification algorithms, and classifier combination. A variety of applications of AR has also been reviewed. In particularly, an application in assisted living where current works are reviewed and the required properties of assisted living in term of practicality and technicality are discussed.

Chapter 3 presents the strategy and approaches used to carry out the research and the system architecture of the multi-sensor AR. It covers the strategy used to collect data and characteristics of the data sets that are collected and used in the research. This chapter also explains the characteristics of sensors and platform used to develop the multi-sensor activity recognition hardware. This is followed by the techniques that are used to measure the performance of the algorithms and strategy used for comparing the results of the study with other works.

In Chapter 4, the results of a feasibility study of using the wrist-worn sensor for activity recognition are presented. Next, the results of the feature and feature selection study are presented. In this Chapter, two feature selection techniques i.e. FC and MRMC are proposed. The chapter presents the results and analysis of the proposed algorithms against the other popular feature selection techniques including MRMR, NMIFS, and Clamping on two multi-sensor activity data sets and benchmark data sets. This chapter also includes the study of contribution of each sensor for AR.

Chapter 5 presents the classification and classifier combination study. In the first part, the results of the study of three classification algorithms including MLP, RBF, and SVM are presented. This includes the analysis of the performances of each algorithm on different activities. The study is carried out on two multi-sensor AR data sets. In addition, this study also tests the hypothesis that using multiple sensors can improve classification accuracy. The second part of the chapter presents the classifier combination study. In this research, GAFW is proposed to use GA to determine the classifier fusion weights. The proposed technique extends previous studies such that all possible combinations are investigated and compared. Also, different fitness functions are investigated. In this chapter, GACM algorithm for selecting classifier, classifier fusion weights, and classifier combiners is proposed. The algorithm is designed to be adaptive for the new data set. The proposed method is compared with manual selection, and the results and analysis are presented. This chapter also presents the application of the proposed multi-sensor AR system. It describes how the proposed method can be used in a home monitoring and decision support systems.

In chapter 6, all the objectives stated at the beginning of the research are revisited. A discussion on how each objective is achieved throughout the research study. A summary of the main findings which are linked with the research questions is presented. This is followed by the research limitations as well as suggestions on how this research can be expanded into new research directions.

Appendix A shows the Barthel Index used for evaluating the independence of the participants. Appendix B shows the informed consent used to obtain the permission from the participants.

Chapter 2

Literature reviews

This research investigates wearable sensors for AR in assisted living application. An extensive review has been carried out on wearable sensor-based AR and its application in assisted living domain with the aim of identifying the research gaps in this particular field. In this chapter, a review on existing researches and state of the art in wearable sensor-based human AR field are presented. This is then linked with how AR are used in assisted living applications and what are requirements of such systems. The chapter is divided into two sections. The first section presents a review in wearable sensor-based AR. Topics reviewed including AR approaches, sensors, features used in AR system, classification algorithms, and AR applications. The second section presents applications of AR in assisted living system.

2.1 Wearable sensor-based activity recognition

The study of human AR has been carried out over the past few decades. The aim is to recognise, classify, or detect a movement, posture, or activity of a human being. Due to its advantages of applications in several domains such as surveillance, health care, security, multimedia, etc., attention on this field has been increasing. Various approaches have been investigated in order to recognise human activities. Based on the literatures, these approaches can be divided into two main categories: visual based and non-visual based AR. Visual based AR mainly focus on interpreting image information to predict activities [86, 87, 88, 89, 160], whereas the non-visual based approach utilises other type of information e.g. body movement, environment, location, etc. Non-visual based AR approach can be further divided based on how the activities are inferred: object-based [163, 165, 166, 170], location-based [172], and wearable sensor-based [85, 139, 152, 155, 177] AR. The classification of AR approaches is depicted in Figure 2.1.

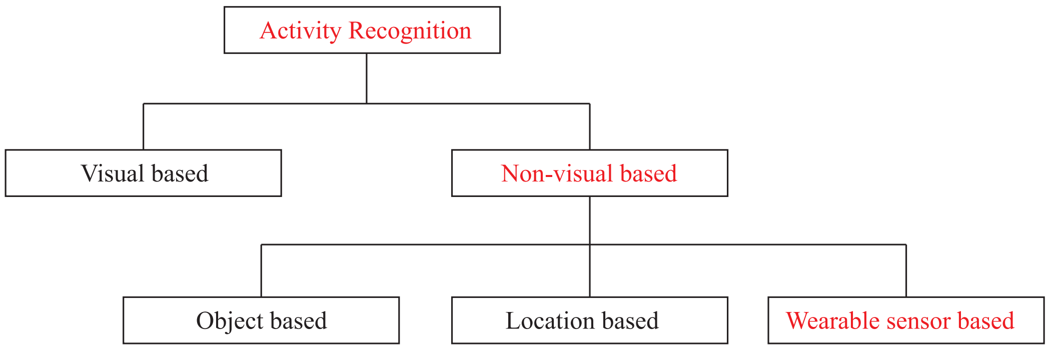

Figure 2.1: Approaches used for recognising human activities

2.1.1 History of human activity recognition

In prior studies, human AR is usually performed using visual sensing i.e. cameras. For example, a hierarchical action decision tree algorithm was proposed for video-based elder care monitoring [160]. A comprehensive review on human AR using visual sensing can be found in [161] and recently in [162]. However, visual signal interpretation can be complicated and may not be suitable in some applications i.e. health monitoring as it may perceived as intrusive and violation of a users privacy.

Over the past decade, Micro Electro Mechanic Systems (MEMS) technology has been advanced leading to an availability of small, inexpensive and low power consumption sensors. Sensor-based activity recognition has received much research attention as using sensors with sensor network and wireless technologies would allow unobtrusive and non-intrusive activity detection. A wide variety of sensors have been investigated and used as inputs for modelling human activities.

Examples of these sensors are accelerometer, microphone, gyroscope, magnetometer, Radio Frequency Identification (RFID), pressure sensor, temperature sensor, compass, heart rate monitor, Global Positioning System (GPS), etc. The approach used in sensor-based AR can be divided into three main categories based on the location of the sensors and how activities are inferred: 1) object-based, 2) location-based, and 3) wearable sensor-based AR. Table 2.1 provides a general taxonomy of approaches in sensor-based AR, showing the main concept, sensors and techniques used in each approach, as well as their prominent advantages and disadvantages.

Table 2.1: A taxonomy of approaches used in sensor-based AR

Approach	Main idea	Example sensors	Classification	Advantages	Disadvantages
Object based	Infer activity from detected objects, or changes in objects status	RFID, state change sensor, binary sensor	Rule-based, logical, reasoning, probabilistic techniques	Activities model is constructed in a semantically way	– Require installation of large set of sensors – Sensor uncertainty – Unable to detect activities which are not involved the use of objects
Location based	Predict activity based on changes in subjects location and activity-location constraint	Wifi, RFID	Bayesian network, Decision tree	Good at detecting transition activities	– Cannot detect detailed activities or activities which can possibly be performed at several locations
Wearable sensor based	Predict activity from body sensor data occurred from changes in movement	Accelerometer, gyroscope, heart rate monitor	Statistical and machine learning techniques, hierarchy, probabilistic techniques	High accuracy on activities with repetitive motion	– Sensors need to be worn on body – Difficult signal interpretation

1. Object-based AR

   infers activity from data collected from sensors installed on every objects such as cup, tooth brush, dish, etc., furniture, appliances, and also in habitant area e.g. rooms. The sensors used are often binary sensors such as reed switch, contact switch, motion sensor, etc. For example, more than 77 state-change sensors are installed on objects such as door, cupboards, food container, etc. within a home [163]. These sensors record a binary state of an object and the time that the objects state has changed. Information such as which sensor fired and temporal information i.e. before, after and duration are used for classifying activities such as preparing lunch, listening to music and taking medication etc. Another type of sensors is RFID which are normally attached to objects in order to identify the objects that a user encounters with. RFID tags placed on forks, plates, pencils, etc. are used to infer meeting, eating and working activities [164]. Similar idea are also found in [165, 166, 167, 168, 169]. Sensors such as temperature sensor, light sensor, pressure sensor are used to detect changes in environment. For example, 15 minutes differential temperature is used to identify the use of shower [170]. Other sensors include analogue sensor to monitor appliances' usage, pressure mat on a floor or chair to determine user's location. Recent works using object and environment sensors include [180, 181, 183].

The object-based AR approach exploits the semantic relationships between objects and activities to automatically classify activities. Firstly, the object is given its basic concept and associated with a higher concept using ontological technique. Based on the objects shared properties, the structure can be organised in a hierarchy manner to form super-classes and sub-classes [171]. For example, an ontology is used to represent underlying concept of objects e.g. pencil and writing tools [164]. The relationship between 'MakeTea' and 'MakeHotDrink' activities through 'DrinkType' property are defined [171]. After the concept and relationship between objects, locations, and activities are defined, the classification can be done through logical and reasoning methods such as rule-based technique, probabilistic techniques e.g. Naïve Bayes (NB), or others techniques e.g. DS, Decision Tree (DT), and sequential pattern search. A survey on an object-based activity recognition can be found in [156].

The advantage of the object-based AR approach is that the classification is done in a semantically way in which the information of how each activity model is constructed from associated objects is described. The drawbacks are firstly, the approach requires a large number of sensors attaching to objects which is infeasible and time consuming process. Secondly, uncertainty of sensors such as false start and unable to detect object etc. could result in poor recognition rate. For example, RFID tags may not be able to detect in some environments e.g. metal, liquid, or other tags not related to a particular activity can be detected by the reader because the corresponding object is accidentally close. Finally, this approach cannot detect activities which not involve the use of objects e.g. standing. Nevertheless, the last shortcoming may be overcome by adding more sensors. For example, an accelerometer is added to the AR system in [166]. They are able to combine posture activity with activity which interact with objects e.g. taking picture standing, toothbrush standing, etc. However, their system requires a user to wear RFID glove all the time which may reduce user acceptance of the system.

2. Location-based AR

   The location-based approach utilises the location of a user to infer activities. Sensors used in this approach are sensors which can identify the location such as wireless access point, RFID tags, pressure mat, and motion capture system. For example, in one study [172], the received signal strength indication (RSSI) received from wireless access point is used. Based on the location of a user, activities such as Office-to-Print-in-Room, Office-to-Seminar-in-Room, etc. can be inferred. Another study [184] uses the RFID tags placed on the objects. Instead of using only the tag ID as commonly found in the object-based AR approach, the estimated distances relative to the antenna, and the area in which the objects are using RSSI, are calculated. DT is used to learn activities such as take and return. In [179], a user wears a mobile sensor and wireless transceivers are deployed on furniture in the bedroom. They also used RSSI as input to detect bed activities e.g. lie on bellies with head turned to the side, sleep on right side with both arms down, etc.

The location-based approach is sometimes used to complement the object-based approach such as in [174, 175]. By cooperating location information, AR accuracy is improved [173]. This approach can be used to detect transition activities well. However, in the case of detailed activities e.g. reading, brushing teeth, a large number of tags or transceivers are required which would increase calculation complexity.

3. Wearable sensor-based AR

   This approach is the most popular in sensor-based AR [85, 139, 152, 155, 177]. Sensors are attached directly to the person being monitored. For example, accelerometers are used on wrist, arm, thigh, and ankle to detect daily activities such as walking, running, bicycling, etc. [177]. The glove with magnetic sensors is used to detect activities which involved the use of hand e.g. brush teeth, use hair dryer, vacuum, shave, etc. [178]. The sensors used normally have the capability of reflecting changes in different movement. Statistical techniques are often employed for classification in this approach. An activity is expressed as a set of statistical measurements which often referred as features. These features come from the statistical calculation e.g. mean, standard deviation, etc. of the collected sensor data and expressed in an m-dimensional feature space. Classification of an activity is achieved by firstly establish decision boundaries that will separate a feature space into classes regions. By studying the distribution of these features and the statistical properties of the classes, a decision on classification can be made regarding the possibility of belonging to each class [176]. Mathematic and statistic theories such as probability, distance function, etc. are utilised in the development of the classification functions. Classification models include generative models i.e. NB, HMM, etc., discriminative functions i.e. Logistic Regression (LR), SVM, etc. and Neural Network (NN).

   This approach can provide a good recognition for activities with repetitive motions [163] and high accuracy activity detection can be achieved providing sensors are installed at suitable locations. The disadvantages of this approach are such as difficulties in signal interpretation of activities with vary motions i.e. cooking and that sensors are required to be worn at all time which may interrupt or reduce mobility of a user or even obstruct daily activities routine. In some cases such as in elderly people, these sensors may be perceived as stigmatisation. Due to the appropriateness in the elderly care application studied in this research, the review will be focusing only on wearable sensor-based AR approach.

2.1.2 Multiple VS single sensor location

The approaches used in wearable sensor-based AR can be divided into two categories based on the number of location of the sensor. In earlier study of wearable sensor-based AR, multiple sensors are used in different locations of human body. This includes both the use of one type e.g. accelerometer and multiple types of sensors e.g. accelerometer and gyroscope, etc. The data collected from different parts of the body would yield a large information used for activity classification. Examples of previous works are shown in Table 2.2.

The advantage of the multiple sensor location (MSL) approach is high classification accuracy can be achieved given appropriate sensors are used in the appropriate locations. Nevertheless this approach is mainly focus on the classification accuracy, overlook practicality issues such as acceptability, cost, etc. In order for the AR to be used in reality, practicality issues need to be taken into account. The approach which uses sensors in various location of human body could obstruct or prevent the way human perform daily activities normally. Sensors worn on many parts of body may not look appealing and not easy to be accepted by a user. Hence, in later years, some of researches aim to overcome these limitations by focusing on using sensors on a single location. Majority of the studies using the single sensor location (SSL) approach used only one type of sensor i.e. accelerometer. Table 2.3 presents some of previous studies using this approach. Although, the SSL approach overcome the disadvantages in the MSL approach, some limitations still exists. Firstly, the activities recognised using this approach are still limited to mainly posture e.g. lie down, sit, stand and transition activities e.g. sit-to-stand, stand-to-sit. Secondly, the accuracy of the single location approach is still lower comparing to the MSL approach.

Table 2.2: Wearable sensor-based AR studies which use sensors in multiple locations of human body.

Author	Sensor	# Sensor	Sensor location	Recognised activities	Accuracy
Lee et al. [84]	Biaxial accelerometer, digital compass sensor, angular velocity sensor	3	Waist and leg	Sitting, standing, Different styles of walking	86.7%
Bao et al. [177]	Accelerometer	5	Arm, wrist, hip, thigh, and ankle	walking, running, climbing stairs, standing still, sitting, lying down, working on a computer, bicycling and vacuuming (N=20)	80%-95%
Ward et al. [182]	Microphones and accelerometer	4	Wrist and upper arm	Wood workshop activities (N=21)	63% - 98%
Parkka et al. [155]	Accelerometers, compass, temperature, GPS, heart rate, audio, altitude, humidity, light, pulse, EKG, skin resistance, SaO2	-	-	Lying, sitting/standing, walking, Nordic walking, running, rowing, cycling (N=7)	82%-86%
Junker et al. [186]	Inertial sensors	5	Upper arm, upper torso, wrists	Case1 (light button, hand shake, phone up, phone down, door, coin) Case2 (cutlery, drink, spoon handheld)	Case 1 98.4%. Case 2 97.4%
Yin et al. [85]	Light, temperature, microphone, two-axis accelerometer, two-axis magnetometer	5	Shoulder, waist, leg	Sitting down, walking, walking down stairs, walking upstairs, running, slipping on the ground falling down backwards, falling down forwards (N=7)	98.5%
Ermes et al. [124]	Accelerometer and GPS	3	Hip, wrist, rucksack	Lying, rowing (with a rowing machine), cycling (with an exercise bike), sitting, standing, running, Nordic walking, and walking	89%
Amft and Trster [146]	Inertial sensors, Ear microphone, stethoscope microphone, Electromyogram	6	Ear, neck, arms, wrists	eating meat lasagne with fork and knife, fetching a glass and drinking from it, eating a soup with a spoon, and eating slices of bread with one hand only	80%-90% recall
Luštrek et al. [117]	Radio tags	12	Shoulders, elbows, wrists, hips, knees and ankles	Falling, lying down, sitting down, standing/walking, sitting and lying	Over 95%
Györbíró et al. [152]	Accelerometer, a magnetometer, gyroscope	3	Wrist, hip, ankle	Resting, typing, gesticulating, walking, running, and cycling (N=6)	79.76%-81.63%
Altun et al. [139]	MTx 3-DOF orientation trackers equipped with tri-axial accelerometer, a tri-axial gyroscope, a tri-axial magnetometer	5	Knee (2), chest (1), wrist (2)	Sitting, standing, lying on back and on right side, ascending and descending stairs, standing in an elevator still and moving around, walking in a parking lot, walking on a tread mill, running on a tread mill with a speed of 8km, exercising on a stepper, exercising on a cross trainer, cycling on an exercise bike in horizontal and vertical positions, rowing, jumping, and playing basketball (N=19)	99.2%

Table 2.3: Wearable sensor-based AR studies which use sensors in single location of human body

Author	Sensor	# Sensor	Sensor location	Recognised activities	Accuracy
Najafi et al.[101]	Piezoelectric gyroscope and two accelerometers	3	Chest	Lying down, walking, as well as SiSt and stand-to-sit (StSi) transitions using different types of chairs (standard wooden chair, armchair, and upholstered chair), with and without armrests (N=6)	-
Karantonis et al.[108]	Accelerometer	1	Waist	Sit-to-stand, stand-to-sit, lying, lying-to-sit, sit-to-lying, walking (slow, normal, fast) fall (active, inactive, chair), circuit (N=12)	90.8%
Maurer et al. [75]	Light, 2D accelerometer	3	Wrist, the belt, shirt pocket, trouser pocket, backpack, and necklace	Sitting, standing, walking, ascending stairs, descending stairs and running (N=6)	78.6%–87.0%
Yang et al. [157]	Accelerometer	1	Wrist	Walking, running, scrubbing, standing, working at a computer, vacuuming, brushing teeth and sitting (N=8)	93%
Pawar et al. [135]	Electrocardiogram recorder	1	-	Sitting still, arm movement, walking and climbing down stairs, climbing upstairs, twisting movement at waist. The arm movement is a combined class of three separate movements of left arm, right arm, and both arms	92.44%
Yang et al. [116]	Accelerometer	1	Wrist	Walking, running, scrubbing, standing, working at a computer, vacuuming, brushing teeth and sitting (N=8)	95.24%
Choudhury et al.[107]	Electret microphone, Visible light phototransistor, 3-axis digital accelerometer, Digital barometer temperature, Digital IR and visible+IR light, Digital humidity/temperature, Digital Compass, 3D magnetometers, 3D gyros, and 3D compass	10	Waist	Walking, sitting, standing, taking stairs up and stairs down, taking elevator up and down, brushing teeth	93.8%
Chen at el. [130]	Accelerometer	1	Wrist	Standing, sitting, walking, running, vacuuming, scrubbing, brushing teeth, and working at a computer	92.86 ± 5.91%
Zhang et al. [158]	Accelerometer	1	Wrist	Eating and drinking	88.139%
Bonomi et al. [106]	Accelerometer	1	Lower back	Lie, sitting or standing (Sit-Stand), active standing (AS), walk, run, and cycle	91.67%
Khan et al. [129]	Accelerometer	1	Chest	Sitting, sit-stand, standing, stand-lie , lying, lie-stand, walking, walk- stand, walking-upstairs, walking downstairs, stand-sit, sit-lie, lie-sit (N=15)	97.65%
Han et al. [105]	Accelerometer	1	Waist	Standing, walking, running, falling, lying and jumping (N=6)	93.05%

2.1.3 Type of sensor

A variety of sensors have been investigated in wearable sensor-based AR research. These sensor can be separated into three categories: movement, environment, and bio sensors. Movement sensors are used to capture the changes caused by movement. The sensors should be able to react changes quickly and reflects different type of activity well. These sensors are such as accelerometer, gyroscope, angular velocity sensor, magnetometer, RFID, and orientation sensor. Environment sensors are used to measure changes in surrounding environment near the user. Examples of such sensors are light sensor, temperature sensor, humidity sensor, altimeter, proximity sensor, barometer and GPS. Bio-sensor are sensors which can be used to measure users' biological data. These sensors are such as heart rate monitor, pulse, electrocardiogram (EKG, ECG), skin resistance, electromyogram (EMG) [146], and respiratory sensor [147].

The most popular sensor used for AR is an accelerometer. Accelerometer is an instrument that measures the applied acceleration acting along the sensitive axis [14]. It is widely used for human AR purposes because of its capability to respond to both frequency and intensity of movement, and measure tile as well as body movement [13, 83, 177]. Accelerometers are relatively small and inexpensive which makes them appealing to real-life applications. There are many types of accelerometer for example, piezoresistive, piezoelectric, magnetoresistive, capacitive etc. in which different key technologies are used to measure acceleration [11]. Conceptually, a variation of the spring mass system is used. In this system, when acceleration is applied, a small mass inside the accelerometer responds by applying force to the spring, causing it to yield or compress. Measurement of the displacement of the spring is used to calculate the applied acceleration. Some studies [157, 158, 159, 185] use only accelerometers, while others e.g. [149, 151, 152, 155] use accelerometers in conjunction with other types of sensors. Accelerometer is shown to be the most information-rich and most accurate sensor for AR as it reacts fast to activity changes and reflects well the type of activity [155]. It has advantages over other sensors in quantitatively measuring human movement [150].

Gyroscope and magnetometer sensors are often used with accelerometer to provide additional movement information in term of rotation angle and direction. Gyroscope can be used to estimate the orientation and rotation of the movement. Work by [38] shows that after gyroscope and magnetometer are used with accelerometer, the accuracy of their system is increased by 17%. In wearable sensor based AR, movement sensors are the most important. Environment and bio sensors are used to provide additional information to improve accuracy in AR. For example, microphone and accelerometer are used to detect assembly-related activities [182]. The data from microphone can be used to detect surrounding noise caused by different action such as grinding, using hammer, sanding, etc. Accelerometer, microphone and light sensors are used in the AR systems [37]. Barometer can be used to collect information about pressure and temperature of the environment. Accelerometer and barometer (air pressure differential) are used to detect ambulatory movements considering vertical position shifts [35]. Combining barometer and accuracy can improve classification accuracy in child activities [36]. Temperature could be used to indicate changes in environment when performing certain activities. For example, washing dishes and brushing teeth involve a use of water, or when ironing, the temperature maybe higher than normal. Several works such as [79, 155] use the temperature sensor as part of their AR systems e.g. the difference of temperature of 15 minutes is used to determine the use of a shower [79]. Accelerometer with heart rate monitor and GPS are used in detecting work, leisure time, exercise, entertainment activities [148]. It has been shown that there is a relationship between heart rate and physical activity. Heart rate can be used to measure physical activities indirectly because heart rate is proportional to the intensity of movement and oxygen supplied to skeletal muscles [154]. For example, a subject specific regression model is used to measure the activity intensity level [34, 153]. A study in [143] show that by combining acceleration and heart rate improve accuracy of estimation of energy expenditure by 1.4%. However, the study concluded that the use of the heart rate monitor is difficult it as the users are required to wear the heart rate monitor at all times.

The choice of sensor depends on the type of activity being recognised. Using prior knowledge on the domain can improve the success of AR. For example, for detecting activities which occur in different location, environment sensors such as light, temperature, microphone can provide useful information. In detecting smoking activities as in [147], respiratory sensor which is used to detect gas exchange will provide valuable information for classification.

2.1.4 Sensor location

Studies in wearable sensor-based AR have been carried out investigating the use of sensors on different body locations. These locations include waist, leg, arm, wrist, upper arm, upper torso, shoulder, hip, ankle, chest, hand, thigh, trunk, shank, shin, feet, abdominal, and lower back. Waist is one of the popular locations when whole-body movement AR is desired. This is due to the fact that the waist is near to the center of mass of a human body, and the torso occupies the most mass of a human body therefore can better represent most of human motion [150]. The discriminatory power of different sensor locations is studied [177]. The findings indicate that thigh is the most powerful location in recognising 20 common everyday household activities e.g. running, bicycling, scrubbing, etc., followed by hip, ankle, wrist, and arm. Wrist and arm is better at discriminating activities using upper body movements. The results from their study also showed that using sensors on thigh, hip, ankle, wrist and arm gave the highest classification accuracy. Nevertheless, they suggested that effective recognition of certain everyday activities can be achieved using at least one sensor on the lower and upper body i.e. wrist and thigh or wrist and hip.

The choice of the sensor location is very important for the practical application of the activity recognition system. The location of the sensor is linked with the user acceptance of the system. In wearable sensor-based AR, a user is required to wear the sensors all the time. Certain sensor locations may prevent users from performing activities normally or may cause discomfort. Also, in certain applications such as in elderly care, these locations may be perceived as stigmatisation. Another consideration is how to attach the sensor to human body e.g. using belt clip, wrist band, strap, embedded in glove, etc. Loose attachment or unsecured fit causes vibration and displacement of sensors may produce extraneous signal artefacts therefore degrade sensing accuracy [150].

2.1.5 Sensor fusion

Many wearable sensor-based AR systems use more than one sensor to obtain information of human physical activities. This is known as sensor fusion which is when data from different sensors are integrated to extract more information [113]. It is believed that using multiple sources of information would increase recognition accuracy. Fusion of sensors can be as simple as to concatenate all data together and treat it as one single source or more complicated by associate different sources using probability theory. According to [182], there are two commonly used approaches for fusing sensor data, namely feature fusion and classifier fusion. Table 2.6 and Table 2.7 shows the level of sensor fusion of the existing works in sensor-based AR.

In feature fusion level, data from different sensors are combined and fed into a single classifier. The advantage of this approach is that more information is obtained thus recognition accuracy may be improved. However, sensor fusion at feature level may be difficult to perform for noncommensurate data i.e. data that are not comparable [73]. Different sensor may generate sensor in different form and size. For example, data obtained from camera is image which represents in pixel, while data from accelerometer is acceleration respective to the axis. Also, sensor may have different sampling rate or is deploy different platform which make the fusion more complicated Also, system complexity is increased due to larger input dimensionality [113]. An appropriate pre-processing technique e.g. data normalisation and feature reduction or selection needs to be carried out to normalise and reduce the size of the feature space. This approach is normally employed due to its simplicity. Also, this approach is suitable when the sensors are not useful on its own.

A majority of wearable-sensor based AR performed sensor fusion at feature level. For example, biaxial accelerometer, digital compass sensor, angular velocity sensor worn over waist and leg are used to detect basic activities such as sitting, standing, and different styles of walking [84]. Sensor fusion is performed at feature level where features from different sensors such as a standard deviation over 50 samples of the forward acceleration, upward acceleration, and the thigh angle, etc. are calculated and used for classification. Hierarchy based approach is used to classify the activity. For example, if thigh angle is more than 16 and the thigh angle difference is more than 70 and acceleration at x axis is more than 0.7 G, then the current activity is sitting. Fuzzy logic is used to classify different speed of walking. Kinematic sensor which is composed of one miniature piezoelectric gyroscope and two miniature accelerometers is used in [101]. The sensor is worn over the chest. The data fusion is performed at feature level where features from gyroscope and two accelerometers are calculated and feed into their hierarchy-based classification. Accelerometers at hip, wrist, arm, thigh, and ankle are used in [177]. Data fusion is performed at feature level. Features such as mean, energy, entropy and correlation from each accelerometer are calculated and fed into machine learning-based classifier. Multiple sensors worn over body is used in [155]. The sensor fusion is performed at feature level. They calculate several features based on priori information and literatures. The features are selected based on visual and statistical analysis. Machine learning-based classifiers are used in this study. Accelerometer, gyroscope, and magnetometer on five body locations are used in [139]. The sensor fusion is performed at feature level. Firstly, they calculate a large number of features from the sensors, and then use PCA to reduce the size. The features are then used in machine learning-based classifier. Based on the analysis of literatures, sensor fusion at feature level selects features based on two strategies. (1) The features are selected based on the analysis of the features. The number of features studied from this approach is small and prior knowledge or hypotheses of which feature would be useful for the activities are required. Each feature goes through analysis to discover the distinct characteristic of different activities e.g. changes in acceleration value in X-axis in certain activities, etc. This approach is normally associated with hierarchy-based classification. (2) The features are chosen based on previous studies. The number of features selected varies in size. If the feature set is small, then no selection process is carried out. Otherwise, the feature reduction technique such as PCA to reduce the feature dimension, or feature selection algorithm to select a smaller set of features is employed. In some cases, analysis using visualisation, bar chart, ROC are also used to select the features. Machine learning-based classification is often used with the second approach. Table 2.4 summarises how the features are calculated and selected in previous studies.

Table 2.4: Strategies used for sensor fusion at feature level

Feature calculation	Feature size	Feature selection	Classification	Example studies
Analysis	Small	None	Hierarchy	[84, 101]
Literatures	Small	None	Machine learning	[152, 177]
Literatures	Large	None	Machine learning	[117]
Literatures	Large	Feature selection using Boosting, forward-backward search, Correlation based Feature Selection/Feature reduction using PCA	Machine learning	[75, 79, 107, 129, 139]
Literature + prior information	Large	Analysis using visualisation, bar chart, ROC curve	Machine learning	[124, 155]

Another level in sensor fusion is at the classifier. The classification results based from different sources of information using independent classifiers are combined for final prediction. This approach suggests that there may be one classifier performs better for specific classes thus, by suitably combining multiple classifiers, accuracy could be improved. This approach is employed when it is clear how each sensor will be contributed to the classification. A limited number of wearable-sensor based AR studies performed sensor fusion at classifier level. For example, two microphones and two accelerometers worn on wrists and arms are used in [182]. Their system perform data fusion at classifier level. The sound features are generated from microphone and used in LDA for classification. The features generated from accelerometers are used in the HMM classifier. Each classifier generates class rankings which are combined to give the final prediction. An accelerometer and motion capture system are used in [175]. The sensor fusion is performed at classifier level. The accelerometer is used to obtain the motion information while the motion tracker system is used to provide the location information. The information is combined using the Bayesian technique. Inertial sensors, ear microphone, stethoscope microphone, and EMG worn on ear, neck, arm, and wrist are used [146]. The data fusion is performed at classification level. They calculate features from each sensor, and then feed to the classifier based on the feature similarity technique. Fusion strategy such as cooperative and competitive, and linear regression is used to combine events. The study evaluates three combination methods: comparison with the highest confidence (COMP), agreement of the classifiers (AGREE) and re-weighting the classifiers using LR. They present a study on two comparison schemes: competitive and supportive fusion. The results show that for competitive fusion, LR shows a better result than COMP. LR reduces more insertion errors and has fewer deletions. For supportive fusion, AGREE is better than LR as it achieves higher recall and improves precision.

There are three basic approaches in sensor fusion at classifier level [182]. The first approach simply compares the result of each classifier, discarding any result where there is disagreement. The disadvantage of this approach is that it does not take into account if one classifier may be expertise in particular classes. Second approach employs soft fusion using class probabilities. The assumption that classifiers produce continuous outputs such as class likelihood or class distances is assumed in this approach. Combining continuous results create richer sources of information for final decision making. An example of stochastic approaches appropriate for this fusion approach is DS theory which is a mathematical theory of evidence. It combines several sources of evidences associating with different probabilities and based on that predict final decision with a degree of belief. One drawback of DS is that a counterintuitive result is involved if high conflict between evidences exists. This approach has a disadvantage of high computation. Another technique is to use simple classifier combination rules such as product, summation, maximum, and minimum. The details of this technique are presented in Section 2.1.10. The third approach is a compromise between the first and second approach where class probabilities are converted into class ranking. Computational cost is reduced in this approach without discarding any specialisation that one classifier may have over another. However, some information may be lost during the conversion.

A sensor fusion can also be performed at both feature and classifier levels. However, this concept is only found in [166]. The object-based and wearable sensor-based approach are combined for AR. Accelerometers on thigh, wrist, and waist, and a RFID glove are used. The data fusion is performed at both feature and classifier level. The waist and thigh sensor data are fused at feature level where they calculate the features and feed to the DT to obtain the body state.

The wrist sensor data is used to determine hand state while an RFID sensor is used to detect the touched object. All information is later combined at the classifier level using a decision box. It is noticed that majority of wearable-sensor based AR performed fusion at feature level. Otherwise if there is a clear indication that each sensor used is capable of AR, then fusion at classifier is performed. When two AR approaches i.e. object-based and wearable-sensor are used, data fusion at classifier level is also suitable. Table 2.5 summarises the advantages and disadvantages of the sensor fusion at feature and classifier level.

Table 2.5: A summary of key advantages and disadvantages of fusion at feature and classifier level

Level	Advantages	Disadvantages
Feature	Suitable when a sensor is not useful on their own Create data-rich information for the classifier Easy to implement	Difficult to perform for noncommensurate data May generate a large feature space
Classifier	Convenient for noncommensurate data i.e. data that are not comparable due to form, size, sampling rate, platform, etc. Suitable for combine sensors used in different approaches	If a sensor fails to detect the signal, the full benefit of sensor fusion will not be achieved. If soft combination approach e.g. Dempster-Shafer is used, classifier combination can become complex

2.1.6 Pre-processing and segmentation

AR is composed of several subsystems as depicted in Figure 2.2. In each subsystem, a sub-problem is defined and has to be solved individually. Each subsystem is connected to each other and to develop a pattern recognition system, all sub-problems need to be solved. Acquisition or sensor acquisition or sensing is the first step in wearable sensor-based activity recognition. The collected sensor data is then passed to the pre-processing stage. Pre-processing makes modification to raw sensor data in order to improve them for facilitation of activity recognition. For example, raw sensor data are normally contaminated by noise. By pre-processing data, the noise is removed allowing true data to be used for classifier modelling. Several techniques can be used for pre-processing e.g. weighted average, weighted moving average, low-pass filter, high-pass filter, etc. The choice usually depends on the type of sensors.

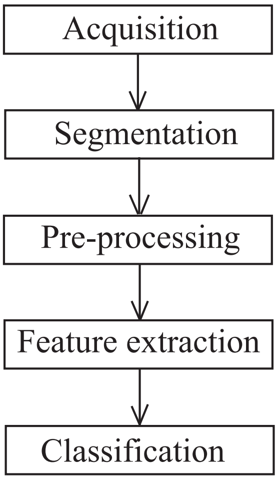

Figure 2.2: Basic AR components

Next, the processed data is passed through segmentation process where the stream of inputs is separated into a single pattern. Segmentation requires identifications of starting and ending points of a pattern which is considering one of research areas in AR. The challenging problems with segmentation are how to determine patterns starting and ending points and when to segment. Determining the beginning and end of an activity is difficult as naturally activity is interleaving and overlapping. In wearable sensor-based AR, a fix-length window based segmentation called sliding window is often used e.g. [145, 149, 155, 177]. This technique is used for separating time series data into the input vector without losing information. Given a sensor data stream, the data is divided into windows consisting of an l elements without overlapping data. The overlapping sliding window is also often used where it divides the data into windows using overlapping data from previous window. Given a sensor data stream, the window is consisted of $[l-\Delta t, l+\Delta t]$ points. Majority of wearable sensor-based AR studies employed 50% data overlapping as it is believed to reduce the edge conditions that occur when dividing the data into independent sequential windows [143].

The length of the window is a trade-off between information and resolution [144]. Normally, a short window length is more preferred. Long window may be suitable for recognising a single activity carrying out over a long period, however this does not resemble how activities occur in reality. The effect of different window length ranging from 64 – 2048 data points or 1.4 to 91 seconds using the C4.5 DT on 2 features calculating from accelerometers are studied [143]. The findings indicate that the best window length is depending on the activity being recognised. A long window length is preferred for periodic activities such as walking, riding, etc. and household activities with high motion variability such as weeding, making bed, etc. A Short window length is preferred in posture activities and short duration activities such as walking up/down the stairs, crunching, sit-up, etc. The 5.6 second window length or 256 data points is used as it allowed good performance in recognising short time and posture activities and fast interventions could be generated as soon as the activity is recognised. The choice of window length also depends on the resolution of the selected sensor. Sensor with a low resolution would require a long window length to ensure that enough information can be captured in that window.

The activity can also be segmented using activity-defined windows. In this technique, the start and the end of activity is identified by observing the changes in sensor data. For example, segmentation is performed using sound from microphone [182]. However, this technique may only be suitable for activities with apparent characteristic difference e.g. sound, movement, etc. so that the segmented data is correctly identified.

2.1.7 Features

After the data is segmented, the features can be extracted from the raw sensor data. The goal of feature extraction is to find the distinctive characteristics of an activity whose values are similar for the same activities but different for others. These characters should remain invariant to irrelevant transformations of the input [141]. These distinctive characteristics are referred as features in AR domain. Features are usually extracted from input by simply selecting some measurements e.g. temperature, pressure, etc. or by calculating some functions on the measurements e.g. mean, variance of body temperature.

A large set of different features have been studied in wearable AR researches.

Examples are mean, number of peak, standard deviation (STD), angle, energy, entropy, correlation, SMA, peak frequency, median, variance, intensity, pitch, roll, speed, zero crossing rate, etc. These features can be separated into 4 categories: heuristic, time-domain, frequency-domain, and time-frequency domain features.

Heuristic features are derived from an intuitive, fundamental understanding or prior knowledge of how a movement, posture, or activity will produce a characteristic signal. Example of these features are the difference, zero-crossing, angle and angular velocity, Signal Magnitude Area (SMA), signal vector magnitude, etc. [140]. The difference which calculates differences between sensor data can be used to distinguish activities where it is believed that different activities have noticeable strength in one or more accelerometer axis. SMA can be used to distinguish between static and dynamic activities using triaxial accelerometer signals. Different dynamic activities, e.g., running and walking, have different SMA values [145]. Intensity is used as input in the activity classification system [152]. The intensity is defined as a proportional to the variation of acceleration [152]. The study shows that the intensity of accelerometer on different location of body i.e. wrist, ankle and hip are different among resting, typing, gesticulating, walking, running, and cycling activities. Euler angles is used to describe rotations or relative orientations of the arm to identify arm movement to detect eating and drinking activities [158].

Time domain techniques use mathematic and statistic function to analyse signal data with respect to time. Using the time domain technique, the basic signal information which represents key signal characteristics can be extracted from the raw sensor data. Because of its small computational complexity cost and memory requirements, time domain techniques are often used in practical AR systems. The most popular time-domain features are mean, correlation, variance, standard deviation, kurtosis, maximum, skewness, minimum, range, and root-mean-square (RMS), respectively. The mean can be used to detect posture and discriminate type of activity i.e. static from dynamic [140]. A study by [177] shows that mean acceleration can be used to classify postures such as sitting, standing still, and lying down. The variance and standard deviation features representing variability and probability distribution of a data are also common features used in several AR systems [138, 139, 146, 155]. Standard deviation represents the amount of motion presented in the signal which can be used to differentiate activities with a very different pattern such as walking and running [136]. The range, minimum and maximum features are used where different activities possess large differences in signal such as running and standing. Signal correlation is also one of popular techniques used in AR system in which a linear relationship of two signals is expressed. It is very useful for discriminating between activities that involve translation in a single dimension [140]. For example, while both walking and running exhibit similar acceleration pattern in all dimensions, climbing stairs has a very different pattern in two dimensions [136].

The frequency domain technique is an analysis of mathematical functions on signal data with respect to frequency rather than time [137]. Information containing at different frequencies can be used to differentiate between different activities. Frequency domain techniques have been used extensively to capture the repetitive nature of a sensor signal which often correlates to the periodic nature of a specific activity such as walking and running [140]. A Fourier series such as Fast Fourier Transform (FFT) is often used to transform signal sensor data from function of time to function of frequency. A Fourier series takes a signal and decomposes it into a sum of sines and cosines of different frequencies. Fourier analysis lets certain frequency ranges to be cut off which allows intensive investigation on those frequencies we are interested. For example, the energy of accelerometer signal between 0.3 Hz and 6 Hz includes most of the information found in daily activities signal [144]. The energy of the signal can be used to represent the dynamics of the motion [136]. Hip acceleration energy can be used to classify ambulatory activities and bicycling [177]. Summation of the accelerometer signal coefficients from 0.5 Hz to 3 Hz can discriminate between activities like running and walking [140].

Frequency entropy and correlation can be used to separate activities with similar energy e.g. biking and running. Biking involves a nearly uniform circular movement of the legs, an entropy of hip acceleration in the vertical direction would be low as it contains only a single dominant frequency component at 1 Hz. Running, on the other hand, may show higher entropy as it contains more FFT frequency components between 0.5 Hz and 2 Hz [177]. In [177] work, bicycling shows low entropy hip acceleration and low arm-hip correlation while running showed higher entropy in hip acceleration and higher arm-hip movement correlation. Frequency-domains features commonly used in wearable sensor-based AR include spectral energy, information entropy, coefficient sum, dominant frequency, amplitude, and peak frequency.

As frequency-domain techniques cannot extract changes in spectral information in respect to time, using time-frequency domain techniques allow both time and frequency information to be extracted. To extract time-frequency information, a wavelet transform is carried out. Filter bank is a common technique used for Discrete Wavelet Transform (DWT) [142]. It decomposes the original signal into a detailed coefficient using a high pass filter and an approximation coefficient using a low pass filter. The higher frequency resolution, the more level the signal is decomposed allowing the signal to be decomposed into different coefficients. Wavelet features have been used in some studies. Daubechies wavelet decomposition which is a type of DWT is used in the hierarchy classification algorithm to classify posture and transition activities [101]. The study shows that DWT is a powerful technique to detect posture and walking period even when the subject is using walking aids such as a cane or walker. Other recent works which use DWT include [185] and [118].

2.1.8 Feature space manipulation

Normally in wearable sensor-based AR studies, researchers select a set of features they believe are essential for classification e.g. from previous studies or intuition. Therefore, many of the studies does not employ the feature dimension reduction or feature selection process in their systems. Nevertheless, in some systems where several sensors are used, there is a need to perform such process as the feature space becomes large. The aim of feature dimension reduction is to reduce the size of the feature space while feature selection aim to select important and relevant features for classification. This process would allow effective classification and reduce computational cost.

2.1.8.1 Feature dimension reduction

There are two popular techniques for feature dimension reduction in wearable sensor-based activity recognition: PCA and LDA. PCA is a process to reduce the variable dimensionality when correlated variables exist. These correlated variables are converted into principal components by a orthogonal transformation process. A principal component contains a linear combination of optimally-weighted of the interested variables. The first component always have the largest variance of the interested variables, followed by the second component and so on. The number of the component is less than or equal to the number of the interested variables. PCA has been used in wearable sensor-based activity recognition studies such as [132, 133, 134, 135, 136, 139]. However, PCA has shortcomings such as it only tries to preserve the data variance without cindering the discriminant ability. Other techniques e.g. Generalised Discriminant Analysis [119], are proposed to overcome this.

LDA tries to reduce the feature dimensionality while still preserving the separability of the classes. It projects the interested variables on to a line with the highest separability. There are two approaches when projecting the variables into a new space: class-dependent transformation and class-independent transformation. Class-dependent tries to maximise the ratio of between class variance to within class variance while class-independent maximises the ratio of overall variance to within class variance [131]. Wearable sensor-based AR studies which employ LDA include [127, 128, 129, 130, 136, 145, 157]. The difference between PCA and LDA is that in the process of transformation, the shape and location of the variable are changed in PCA, but only the shape in LDA [131].

2.1.8.2 Feature selection

There are three main approaches in feature selection found in wearable sensor-based activity recognition applications: intuition, filter, and wrapper. Intuition based feature selection requires a domain knowledge or understanding in what is required in the classification of the interested activities. This approach is often used in conjunction with visual inspection, statistical analysis of the features e.g. histogram, distribution graph, or observation made during activity occurrence.

Examples of studies which employed this approach are [155] and [182]. Filter based-feature selection measures the relevance between features and the outputs by using techniques such as information theory, distance, correlation, ROC, etc. Each feature is evaluated for its relevance then given a ranking score. For example, features which have the best performance in discriminating the interested activities using ROC are used [124, 125]. Many of the statistical tests are used with this approach e.g. Chi-square, T-test, etc. Mutual information (MI) is also another popular measurement used for measuring the relationship between two variables. Feature selection techniques which use MI are such as Mutual Information Based Feature Selection [20], Maximum Relevance Minimum Redundancy [62], Normalized mutual information feature selection-feature space 2 [65], etc. Some techniques are based on NN to rank the features e.g. Neural Network Feature Selection (NNFS) [19], Clamping technique [18], Constructive approach for feature selection [58], etc. The main advantages of the filter approach are simplicity, fast and independence of the classification algorithms [126]. However, most of the techniques in this approach usually consider two variables i.e. a feature and class output, thus ignoring dependencies among a set of features. This may lead to a selection of redundant features resulting in low classification accuracy.

Wrapper based-feature selection is the most popular technique for feature selection in wearable sensor-based AR. In this technique, various set of feature subsets are generated and evaluated using classification algorithms. The most optimum feature subset is selected using search techniques. Examples of this approach are forward selection [121, 122, 123], backward selection, forward-backward selection [129], exhaustive search [120], etc. In forward selection, one feature is added into a feature subset at a time and the subset is evaluated for its performance. On the other hand, backward selection removes one feature from the feature subset at a time and evaluates the subset performance. Forward-backward selection employs both directions where forward selection is carried out first then the subset is refined using backward selection. The wrapper approach is computationally extensive than the filter method, however it can provide a better result as it take into account the features dependency and interaction with the classification algorithm.

2.1.9 Classification algorithms

Majority of classification algorithms used in wearable sensor-based AR is based on statistic such as LDA or machine learning techniques such as SVM, DT, NN, NB, k-Nearest Neighbour (k-NN), HMM, etc. SVM is one of the popular techniques used. The main concept of this technique is to find non-linear decision boundaries which separate the data with the largest margin as possible. In order to help discriminate data easier, SVM maps inputs into a new higher dimensional space using some kernel functions such as linear kernel, Gaussian kernel and polynomial kernel, etc. It then finds a hyperplane with maximal margin to separate the data. The advantages of SVM are that it can produce a global optimal solution and work well on small data set [115]. [117] carries out AR using RFID tags on human body such as on shoulders, elbows, wrists, hips, knees and ankles. They perform classification using eight machine learning techniques such as DT, NB, SVM, Random Forest, etc. Their results showed that SVM achieved the highest result of on classifying falling, lying down, sitting down, standing/walking, sitting and lying activities. [139] studies human activity classification using accelerometer, gyroscope, and magnetometer. They compare different classification algorithms such as SVM, Bayesian decision making (BDM), Rule-based algorithm (RBA), Least-squares method (LSM), etc. The activities studied are mainly toward on exercise related such as exercising on a stepper, exercising on a cross trainer, playing basketball, etc. Their results show that SVM produced high accuracy in leave one subject out validation, however SVM requires longer time to train. They show that in general BDM achieve the best result and LSM is the most appropriate for online learning.

Bao et al. [177] carried out experiments on different classification techniques such as DT, decision table, instance-based learning and NB on 20 daily activities using accelerometer on arm, wrist, thigh, ankle and leg. They found that overall DT performed best. DT is a hierarchical model that uses divide-and-conquer strategy to recursively separates the input space into class regions. It composes of decision nodes and leafs in which each node has a test function. Given a node, a test function is applied to the input and depending on the output one of the branches is taken. This process is repeated until the one of the leaves is reached.

DT has several advantages over other classifiers such as easy to understand and interpret the rules, allow both numerical and discrete features to be used, quick classification for a large data set, and etc. DT and Artificial Neural Network (ANN) are compared on classification of lying, sitting/standing, walking, Nordic walking, running, rowing, cycling activities using a variety of sensors such as compass, temperature, GPS, heart rate, etc. [155]. The study shows that DT performance is better than ANN and that ANN is easily overfit due to the noisy nature of sensor data. However, DT performance degrades in live experiments [152].

ANN has also been used extensively in wearable sensor-based AR. ANN employs the structure of a neuron system where several input nodes (dendrites) are connected to several output nodes (axons). The basic processing unit in ANN is perceptron which has inputs that are associated with connection weights. The output of the network is calculated from an activation function of the weighted sum of the perceptrons that are linked to the output plus a bias weight. An activation function is usually a sigmoid function such as hyperbolic tangent, algebraic function, arctangent function, etc. The NN can be trained so that it can automatically adjust its weights to model the relationship between given inputs and outputs. The weights are updated in order to minimise the error of the output. ANN has advantages of its fast execution and work as a universal approximator in which anything learnable could be taught to the network. The drawback is that it can be slow to train however techniques such as momentum and adaptive learning can be used to improve the performance of the gradient descent. Feed-forward ANN is used to classify 15 postures and transition activates from accelerometer worn on chest [129]. ANN is used for classifying 8 ADL e.g. working at PC, vacuuming, brushing teeth, sitting, etc. from acceleration data [116]. The study shows that the model using ANN outperforms k-NN. Prior knowledge from DT and ANN are combined for classifying exercise related activities [124]. The results show that a combination of DT and ANN improves classification accuracy.

Other algorithms have also been applied in wearable sensor-based AR. [182] employs HMM for acceleration classification. Gaussian mixture is used for the observation probabilities. They modify the number of mixtures and hidden states for each class model. The activity is predicted based on the model that produces the largest log likelihood. A fuzzy basis function classifier is used for AR [130]. Hierarchical Temporal Memory which is normally used in image processing is used in [158]. Rotations or orientations of the arm from accelerometers worn on wrists are calculated as AR inputs. The idea of this technique is to construct a coincidence matrix which discovers meaningful coincidence in training data. The activity can be inferred by comparing the unknown input and the coincidence matrix.

2.1.10 Classifier combination

As it can be seen from previous section, different classification algorithms have different advantages and disadvantages. Also, due to different sensor type, sensor location, features used, some techniques may be superior to others. The hypothesis is that by combining classifier result, the performance of classification model can be improved. Although this observation seems apparent, to the best of our knowledge, work in wearable sensor-based AR has not yet been investigated on this.

A construction of classifier for combination can be carried out in several ways such as using different feature sets, training sets, classification algorithms, classification architectures, or parameter values. Classifier combination methods can be divided into different categories depending on criteria used. For example, based on output type, the combination methods can be separated into three approaches as presented in Section 2.1.5. Using type of combination criteria, the method can be divided into two approaches: static and dynamic combination. Static combination employs a rule to combine output from the classifiers. Popular classifier combination rules are product, summation, maximum, minimum, majority voting, and weighted average. The majority vote combines all the votes given by each model and selects the class which has the highest vote. Using the product rule, the classifiers' outputs are combined using a vector product. The product rule is more sensitive to objection than support where the class with low probability has more influence to the decision than the class with high probability. Using the summation rule, the classifiers' outputs are combined using the sum function and the class which has the highest maximum of the average output is selected. The summation function generates the result of the average decisions of all classifiers. This is similar to the majority voting function however continuous output i.e. class probabilities can be used in sum function. The maximum function decides the result based on the most confident classifier where it selects the class with the highest output from all the models. The minimum function combine classifiers results by selecting the class which is least objection by all the models.

The static combination approach is simple, easy to apply, and uses low computation. The disadvantage is that optimum result cannot be guaranteed and over-confidence classifier could affect the overall accuracy. For dynamic approaches, the combination can be trained so that optimum combination can be achieved. Example techniques used are such as NN, linear regression GA, etc. The dynamic approach requires higher computation cost than a fixed rule, however better performance is normally expected. For example, a method to find a combination model between features, classifier, and combiners using GA was proposed [103]. The method was evaluated on two data sets and the results show that the method outperforms other methods including single best classifier, GA-optimised weighted soft linear combiner, GA-optimised class independent soft linear combiner, and GA-based classifier selection only. However, this approach [103] has some limitations. First, since the features are determined on the fly, the optimal parameters for that features and classification algorithms may not be able to obtained. For example, C and γ parameter need to be determined before constructing the SVM classifier in order to obtain best results. Second, this method may not be suitable for the classification algorithm that requires longer time to train. Thirdly, the method may suffer high computation when involve with a large feature space and complex classification algorithm. Fourth, although the study compares the performance of the method with several other methods, it does not compare with all possible combination to demonstrate that the combination model selected by the method is optimum.

Classifier combination can be enhanced by cooperating weights. Weights can be defined so that a classifier with better performance is associated with a higher weight. Weights can be calculated using by simple techniques such as simple average, weighted accuracy, or some techniques from other domains e.g. forecasting such as variancecovariance and discounted combination methods [104]. Simple average gives the average weights to all classifiers. Variance-covariance (VACO) uses the mean square error to calculate the weights. Discounted mean square forecast error (DMSFE) is the modified version of VACO where an additional β parameter is introduced to discount the factor of the error. DMSFE is suitable for error that is associated with time such that the recent error has higher weight than the older error. Accuracy of the model can be used as weight where higher weights are given to the classifier that are more accurate.

Weights can also be obtained by learning from data set using techniques such as NN or from search techniques e.g. GA. Studies [92, 93, 103] indicate fusion weight determined by GA improve the classifier fusion accuracy. For example, classifier combination using 8-10 ensembles generated from different techniques was studied [92]. A weight combination using GA was investigated for combining several Bayesian classifiers [93]. However, some factors were not included in these studies. Firstly, the investigations were not complete as all combinations were not investigated [92, 93]. For example, six classifiers are produced, then GA is used to combine all classifiers' results. Based on this, the conclusion that GA improves classifier combination accuracy is not always true as all possible combinations have not been tested. Secondly, the weights determined are often from optimising the accuracy or error of the weighted average fusion technique i.e. $f(w) = w_1x_1 + w_2x_2 + ... + w_Kx_K$. Other fusion functions which reflect on the combination function e.g. summation, minimum, maximum, product, etc. have not been applied before. Finally, some of these results are often compared with the mean accuracy of a set of classifiers rather than to the best individual classifier. However, the mean accuracy is always equal or less than the accuracy of the best individual classifier (equal accuracy is only occurred if and only if all classifiers have the same accuracy). For example, if there are three classifiers with accuracies of 90%, 85%, 95%, the mean accuracy is 90% which is less than the best individual (95%). This weakens the conclusion that the classifier combination is better than a single classifier.

2.2 Activity recognition approaches discussion

Tables 2.6 and 2.7 classify the approaches, sensors, fusion level, classification techniques, application used in previous studies. From the tables, it can be seen that a variety of approaches have been employed. From the tables, it can be seen that using location approach for activity recognition is not popular. This is due to the difficulty in determining the exact location of a user indoor. Also, the RSSI can be impacted by furniture, objects, and layout of the house. Due to these limitations, using location approach is not practical for assisted living application. The object approach can provide good activity recognition, providing that enough sensors are installed in homes. However, this approach is not popular due to its feasibility in deploying and maintaining a large set of sensors in homes. If this approach were to used, it is recommend to perform fusion at feature level, and employ machine learning technique such as DT or reasoning technique for AR. Using wearable sensor is the most popular approach for AR. Sensor fusion using this approach can be done at feature and/or classifier level. This is depend on the nature of the sensors as discussed in Section **??**. Machine learning techniques are often used with this approach and have been shown to provide high classification accuracy. In term of assisted living application, it is suggested that wearable sensor-based at single location should be used to increase the acceptance and usability of the system.

Table 2.6: A classification of sensor-based AR works regarding their approach, sensor, fusion, classification, and application.

Approach: WS = Wearable sensor-based at single location, WM = Wearable sensor-based at multiple location Ob = Object-based, L = Location-based, V = Visual-based. Sensor: MT = Multiple type, ST = Single type. Fusion level: F = Feature level, C = Classifier level. Classification: H = Hierarchy, ML = Machine learning and statistical, L = logic reasoning, PM = Pattern matching, O = Other. Application: P = Postures, T = Transition, B = Basic activities such as walking, running, etc., ADL = Activities of daily living, F = Fall, S = Specific

Author	Approach	Sensor	Fusion level	Classification	Application	#Participants	Elderly?
Ward et al. [66]	WM	MT	C	ML	S (Kitchen)	-	-
Lee et al. [84]	WM	MT	F	H	B + P	8	N
Najafi et al. [101]	WS	MT	F	H	P+T	11, 24, 9	Y
Bao et al. [177]	WM	ST	F	ML	ADL	20	N
Tapia et al. [163]	Ob	MT	F	ML	ADL	2	Y
Wilson et al. [67]	Ob	MT	F	ML	N/A	1-3	N
Karantonis et al. [108]	WS	ST	-	H	T+F	6	N
Maurer et al. [75]	WM	MT	F	ML	B	6	N
Maurer et al. [74]	WM	MT	F	ML	B	16	N
Ward et al. [182]	WM	MT	C	ML	S (Workshop)	5	N
Yang et al. [157]	WS	ST	-	ML	ADL	7	N
Yamada et al. [164]	Ob	ST	-	L	S (Work)	-	-
Pawar et al. [135]	WS	ST	-	O	B	23	N
Junker et al. [186]	WM	ST	F	ML	S (Gesture)	4	N
Yang et al. [116]	WS	ST	-	ML	ADL	7	N
Chen et al. [130]	WS	ST	-	ML	ADL	7	N
Sanchez et al. [68]	Ob	N/A	-	ML	S (Hospital)	15	N
Yin et al. [172]	L	ST	-	ML	T	-	-
Choudhury et al. [107]	WS	MT	F	ML	ADL	15	N
Ermes et al. [124]	WM	MT	F	ML	S (Sport)	12	N
Yin et al. [85]	WM	MT	F	ML	B + F	-	-
Landwehr et al. [165]	Ob	ST	-	L	ADL	12	N
Amft et al. [146]	WM	MT	C	PM	S (Dietary)	4	N
Luštrek et al. [117]	WM	ST	F	ML	B + F	3	N
Diermaier et al. [83]	Ob	MT	F	Manual	ADL	2	Y

Table 2.7: A classification of sensor-based AR works regarding their approach, sensor, fusion, classification, and application (cont.).

Author	Approach	Sensor	Fusion level	Classification	Application	#Participants	Elderly?
Szewcyzk et al. [69]	Ob	MT	F	ML	ADL	2	N
Zhang et al. [158]	WS	ST	-	H	S (Dietary)	-	-
Cook et al. [174]	Ob	MT	F	ML	ADL	60	N
Györbíró et al. [152]	WM	MT	F	ML	ADL	-	-
Lu et al. [173]	Ob	MT	F	ML	ADL	11	N
Hong et al. [82]	Ob	MT	F	L	N/A	-	-
Bonomi et al. [106]	WS	ST	-	ML	B	15	N
Hong et al. [166]	WM+Ob	MT	F+C	ML	ADL	15	N
Khan et al. [129]	WS	ST	-	ML	P+T	6	N
Khan et al. [145]	WM	ST	F	ML	ADL	8	Y
Sarkar et al. [71]	Ob	MT	F	ML	ADL	1-2	N
Iglesias et al. [70]	Ob	MT	F	ML	ADL	24	N
Han et al. [105]	WS	ST	-	ML	B	-	-
Altun et al. [139]	WM	MT	F	ML	S (Sport)	8	N
Cheng et al. [80]	Ob	N/A	-	ML	ADL	1	N
Zhu et al. [175]	WS+Ob	MT	C	ML	B	1	N
Martine et al. [168]	V	ST	-	O	S (Care)	2	Y
Fleury et al. [79]	WS+Ob	MT	F	ML	ADL	13	N

2.2.1 Activity recognition application

2.2.1.1 Elderly care applications

Due to the strength in providing personalised support, AR has been used in many healthcare-related applications especially in elder care support, long-term health monitoring and assisting those with cognitive disorders [107]. AR enables new model of care that is a home-based preventive system which will allow people to age in their own home. The quality of life for people remaining in their own homes is generally better than for those who are institutionalised. Furthermore, the cost for institutional care can be much higher than the cost of care for a patient at home [78]. AR in home can be used for monitoring patient care, judging independence of elderly people, detecting changes in behaviour over time and human-computer interfaces can motivate healthy behaviour [163]. Moreover, other monitoring sensors data such as heart rate, temperature, pressure, etc. would allow patient to be monitored at home without disturbing their daily activities.

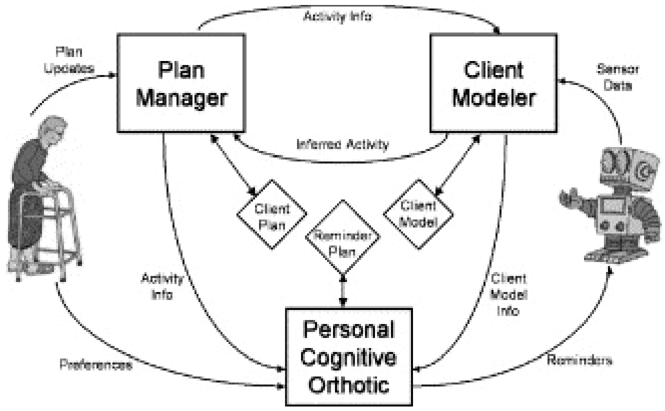

Figure 2.3: Autominder architecture [78]

A research by [78] investigates AR in order to help the elderly recognise and cope with the cognitive decline associated with illness and aging by sending adaptive personalised activity reminders. Their Autominder architecture is depicted in Figure 2.3. The system helps older persons adapt to cognitive decline and continue the satisfactory performance of their routine activities and potentially enabling them to remain in their own homes longer. In addition, automatic AR also can allow older people to live at home safely by detecting abnormal activities. If a safe and smart house can be instrumented with a sensor network, the occupants would have a better chance to live safely and independently, especially for those who suffer from severe illnesses e.g. Parkinsons or Alzheimers. Detection of unusual activities can also be used as a first indicator when the elderly develop cognitive decline or symptom of illness or even injury. Some work such as by [179] specifically investigates bedside activities in order to prevent bedsores.

In elderly care where falls are major health hazard, activity classification can be very useful for fall detection and prevention. An issue such as patient falling from their hospital beds could be prevented. The early detection of fall is a crucial step to alert and protect the person, so that serious injury can be avoided [27]. AR also allows patients who are at higher risk of falls to be identified offering an opportunity to intervene early to help prevent fall events from occurring, thus improving patients quality of life, increasing survival, and cutting the staggering costs related with falls and fall-related complications [28]. A number of researches including [27, 28] investigate specifically on fall detection by using small, non-invasive sensors which allows practical, inexpensive way for monitoring ambulatory movement of elderly people.

AR is used in robot applications. In elderly care domain, robots can work as companions of elderly living alone in their homes. Also, robots can provide services e.g. make a phone calls, tele-monitoring, etc. to older persons in home. An example is a robot-assisted living system introduced in [31]. AR is applied in robot system to determine users activity and intention and goal is then inferred in order that the robot can respond in an appropriate way. Also, interaction and communication between elderly people and robot can be achieved by AR. AR has a major advantage in facilitating intelligent elderly care. The activity monitoring system can be used to intelligently monitor the elderly living independently, providing a peace of mind for their relatives and friends. It can facilitate a new model of care in which ageing in home is encouraged. It can also benefit nursing home or any care institutes by providing real-time monitoring so that care can be done in a more effective and efficient way.

2.2.1.2 Physical health and fitness applications

In physical health and fitness applications, knowing a person is currently working out, information of energy expenditure, activity intensity level, etc. could be calculated and used to provide further health and fitness guidance which suits the user. Wearable sensors based AR is used to detect several sport activities such as rowing, cycling, etc. [124]. In this study, the authors suggest that a more detailed analysis of physical effort can be obtained by detecting the exact form of activity the subject is performing. A feedback can be provided to the user about his/her lifestyle regarding physical activity and sports therefore promote a more active lifestyle. An automated method of updating the exercise diary was proposed in [77]. Their method can detect various sports including racket sports i.e. tennis, team sports i.e. football, Nordic walking, gym training and aerobics etc. The diary recoding sport exercises and personal training activities can help motivate people to exercise more regularly and actively.

2.2.1.3 Assembly and maintenance application

Another application of AR is in the domain of industrial production. The captured information of workers activities can be used to provide guidance and support on the tasks. The information can help to perform intricate, tedious or critical tasks and improves productivity, decreases error rates, reduces labour cost [30]. Thus, works can be done in an effective and efficient way. A wearable computing prototype which enables a context-sensitive provision of necessary information is developed to the training workers supporting training and qualification activities at the SKODA production facilities in Czech Republic [29]. The experimental results of their system show that the performances of workers have improved. The assembly tasks are completed faster and with less error. Moreover, the system provides autonomous relevant information to the preformed activity resulting in elimination of dispensable movements when workers need to check assembly information. Similarly, AR is used on assembly of the front lamp of a car i.e. mounting and adjusting the lamp as depicted in Figure 2.4 [30].

Figure 2.4: Application of AR in industrial domain. AR on assembly tasks [30].

AR has also been applied in maintenance tasks. In critical maintenance such as in an aerospace industry, missing a verification step is always possible and could be avoided by tracking workers activities [30]. Example of work can be found in [30] research.

2.2.1.4 Dietary-related applications

Another interesting application of AR is in the dietary-related area. Balanced nutrition intake is important part of a healthy life. Automatic detection of food-related behaviour can be used in the development of an intelligent system to promote better health and well-being. AR with ubiquitous technologies can provide a mean for individuals to proactively monitor their food as well as water intake and act upon it, leading to a better food selection and sensible eating [24]. A dietary-aware dining table proposed in [24] is an intelligent table that can automatically track what and how much the individual eats from the dining table over a course of a meal. The system can detect activities such as transferring food among containers and eating food into individual mouths as well as the amount and type of food consumed. The system provides effortlessly way for individuals to quantify and acknowledge their dietary intakes. Moreover, understanding current users dietary behaviour can help improve customer service satisfactory in restaurants. An innovative research called Future Dining Table which can recognise users dining activities in order to provide recommendation on food which is related to the users current dining status [23]. The system stores information such as dining action history and current dining status and uses it to recommend the dishes that would fit the current meal in the right timing for additional order (See Figure 2.5). The system can be used in a restaurant to help waiters when human resources are limited.

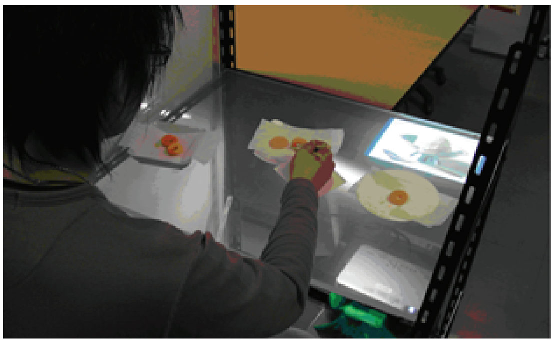

Figure 2.5: An innovative research, Future Dining Table, by [23].

2.2.1.5 Robotic applications

AR is an essential part for a personal service robot [21]. These robots must have the ability in detecting and recognising human activities in order to decide next appropriate actions. It is important for a service robot which offers help to a person to be able to detect and understand the user's intentions and infer his goals. This is particularly important when the robots confront with humans who are not acquainted with service robots and their behaviour. For example, AR allows the robot to be able to detect if the person needs help or guidance and approach him in a proper way and also able to recognise normal activities thus not interrupting a user in current activity. In addition, understand human activity would allow robot to successfully communicate with human.

2.3 Applications in assisted livings

2.3.1 Activities of daily living

A variety of activities have been investigated in wearable sensor-based AR research. This research is focused on the application in assisted living domain. The activities recognised can be divided into two categories namely ambulatory activities and ADL. Ambulatory activities are activities that related to walking including static postures e.g. standing, sitting, transition activities e.g. sit-to-stand, stand-to-sit, and dynamic activities e.g. walking. Examples of works which used wearable sensors to recognise ambulatory are such as [101, 106, 108, 129]. ADL, on the other hand, cover a broader range of activities often found in daily living. ADL is more complex than ambulatory activities and normally contain several movements. ADL activities can be divided into two types which are basic ADL and instrumental ADL (I-ADL). The basic ADL are activities necessary for self-care while IADL are other activities which involved the use of an instrument. Basic ADL include feeding, bathing, dressing, grooming, stairs, toilet use, bowel, bathing, bladder, transfer, mobility and stairs. Examples of I-ADL are such as using telephone, house work, doing laundry, watching TV, typing, vacuuming, cooking, etc. Majority of wearable sensor-based AR aim to recognise ADL as it has wider applications in various domains.

2.3.2 Activity recognition for assisted livings

General speaking, any AR system could potentially be applied in assisted living applications. This depends on the types of activities that the system can recognised whether they are suitable for assisted living applications or not. However, in this section only AR studies that are focused on applications in assisted living are reviewed. Tables 2.8, 2.9, 2.10, and 2.11 present descriptions such as sensor, sensor location, features, activities and algorithms about these studies. Based on the sensor type, these works can be classified into two main groups i.e. (1) on-object sensor, and (2) wearable sensor. The studies which use (1) on-object sensors in their AR system are [79, 80, 81, 82, 83, 166, 173, 174]. This approach predicts the activity of a person by using information from the status of objects and surrounding environment. The sensors are placed on objects or environment around home to monitor their status. The main sensors used are binary sensors which are used for detecting objects status. For examples, RFID is used to detect whether the object is touched by the user, contact switch to detect objects status i.e. open/close, motion sensor to detect the presence of a user, pressure mat, and door entry sensor. More specialised sensors such as sensors to monitor the use of water and stove burner, and phone usage have also been used. Sensors which are used to monitor changes in environment including temperature, light, and vibration are often used with binary sensor in the on-object sensor based approach. Using this approach, a detailed activity can be detected e.g. using PC, using microwave, hand washing, making oatmeal, put on etc. Also, the on-object sensor based approach shows high classification accuracy. However, in order to detect very fine detailed activities using the approach, a vast number of sensors need to be installed in home. Some objects also need to be replaced over times e.g. toothpaste, skin lotion, etc., therefore the sensors attached to them must also be replaced. Some systems i.e. [83, 173], require specialised sensors e.g. floor sensor, to be instrumented throughout the house. This could make the system more expensive as these sensors cannot be simply deployed and require retrofitting. Normally in a system that objects are attached with RFID tags, a user needs to wear or hold an RFID reader in order to detect the object status. For example, a user is required to wear an RFID glove all the time [166]. This would make the system impractical to use and not easy to accept by the elderly people.

The (2) wearable sensor based approach use information from the sensors which are worn over human body. The studies which use wearable sensors include [85, 101, 108, 117, 135, 145]. This approach uses sensors which can detect changes due to movement such as gyroscope, accelerometer, magnetometer, and ECG, and sensors that can detect surrounding environment such as light, temperature, and microphone. The wearable sensor based approach is focused on detecting activities that involve movement such as posture e.g. standing, sitting, transition movement e.g. sit-to-stand, stand-to-sit, repetitive movement e.g. running, walking, and ambulatory movement such as fall. The wearable sensor based approach can be further divided into sub-categories based on the number of sensors locations i.e. multiple and single location. The multiple location approach involves using sensors on top and bottom parts of the body. For example, sensors are used on shoulder, waist, and leg [85]. Radio tags are used on 12 locations over the body e.g. shoulders, elbows, wrists, etc. [117]. However, this approach may not be suitable for elderly people in term of usability and acceptance. A single location approach, on the contrary, uses sensors on a single location such as waist[108] and chest[101, 135]. This approach helps reduce the possibility of sensor interrupting with daily activities. Nevertheless, not all locations are suitable for AR and some locations may have higher usability and acceptance than others. The disadvantages of the single location approach are that the accuracy for the multiple location approach is normally higher and the activities studies are normally posture and transition.

Table 2.8: Studies in AR for applications in assisted livings

Author	Sensor	# Sensor	Sensor location	Features	Classification method	Recognised activities	Accuracy
Najafi et al. [101]	Piezoelectric gyroscope and two accelerometers	3	Chest	Discrete Wavelet Transform features	Hierarchical	Lying down, walking, as well as SiSt and stand-to-sit (StSi) transitions using different types of chairs (standard wooden chair, armchair, and upholstered chair), with and without armrests (N=6)	-
Karantonis et al. [108]	Accelerometer	1	Waist	Low pass, filtering, signal magnitude area	Hierarchical	Sit-to-stand, stand-to-sit, lying, lying-to-sit, sit-to-lying, walking (slow, normal, fast) fall (active, inactive, chair), circuit (N=12)	90.8%
Pawar et al. [135]	Electrocardiogram recorder	1	-	Mean of ECG beats	Proposed Body Movement Analyse classifier	Sitting still, arm movement, walking and climbing down stairs, climbing upstairs, twisting movement at waist. The arm movement is a combined class of three separate movements of left arm, right arm, and both arms	92.44%
Yin et al. [85]	Light, temperature, microphone, two-axis accelerometer, two-axis magnetometer	5	Shoulder, waist, leg	-	SVM and Kernel Non-Linear Regression	Sitting down, walking, walking down stairs, walking upstairs, running, slipping on the ground falling down backwards, falling down forwards (N=7)	98.5%
Luštrek et al. [117]	Radio tags	12	Shoulders, elbows, wrists, hips, knees and ankles	Coordinate and angle from the reference points	SVM	falling, lying down, sitting down, standing/walking, sitting and lying	Over 95%

Table 2.9: Studies in AR for applications in assisted livings (cont.)

Author	Sensor	# Sensor	Sensor location	Features	Classification method	Recognised activities	Accuracy
Diermaier et al. [83]	Accelerometer, reed contact switch, light sensor and temperature sensor	13	Floor, door, environment	Discretised data	Manually analysis of data	Laying down, getting up, being absent from the flat, being present in certain rooms of the flat and eating, etc. (N=151)	-
Cook et al. [174]	Motion and temperature sensors, analogue sensors that monitor water and stove burner use, VOIP captures phone usage, contact switch sensors to monitor usage of the phone book, a cooking pot, and the medicine container	-	Objects in homes	room location of the individual, on/off status of the water and burner, the open/shut status of the cabinet, and the absent/present status of the item sensors, as well as the number of seconds that elapsed since the previous sensor event	Markov Model	Telephone Use, hand washing, meal preparation, eating and medication, cleaning	-
Lu et al. [173]	Current sensor, flood sensor, contact sensor (reed switch, mercury switch), pressure mat, 3D accelerometer, motion sensor, vibration sensor, RFID	-	Objects in home	Mean, variance, area under curve, maximum, and minimum, frequency domain features	Bayesian Network	Using PC, using phone, studying, listening to music, watching TV, using microwave, using refrigerator, making tea, using printer, using other appliance with RFIDs, walking, sitting	92.43% (with location) 88.43% (Without location)
Hong et al. [82]	Movement detectors, contact switches and pressure mats	-	Objects in home	Binary sensor data	Ontology + Dempster Shafer	-	-

Table 2.10: Studies in AR for applications in assisted livings (cont.)

Author	Sensor	# Sensor	Sensor location	Features	Classification method	Recognised activities	Accuracy
Hong et al. [166]	Accelerometers and RFID	-	Thigh, wrist, waist and objects in home	Mean, energy, entropy and correlation	DT (body motion), DT (movement of hand), object used	Sitting, brush hair standing, standing, phone calling sitting, walking, taking picture standing, lying, reading sitting, running, wiping with cloth standing, hand shaking, running a vacuum cleaner, rope jumping, put on an umbrella standing, put on skin conditioner, toothbrush standing, pushing a shopping cart, cutting standing (N=18)	95%
Khan et al. [145]	Accelerometers	5	Chest pocket, front left trousers pocket, front right trousers pocket, rear trousers pocket, and inner jacket pocket	Spectral entropy, Autoregressive, Signal magnitude area	ANN	Resting (lying/sitting/standing), walking (along the corridor), walking upstairs, walking downstairs, running, cycling, and vacuuming	94%
Cheng et al. [80]	RFID	-	Objects in home	Sensor ID	Adaptive Learning Hidden Markov Model	Initial state 6 Making tea, Using the bathroom, Making or answering a phone call, Making oatmeal, Taking out the trash, Making soft-boiled eggs, Setting the table, Preparing orange juice, Eating breakfast, Making coffee, Clearing the table	89% (precision)

Table 2.11: Studies in AR for applications in assisted livings (cont.)

Author	Sensor	# Sensor	Sensor location	Features	Classification method	Recognised activities	Accuracy
Fleury et al. [79]	Infrared presence sensors, door contacts, temperature and hygrometry sensor in the bathroom, microphones and a wearable kinetic sensor (accelerometers and magnetometers)	-	Objects in home and environment	% of time spent in postures and walking, number of events per class, number of event per microphone, % of time in each room, % of time open and predominant position in the time frame, differential measure for the last 15 minutes for temperature and hygrometry	SVM	Hygiene, toilet use, eating, resting, sleeping, communication, and dressing or undressing	86.2%

2.3.3 Requirements of assisted living systems

The goal of assisted living solution is to enable elderly people to live longer in their preferred environment, to enhance the quality of lives and to reduce costs for society and public health systems [112]. Especially with the population ageing phenomenon, assistive technology will be the key component of care of elderly persons who require help with their daily activities within their own homes. A report by [100] shows that majority of elderly people prefer to remain in their own homes for as long as possible. Also, the cost of care home can be expensive comparing to assisted living facilities. At the present, there are a number of off-the-shelve products available in the market e.g. fall monitoring system on mobile phone, emergency alarm, etc. Usually they are closed, stand-alone systems with limited ability to describe actual situation, often too difficult for elderly people to use and useless in emergency situations [112]. There is a need for the assisted living solutions to become intelligent in order to actively assist elderly people.

There are three major requirements for assisted living systems which need to be met in order to fulfil its purpose and potential to assist vulnerable people [112]. Firstly, the system needs to be ambient and unobtrusive for high acceptance purpose. Secondly, the system must adapt themselves to changing situations or capabilities of the individual and environment to fulfil individual needs. Finally, the system must provide services in an accessible way. To sum up, the assisted living system must have these characteristics: adaptive, accessible, high usability, and high acceptance.

However, some of the currently available solutions only focus on the technical solution neglecting user acceptance and usability issues, especially for elderly people who are the most demanding stakeholders for IT development [112]. For example, a system which requires users to wear special equipment may be perceived as stigmatisation or too complicated to use resulting in low acceptance. For example in a mobility aid system [99], user interface is critical requirement as it has direct physical interaction with the users. An interview-based investigation by [102] also shows that elderly people are concerned about privacy violation, visibility and accuracy of the assisted living systems. Even if the systems could deliver the best services for assisting people but unless they are easily accessible and usable and address the real need and concerns of the users, they will not be accepted.

Another issue in current assisted living systems is lack of human and social interaction [96]. Over-using technology could reduce interaction between elderly people and outside community. Many of the Ambient Assisted Living reports emphasised on the importance of bridging distance and preventing loneliness and isolation of elderly people and combining services with formal and informal care [94]. Some assisted living systems [96] have taken this issue into consideration by combining support from patients family, friends and all care team e.g. doctors. By utilising human participation, effective assisting services could be achieved. Simulation results revealed that informal care helps reduce the social resources and provide timely assistance [98]. Elderly people social connection strengthened while the dependence on social resources is reduced when they are actively involving in group activities [97].

The cost of an assisted living system is another important issue [95]. The cost of current assistive technology equipment varies from £6 (Talking medicine label) to £3299 (Special magnifier). For a practical solution in assisted living, the systems need to be cost-effective to make it affordable for general population. With current sensor technology, small and low-cost sensors can be embedded in everyday objects such as cloth, watch, etc. to provide cost-effective assisted living solution. The assisted living domains are still in an immature state, nevertheless with population ageing it will soon be a huge market and in order to compete in such market, the cost will be a vital factor. There remain many issues and challenges in activity recognition for applications in assisted living other than technical perspective. These include user acceptance, usability, privacy, visibility, systems accuracy, lack of human and social interaction and cost.

2.4 Identified research gap

This section discusses and identifies the gaps attained from the analysis of literature reviews in sensor-based AR for assisted living. It also discusses how the research is different from previous studies.

Earlier approaches in AR have been through visual sensors. However, this may not be practical for home-based care due to privacy concerns. Later, the research in AR moves toward using non-intrusive sensor for AR. However, the works in early years are mainly focused on the technical aspect of the system, that is, to recognise the activity. Usually they are closed, stand-alone systems with limited ability to describe actual situation, often too difficult for elderly people to use and useless in emergency situations [112]. Factors such as location of the sensor, number of sensors are linked with the acceptance and usability level of an assisted living system. Certain sensors location or multiple sensor locations may prevent elderly people to perform activities normally or may cause discomfort. Also, some sensor types may be perceived as stigmatisation or too complicated to use resulting in low acceptance. In a system which distributes sensors in environment normally requires a larger number of sensors. This approach may be time consuming and not feasible to set up. For example, RFID tags are attached on numerous objects in homes [76, 166]. Similarly, home objects e.g. cups, fridge, tea, etc. must be equipped with contact switch sensors in [82]. Cost of the system is also another important factor. The AR system needs to be affordable in order it to be useful to improve health care of general population.

Therefore, in this research, a wrist-worn sensor is proposed and developed to detect activity of an elderly person. Wrist is an ideal location which should not interrupt normal activities. Earlier studies such as [116, 130, 157, 158] investigated the use of wrist-worn sensors to detect activities. Some limitations are persisted in these studies. Firstly, the activities are limited to mainly posture e.g. lie down, sit, stand and transition activities e.g. sit-to-stand, stand-to-sit. Secondly, the classification accuracy is lower than systems which use sensors at multiple locations. Thirdly, these studies only use one sensor i.e. an accelerometer. To overcome these limitations, multiple sensors worn on wrist will be used. A rich data can be acquired by using more sensors which will help improve the classification accuracy.

A study which is closely related to this research is the work by [75] where light and two-axis accelerometer embedded on wrist watch is used. Activities include walking, sitting, standing, taking stairs up and stairs down, taking elevator up and down, and brushing teeth are studied. The sensor fusion is carried out at feature level where several features are calculated then a Correlation based Feature Selection algorithm is used to select a subset of features. The study investigates the classification performances of the sensors used in different body locations such as wrist, the belt, shirt pocket, trouser pocket, backpack, and necklace. The study shows that sensor worn on wrist location achieved the highest accuracy of 87%. The results of the study indicate the feasibility of using multi-sensor on wrist for AR. Also, the results show that combining light and accelerometer increased classification accuracy. However, this study only concentrates on basic activities such as sitting, standing, walking, ascending stairs, descending stairs and running.

Another related study is the one by [107] where multiple sensors worn over the waist are used. Their sensor fusion is done at feature level. Around 600 features are generated and feature selection based on Boosting technique is performed. In Boosting technique, each feature is associated with one classifier. The classifier is then trained for all features and the feature which is associated with the best model is selected. The weak feature is associated with a lower weight. The study shows that accelerometer and microphone yield the most discriminative information for the activities studied. The results also indicate automatic feature selection helps in recognising activities, and information of users context can improve the inference. The study [107] differs from this research in term of sensor location. In addition, their work is concentrated on the use of Conditional Random Field for AR. It is unclear how their feature select method can be applied with other classification techniques.

Another important aspect of the multi-sensor based AR is the sensor fusion. Based on the review, it is found that a majority of wearable-sensor based AR performed sensor fusion at feature level. Using this strategy, the features from different sensors are concatenated together and fed to the classifier. Techniques such as manual analysis, feature selection, and feature reduction are used to select and reduce the feature space. However, manual analysis is not suitable for a large feature space. Also, some feature selection techniques such as Boosting, PCA, and Clamping only concern the relationship between the feature and the classes. The relationship between features are neglected which may result in the selection of redundant features. Although, in other popular feature selection techniques such as MRMR and NMIFS, relationship between features is considered, it is only one-to-one relationship i.e. feature to feature. There is still a gap regarding the feature selection technique which also taken the relationship between a set of features and the classes into account.

Sensor fusion can also be done at classifier level. A limited number of wearable-sensor based AR studies performed sensor fusion at this level. Features from each sensor are used in each classifier and the final prediction is the combination of the predictions generated from each classifier. However, this method requires that each sensor is capable of activity recognition with good accuracy. This method is not suitable the choice of sensors in this research, the . In contrast, the research will use different classifiers to generate different predictions and combine them for the final prediction. Two main approaches i.e. static and dynamic can be used for classifier combination. The static approach is simple and uses low computational cost. However, the optimum result cannot be guaranteed and over-confidence classifier could affect the combination result. Also, the combination model generated from this approach may not be suitable for different data sets.

A previous study [103] used GA to find combination between features, classifiers, and classifier combiners. The study shows an interesting idea of using GA to find the combination model. There are some gaps which can be extended from this study. First, other combination criteria can be added when selecting the combination. For example, less number of classifiers or different classifiers is more preferable, etc. Secondly, feature and classifier selection is performed in this study which may not be suitable for classification algorithms which require different values of parameters depending on the features e.g. SVM. Although this can be solved by integrating the process of determining parameters into the algorithm, it will require more computational cost. Thirdly, adaptive algorithm can be added so that the combination model is periodically adapted to the new data.

Classifier can be associated with weight to improve the combination result. Section 2.1.10 reviews some weight strategies. The review indicates that static and deterministic weight methods cannot guarantee the improvement in classification accuracy. Other technique e.g. GA is used to find weights [92, 93, 103]. However, there are still some limitations regarding these studies. Firstly, the performance of GA is based on the combination of all classifiers. Based on this, the conclusion that GA could improve classifier combination accuracy is not always true as all combinations have not been tested. Secondly, the accuracy or error of the weighted average fusion technique i.e. $f(w) = w_1x_1 + w_2x_2 + ... + w_Kx_K$ is the only fitness function studied. Other fusion function which reflects on different combination function e.g. sum, minimum, maximum, product, etc. have been studied before. Finally, combination results are often compared with the mean accuracy of a set of classifiers rather than to the best individual classifier. The mean accuracy is always equal or less than the accuracy of the best individual classifier. Equal accuracy is only occurred if and only if all classifiers have the same accuracy. This weakens the conclusion that the classifier combination is better than a single classifier.

To summarise, the following identifies the main gaps in sensor-based AR for assisted living:

- Practical aspect

- acceptance, usability, cost and privacy -

- Technical aspect
  - Sensor-based AR at single location
    * Limited activities
    * Low recognition accuracy
    * Multi-sensor fusion for AR
  - Sensor fusion at feature level
    * Analysis of feature manually is not suitable for a large number of features.
    * Some feature selection techniques only concern the relationship between the feature and the classes. They do not concern the relationship between features or group of features and classes.
  - Sensor fusion at classifier level
    * Using static techniques cannot guarantee the improvement in classification performances of the combination model
    * The combination model generated cannot be applied to different data set.

Due to time constraint in this research, only technical aspects will be focused.

2.5 Summary

This chapter presents a state of art in wearable sensor-based AR research. Previous researches mainly focus on the investigation of the possibility of recognising human activities using wearable sensors. The literatures have shown that using wearable sensors for AR is possible. A number of sensors and classification techniques have been investigated. However, there is still a gap due to the lack of practicality concern in the development of AR model. This is particular important especially in the assisted living application. Based on the review, there are two aspects that need to be met in order to fulfil its purpose and potential to assist vulnerable people. These are practical e.g. acceptance, cost, privacy and technical i.e. accuracy aspects. Even if systems could deliver the best services for assisting people unless they are easily accessible and usable and address the real needs and concerns of the users, they will not be accepted. The practical issues can be overcome by the use of appropriate sensors and location. In this research, small, low-cost, non-intrusive non-stigmatize wrist-worn sensors are investigated. Previous studies which use wrist-worn sensor only covered limited activities e.g. mainly ambulatory and transition activities and only single sensor i.e. accelerometer is often used. From the literatures, it can be seen that multiple sensors can yield more information. This research is interested in using a multiple wrist-worn sensor for AR of an elderly person with the aim of achieving practicality in term of user acceptance, privacy (non-visual) and cost and high accuracy.

Chapter 3

System architecture and approach

Chapter 2 provides a review of prior studies in wearable sensor-based activity recognition and identifies the shortcomings on previous activity recognition for assisted living applications. To overcome the practicality issues in term of cost, privacy (non-visual), and acceptance, and to extend the types of recognised activities and improve classification accuracy, an investigation on multi-sensor activity recognition is carried out. The aim is to develop an activity recognition model which is practical with high accuracy. This chapter presents overviews of research design, research approach and system design. The design and development of wearable sensors and justification of sensor location and choices of activities are presented. This is followed by the details of sensor data set acquisition. The chapter also describes the multi-sensor activity recognition framework proposed in this research. followed by the descriptions of how the proposed work is assessed and compared with other studies.

3.1 Research design

There are two main aspects of the research gaps in sensor-based AR for assisted living as discussed in Section 2.4. The first gap is related to a practical aspect including cost, usability, acceptance and privacy. The other gap is related to a technical aspect including classification accuracy and sensor fusion. To overcome these limitations, the research on how to recognise an older persons activities using non-visual, non-intrusive, small, and low cost wrist-worn sensors is carried out. The results of the study can help identify the activity recognition method suitable for a practical assisted living application. This section discusses the research design. Firstly, to understand the process of sensor-based AR and check the feasibility of the proposed concept, a feasibility study needs to be carried out. It was decided to follow the methodology proposed by **??** as the paper is very related to the proposed method in the research, and the methodology is explained in great detail. The results of the feasibility will help preliminary identify features and techniques, and limitations of activity recognition. Next, the sensor data generated from older adults performing activities are required to investigate and understand their characteristics. Based on the literature review, sensor data can be obtained from either data collection, or public data set. In this research, it was decided to collect the data as it was not possible to find a suitable data set which contains a variety of sensor types and activities. From the literature review, there was only one study which used wrist-worn sensors, and their number of sensor types and studied activities were very limited. The sensor types must be selected to use in the system. The sensor selection can be identified based on the literature review e.g. sensors which are successfully used by other studies, and sensors which can potentially provide useful information for activity recognition. Also, due to the limitation of electronics skills, the sensors and their platforms should be easy to implement, and/or ready-off-the-shelve. After the data are collected, a series of experiments can be carried out to answer research questions. Tools including Matlab and SPSS are used for data analysis, and model evaluations. The reason of software choice are Matlab is a popular platform which can be used for exploring, visualising, and modelling data, and SPSS is a popular platform for statistical analysis. The choices of techniques used and investigated in the research are selected based on literature reviews e.g. techniques that are successfully applied and popularly used in related problems. As part of this research is concerned on the practicality of the AR, the evaluation and comparison will be carried out such that issues such as cost, usability, privacy, and acceptance are considered.

3.2 Research approach

The methodology used in carrying out the research can be separated into three main tasks. (1) First, three activity data sets were collected to develop and test the activity recognition algorithm. The first data set were collected from a group of young participants which is used for a feasibility study on using a wrist-worn sensor to detect human activity. The activities collected from this group are basic activities such as walking, standing, sitting, etc. in both indoor and outdoor environments. The second data set were collected from a group of elderly people at a residential home to develop, train, and test the activity recognition algorithm. The participants wore three wrist-worn sensors and performed 10 daily activities such as brushing teeth, feeding, dressing, etc. under natural settings. The final data set are also collected from a group of elderly people. However, seven sensors were used in this data collection. The participants performed 13 activities such as wiping, reading, exercising, etc. (2) Once the data were collected, a series of experiments were carried out to develop an activity recognition algorithm that can detect a range of activities of daily living, practical (use non-visual, low-cost, low-profile, wrist-worn sensors) and high accuracy. These experiments were performed to determine parameters for the algorithm e.g. the sliding window length, features to extract, sensors, classification algorithms parameters as well as to develop and evaluate the proposed feature selection and classifier combination techniques. (3) Finally, the completed activity recognition model was evaluated on several criteria in term of performance and practicality. The design and justification on the choice of sensors, sensor location, and activities recognised are described in detail in Sections 3.3. Then, data analysis was carried out with the aim of identifying suitable features and activity recognition model development.

3.3 System design

In this research, a practical activity recognition method which uses wrist-worn sensors is proposed to recognise activities of daily living of an elderly person. The activities cover basic daily routines, and household activities. Six wrist-worn sensors are investigated namely accelerometer, temperature sensor, altimeter, gyroscope, light sensor, barometer. Also, a heart rate monitor is investigated and evaluated if its usage justifies the reduction in term of the wearability of the system. The following sections describe the wearable sensors used, software and hardware implementation, the sensor locations, and data collection procedures used to collect necessary data for the development of a multi-sensor based activity recognition algorithm.

3.3.1 Wearable sensors

This research involves the use of multiple wrist-worn sensors to detect daily activities. It can be seen from previous chapter that there are various sensors available for activity recognition. Therefore, the following criteria has been set in order to select suitable sensors that meet the research's aim.

1. The sensor must be able to capture human movement or changes in environment necessary for activity recognition.

2. The sensor must be inexpensive, and easy to acquire (preferably off-the-shelve).

3. The sensor must be easy to implement and/or develop and/or integrate and/or extend on an existing wrist-worn sensor board or system.

4. The sensor must be low power consumption.

Based on the above criteria and literatures, seven sensors are selected for activity recognition in this research. The selected sensors and their justifications are presented in Table 3.1.

Table 3.1: Sensor choice justification

Sensor	Justification
Accelerometer	Accelerometer has a capability to respond to both frequency and intensity of movement, and measure tile as well as body movement [13, 83, 177]. Its capability in activity recognition has been proved in many systems [157, 158, 159, 185].
Temperature sensor	Temperature sensor measures temperature of the environment or human body (depending on the sensor location). It can be used to indicate changes in environment which may occur when performing certain activities. Studies have used temperature sensor in conjunction with other sensors [79, 155].
Altimeter	Altimeter measures the altitude of an object from a fixed level. The information can provide information for detecting certain activities which involve changes in heights e.g. using stairs, using lifts, etc.
Gyroscope	Gyroscope are often used with accelerometer to provide additional movement information in term of rotation angle and direction. It can measure the orientation and rotation of the movement. A study has shown that a combination of accelerometer and gyroscope can improve activity classification's accuracy [38].
Barometer	Barometer is used to measure information about pressure and temperature of the environment. Studies have indicated that using accelerometer with barometer can improve activity classification's accuracy [35, 36].
Light sensor	Light sensor measures the intensity of the light in the environment. It can provide additional information for activity recognition of certain activities which have changes in lighting condition. A study has used light sensor in conjunction with other sensors to detect activities [37].
Heart rate sensor	A study has shown that there is a relationship between heart rate and physical activity which can be used to measure physical activity indirectly [154]. Some studies has combined accelerometer and heart rate sensor data to recognise activities [34, 153].

These sensors are presented into three different platforms including (1) EZ-430 Chronos watch, (2) Gadgeteer sensors, and (3) BlueRobin heart rate monitor. The following sections describe these platforms and their implementations in further details.

3.3.1.1 EZ-430 Chronos watch

The EZ-430 Chronos watch is developed by Texas Instrument. It is a fully functional sport watch which has integrated accelerometer, temperature, pressure and altimeter sensor, and battery and voltage sensor on board. The watch is light, small, and easy to wear which will not disrupt the elderly persons movement or create stigmatisation. The EZ-430 Chronos is based on the CC430F6137 Micro-controller with the MSP430 CPU which is the industrys lowest power MCU. The watch also contains 8 KB of flash memory available for data logging of altitude, temperature, and heart rate. The on-board accelerometer can measure acceleration in three dimension at a range of up to $\pm$2G (G = 9.81 m/s^2) with 8-bit resolution and the sensitivity of 56 count/G. The accelerometer sampling rate is 100 Hz. However, to reduce the energy consumption, the watch only transmits the third data set. Thus, the accelerometer sampling rate is set to 33 Hz. For continuous acceleration measurement, the watch consumes about 166.0 μA. The altitude sensor has 30 kPa - 120 kPa measuring range with 19 bits resolution. For continuous altitude measurement, the watch consumes about 18.0 μA. The temperature sensor can measure the range of -20 to 70 degree Celsius with 14 bits resolution. For continuous temperature measurement, the watch consumes about 10.0 μA. In this research, the sample rate of temperature sensor and altimeter is set to 1 Hz. The EZ-430 Chronos watch uses a CR2032 battery. For continuous temperature measurement, the watch would last 25.0 months. For continuous altitude measurement, the watch would last 13.8 months. For continuous acceleration measurement, the watch would last 1.5 months. The EZ-430 Chronos watch has an integrated 868 MHz wireless transceiver which allows communication between the computer through a USB RF access point wirelessly. In this research, an application is implemented using Matlab to collect accelerometer data in real time. It recorded date, time, acceleration in X-axis, acceleration in Y-axis, and

Figure 3.1: The EZ-430 Chronos watch. The watch and its module is compared to a pound coin. This figure also illustrates when the watch is worn by a person

acceleration in Z-axis. The watch flash memory is used to log date, time, temperature, and altitude data. The data in the flash memory are later transferred to the computer via radio frequency using a DataLogger Software provided by Texas Instrument.

3.3.1.2 Gadgeteer sensors

The Gadgeteer platform is an open-source toolkit for building small electronic devices developed by Microsoft. It has a wide variety of hardware modules which can be programmed using the .NET Micro Framework and Visual Studio/Visual C# Express. In this research, the data is collected from three Gadgeteer sensors namely gyroscope, barometer, and light sensor. The gyroscope can measure up to $\pm$ 2000 °/s with 14.375 LSBs per °/s sensitivity and 16-bit ADC. The sampling rate of gyroscope is set to 33 Hz. The barometer is based on piezoresistive sensor. It can measure between 300 and 1100 hPa absolute Pressure Range with 14 Bit ADC resolution. The sampling rate of light sensor and barometer are set to 1

Hz.

The barometer, gyro and light sensors are implemented on Gadgeteer FEZ Cerberus board as shown in Figure 3.2. The FEZ Cerberus is 168 MHz 32-bit Cortex M4 processor with 1 MB FLASH and 192KB RAM board. The sensor data including date, time, rotation X-axis, rotation Y-axis, rotation Z-axis, light intensity, barometric pressure, and barometric temperature are sent via I2C bus and recorded on a SD card. The data are later transferred to the computer via SD card reader. The board is powered using an 800 mAh power bank for a light weight application. The board is placed on the power bank which is placed on top of the wrist watch. Figure 3.3 shows the developed wearable sensors used in this research.

3.3.1.3 BlueRobin heart rate monitor

The heart rate monitor is developed by BlueRobin. It has built-in 868 MHz radio frequency which can transmit a range up to 800 meters (depending on environment). It has a built-in data collision prevention allowing multi-user systems with up to 200 chest straps and provides a 24-bit ID to uniquely identify each chest strap. The heart rate monitor uses a CR2032 battery. The chest strap is made of elastic rubber and is waterproof. The heart rate monitor sampling rate is set to 1 Hz. The heart rate monitor can communicate with the EZ-430 Chronos watch. The heart rate data is transmitted to the watch and logged in the watchs internal flash memory which is later transferred to the computer via radio frequency.

3.3.2 Locations of sensors

As the aim of this research is to propose a practical activity recognition method for detecting an elderly person ADLs in term of user acceptance, privacy and low-cost, it is decided that the sensors should be worn at a users wrist. The justification of the system design on this work has been based on the literatures and innovative ideas. For example, the justification that using the accelerometer on the wrist is due to the practicality issue that, from literatures, wrist is the optimum location for wearable sensors as it does not interrupt daily activities. Also, the literatures indicate that it is possible to predict activities based on a

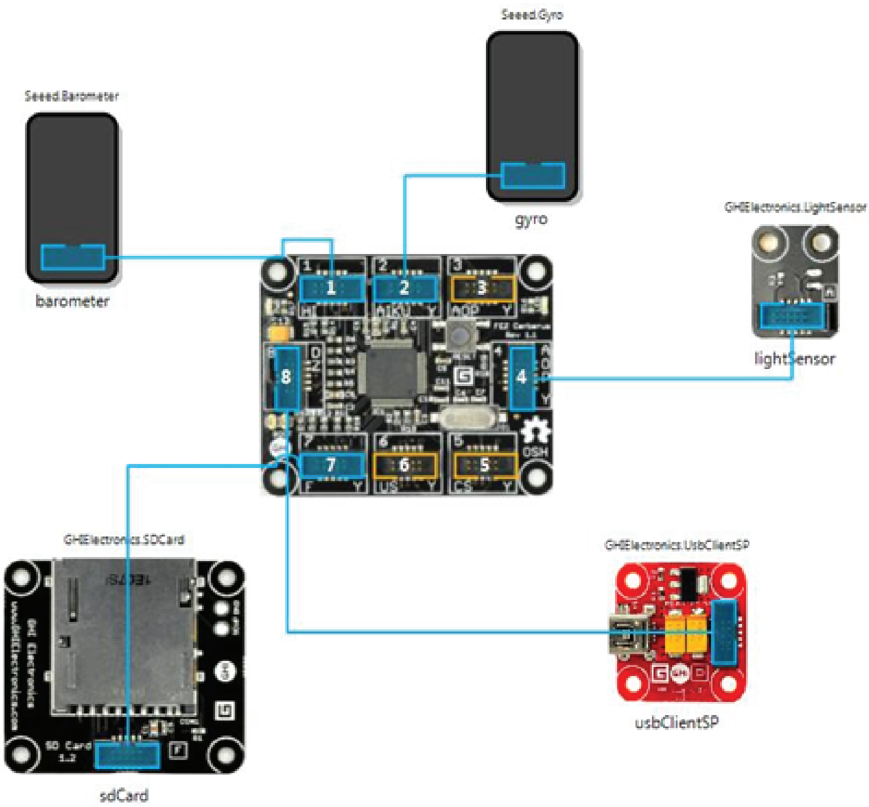

Figure 3.2: The design of the multi-sensor hardware based on Microsoft Gadgeteer. The FEZ Cerberus board is connected with barometer, gyroscope, and light sensor. The SD card is used to log the data. The board is powered by a USB power bank.

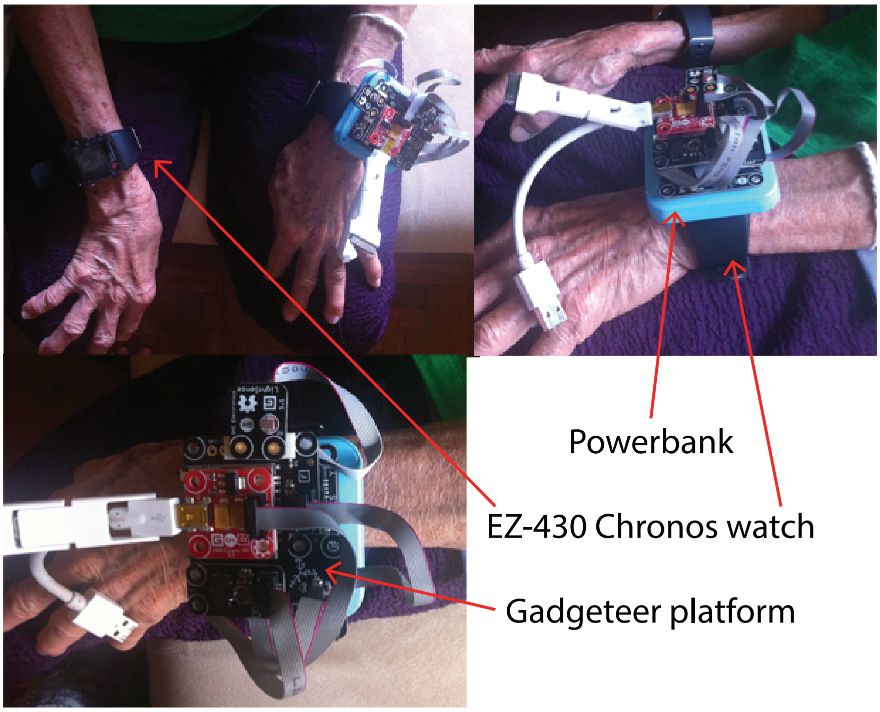

Figure 3.3: The gadgeteer sensor board is powered by a power bank through a USB. The gadgeteer platform is mounted over the EZ-430 Chronos watch.

Figure 3.4: Heart rate monitor. The heart rate monitor is worn over the chest using chest strap.

wrist-worn accelerometer.

However, due to hardware and time limitation, it is not possible to implements all the sensors on the single watch. Therefore, it is decided to separate the sensors between two wrists. We separate the sensors in a way that it should not interfere with the activity recognition. The sensors which are related to the movement i.e. accelerometer and gyroscope are worn on the dominant wrist in order to capture the users movement. Also, barometer and light sensors are also worn on the dominant wrist as they are parts of the Gadgeteer platform. The temperature sensor which captures the body temperature and altimeter are worn on the non-dominant wrist. In a real application, it is expected that all the sensors will be implemented on a single watch and worn on the dominant wrist of the elderly person. This location will not disrupt a user from performing an activity and/or cause discomfort in wearing sensors. The heart rate monitor needs to be worn over a users chest using a chest strap. Figure 3.5 shows the location of the sensors on a participant. Although the chest strap is made from elastic fabric, wearing the sensor for a continuous time might cause discomfort. The study will evaluate the trade-off between discomfort and the obtained accuracy.

3.4 Choice of activities

The choice of activities depends on the objective of a particular system. In the context of this research which is assisted living, a recognition of ADL is of interest. Recognised ADL can be used for evaluating elderly independence [91] to make sure that the elderly can carry out basic activities in their daily life. This research investigates both basic ADL and I-ADL activities in attempt to cover majority of activities occur in independent living situation. For the basic ADL, five activities from Barthel Index [90] are selected namely feeding, grooming (brushing teeth), dressing, mobility (walking) and stairs. Note that activities that are not selected are due to the difficulty in data collection in term of privacy. In addition, sleeping activity is also selected as it is common activity in everyday life. For I-ADL, housework activities i.e. washing Dishes, ironing, scrubbing, wiping and sweeping and leisure activities i.e. watching TV, reading, and exercising are studied.

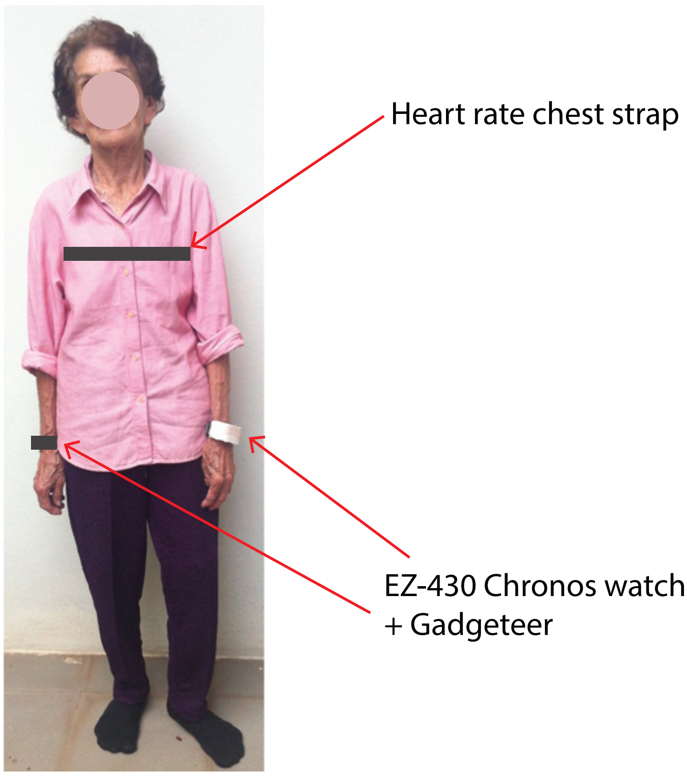

Figure 3.5: The location of the sensors. The gyroscope, barometer, and light sensor on Gadgeteer board are mounted over the Chronos watch. The participant wore two watches and a heart rate monitor on her chest. The participant's face has been blurred to preserve the anonymity.

3.5 Data set acquisition

Three different data sets are collected and used during the development of the activity recognition algorithm in this work. The first one, referred as the Young activity data set, is a small data set consisting of accelerometer data of five activities collected from seven young participants (aged less than 65 years). This data set is used to validate the feasibility of using wrist-worn sensor for human activity recognition and investigation on features for activity recognition. The second data set, referred as the Multi-sensor activity data set, consists of the data collected from three sensors including accelerometer, temperature sensor and altimeter. This data set is collected from 12 older adults performing 12 activities to develop the feature selection and activity recognition algorithm. The third data set, referred as the Wearable-sensor activity data set, consists of data collected from seven sensors including accelerometer, temperature sensor, altimeter, gyroscope, barometer, light intensity sensor, and heart rate monitor. This data set is collected from 12 elderly people performing 14 activities. The data set is used for developing the feature selection and activity recognition algorithm. Five benchmark data sets including Iris, Breast Cancer-1992, Breast Cancer-1995, Cardiography, and Chess are used for the evaluation of the feature selection algorithm. These data sets have been used extensively in feature selection and pattern recognition literatures. The data sets are obtained from UCI Machine Learning Repository available at http://archive.ics.uci.edu/ml. Next, the details of each of the data collection session are discussed.

3.5.1 Ethics and participant evaluation

This research study involve the studies with human participants, ethical issues regarding human participation are taken into consideration. Before the data collection sessions, the participants are given brief introduction about the study and an explanation about the data collection processes and written informed consents are obtained from all participants and they are informed they could withdraw at any time from the study. An example of the informed consent used is presented in Appendix A. This research project is approved by the Faculty of Computing, Engineering and Technology Academic Ethics Team, Staffordshire University, UK.

The participants are also asked about their personal health issues and evaluated using the Barthel Index [90] to assess if they are suitable for participation. The score sheet of Barthel Index is available in Appendix B.

3.5.2 Young activity data set

This is a data set consisting of acceleration data of five activities including sitting, standing, lie down, walking, and running. The descriptions of these activities are presented in Table 3.2.

Table 3.2: Descriptions of the activities collected in the Young activity data set

Activity	Description	Goal used
Sitting	Sitting on a chair	Walk to the notice board and read one of the posters.
Standing	Standing still	Walk to the garden and sit on a bench.
Lying	Lying down face up	Running to the Octagon building.
Walking	Walking at subjects normal speed	-
Running	Running at subjects normal speed	-

The participants are asked about their gender, age, weigh, and height prior the data collection. The data set is collected from seven young participants aged between 27 and 35 years. Two participants are females and five are males. The characteristics of the participants are presented in Table 3.3. This data set is used to investigate the feasibility of using a wrist-worn sensor for human activity recognition and identify a set of features used for activity recognition.

Table 3.3: Participants characteristics for the Young activity data set

Gender	Age (year)			Weight (Kg.)		Height (m.)		BMI(kg/m^2)	
	Mean	Std.	Range	Mean	Std.	Mean	Std.	Mean	Std.
female	27.00	1.41	2.00	49.00	8.49	1.66	0.849	17.69	1.27
male	29.00	1.87	4.00	61.20	7.92	1.67	0.043	21.86	2.91
all	28.43	1.90	5.00	57.71	9.45	1.67	0.050	20.67	3.17

The participants are asked to wear the EZ430-Chronos watch which has an on-board accelerometer on their non-dominant wrists. The accelerometer characteristic and detail are discussed in Section 3.3.1. The data collection is conducted outside in natural setting at Staffordshire University, UK. As some of the activities collected in this data set are postures, a goal based strategy is used in order to collect the data as realistic as possible. For example, a goal to read a poster from a notice board is used for collecting standing activity. The goal used in this data collection is shown in Table 3.2. The participants are allowed to complete these goals in their own times to allow the activities to be carried out naturally. The participants were firstly explained about the overall process of the data collection, and given the list of goals that they had to carry out. The participant had time prior the data collection to ask any questions regarding the goals. Before the start of each goal, the participant informed the research the goal they wished to carry out. The researcher marked down the name of the activity, date, and time. The sensor data are sent to the laptop wirelessly via 868 MHz radio frequency. The data set contains the participant ID, date, time, X-axis acceleration, Y-axis acceleration, Z-axis acceleration, and activity name. The total amount of data collected is 35 minutes containing 69,400 items of acceleration data. The distribution of the data are sitting 21%, standing 26%, walking 18%, lie down 27%, and running 9%.

3.5.3 Multi-sensor activity data set

This data set is collected from three sensors including accelerometer, temperature sensor, and altimeter. A total of 12 elderly participants aged ranging between 65 and 78 years old are recruited through advertisement by the representative

Figure 3.6: Example of a data collection session for the Young activity data set. The participant wore the Chronos watch and performed several activities indoor and outdoor.

of Watket Elderly club. Their characteristics including age, weight, height, and body mass index (BMI) are shown in Table 3.4.

Table 3.4: Participants characteristics for the Multi-sensor activity data set

Gender	Age (year)			Weight (Kg.)		Height (m.)		BMI(kg/m^2)	
	Mean	Std.	Range	Mean	Std.	Mean	Std.	Mean	Std.
female	72.11	4.54	13.00	48.26	10.13	1.53	0.060	20.53	4.10
male	71.00	3.61	7.00	51.80	12.51	1.64	0.070	19.18	4.14
all	71.83	4.20	13.00	49.14	10.28	1.56	0.079	20.19	3.96

The data collection is carried out in a real home in Chiang Mai, Thailand in order to replicate a natural living environment. This process is carried out over several different days. The participants are asked to wear two EZ430-Chronos watches on their wrists as shown in Figure 3.5. Each watch has three sensors on board including accelerometer, temperature sensor and altimeter. One of the participants is left-handed, while the others are right handed. The watch on the dominant wrist is set to transmit acceleration data while the other watch recorded temperature and altitude. The participants are asked to perform 12 activities. The list of the activities and their descriptions are shown in Table 3.5.

Table 3.5: Descriptions of the activities collected in the Multi-sensor activity data set

Type	Activities	Activity and independence description
BADLs	Feeding	Feeds self without assistance (using spoon and fork)
	Brushing teeth	Brushes self-teeth without assistance, including the use of toothpaste
	Dressing	Gets clothes and dresses without any assistance except for tying shoes
	Walking	Walks from one place to another without assistance
	Walking upstairs	Walks up the stairs without assistance
	Walking downstairs	Walks down the stairs without assistance
	Sleeping/lie down	Sleeps or lies down on a bed
IADLs	Washing dishes	Washes dishes, glasses
	Ironing	Irons shirt, trousers, pillow case, etc.
	Sweeping	Sweeps floor using broom
	Watching TV	Sits and watches television

The participant is asked to perform each activity for 5 min except for brushing teeth, dressing, walking downstairs and walking upstairs which had no time limit (See Figure 3.7). The participant is allowed to perform these activities in any order and they could take breaks between activities. Before the data collection, the watches had been calibrated and paired with the computer. The researcher marked down the start, stop time and name of each activity. In order to reduce the strain caused by the appearance of the researcher during the data collection process, the participants are left to perform activities at their own paces without direct supervision. The acceleration data is collected using software developed on MatLab. Temperature and altitude data are recorded on the watches internal memory which is later transferred to computer using the provided software from Texas Instruments. The data collected from accelerometer are date, time, acceleration on X, Y and Z axis. The data collected from temperature sensor and altimeter are date, time, temperature and altitude. In total, 19.2 hours of sensor data are collected. The classes' distribution are brushing teeth 8.59%, dressing/Undressing 4.51%, feeding 12.11%, ironing 11.90%, sleeping 14.95%, sweeping 11.12%, walking 10.99%, walking downstairs 1.02%, walking upstairs

0.97%, washing dishes and/or glasses 11.85%, and watching TV 11.99%.

Figure 3.7: Example of a data collection session for the Multi-sensor activity data set. The participants wore two Chronos watches and performed activities in homes.

3.5.4 Wearable-sensor activity data set

This data set is collected from 12 elderly participants aged between 66 and 79 years. The data set contain data from seven sensors which are accelerometer, temperature sensor, altitude, gyroscope, barometer, light sensor, and heart rate monitor. The advertisement is used to recruit the participants with the collaboration from the representative of Watket Elderly club, Chiang Mai, Thailand. Table 3.6 shows the characteristics of the participants including their age, weight, height, and BMI. The data collection session is carried out in a real home in Chiang Mai, Thailand over several days.

Table 3.6: Participants characteristics for the Wearable-sensor activity data set

Gender	Age (year)			Weight (Kg.)		Height (m.)		BMI(kg/m^2)	
	Mean	Std.	Range	Mean	Std.	Mean	Std.	Mean	Std.
female	72.70	4.76	13.00	50.80	10.75	1.58	0.039	20.44	4.48
male	74.50	2.12	3.00	47.00	14.14	1.58	0.035	18.83	4.85
all	73.00	4.41	13.00	50.17	10.72	1.58	0.037	20.17	4.36

The participants wore heart rate monitor strap over their chests for monitoring their heart rate and other sensors are worn on their wrists as shown in Figure 3.5. The temperature and altimeter are collected on the non-dominant wrist, while accelerometer, gyroscope, light, and barometer are collected on the dominant wrist. The participants are asked to perform 13 activities of daily living including brushing teeth, exercising, feeding, ironing, reading, scrubbing, sleeping, using stairs, sweeping, walking, washing dishes, watching TV and wiping (See Figure 3.8). For exercise activity, the participants are asked to perform exercise using elastic stretching band. Nine of the activities are similar to the activities collected in the Multi-sensor Activity Data set (See Table 3.5 for descriptions). The descriptions of the other four activities are shown in Table 3.7.

Table 3.7: Descriptions of the additional activities collected in the Wearable-sensor activity data set

Activities	Activity and independence description
Scrub	Scrubbing floor using cloth or scrubbing brush
Wipe	Wipe table using cloth or sponge
Read	Read a magazine/book/newspaper
Exercise	Exercise using an exercise elastic band for stretching

For each activity, the participants are asked to carry out the activity for 10 minutes. They could perform the activity in any order and are allowed to have breaks between activities. In total, 33.75 hours of activity data is recorded. The 12 raw data including 3 axes of acceleration, heart rate, temperature, altitude, light, barometer temperature, barometer pressure, 3 axis of rotation are recorded. In total there are 64,084 patterns and the classes' distributions are 7.55% brushing teeth, 8.11% exercising, 7.39% feeding, 7.13% ironing, 7.56% reading, 8.11%

scrubbing, 8.64% sleeping, 6.71% using stairs, 7.82%sweeping, 7.02% walking, 7.53% washing dishes, 8.76% watching TV, and 7.66% wiping.

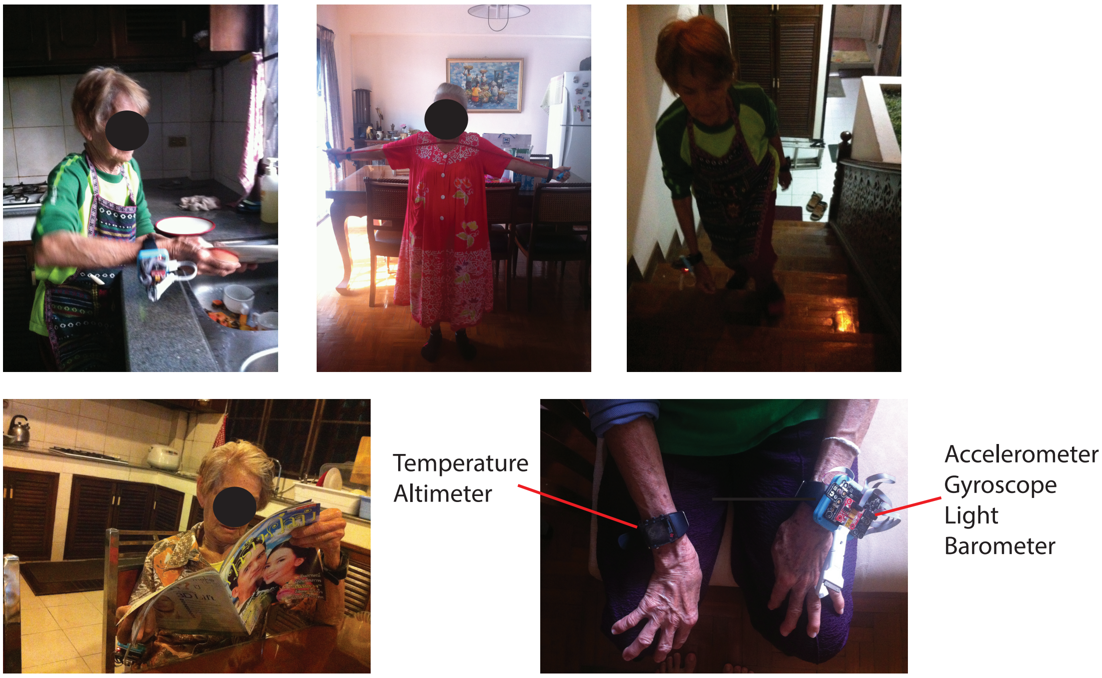

Figure 3.8: Example of a data collection session for the Wearable activity data set. The participants wore two Chronos watches and a heart rate monitor over their chests. Their faces have been blurred to reserve the anonymity.

3.5.5 Benchmark data sets

In this research, benchmark data sets are used to evaluate the proposed feature selection algorithms. The four benchmark classification data sets are used including iris, breast cancer, cardiotocography, and chess which are obtained from UCI Machine Learning Repository [64]. These data sets have been used extensively in pattern recognition literatures. The following sections give details about these data sets.

Table 3.8: Characteristics of the benchmark data sets

Data set	# Features	# Classes	Data type	# Sample
Iris	4	3	Real	150
Cancer-1992	9	2	Integer	699
Cancer-1995	30	2	Real	569
Cardiotocography-fetal	21	3	Real	2126
Cardiotocography-morp	21	10	Real	2126
Chess	36	2	Categorial	3196

1. Iris data set

   This data set has been widely used in classification literatures [60, 61]. The data set contains three type of Iris plant i.e. Setosa, Versicolor, and Verginica. There are 50 samples per each class. One class is linearly separable from the others. Two classes are not linearly separable. Four features in this data set are sepal length (cm), sepal width (cm), petal length (cm), and petal width (cm).

2. Wisconsin diagnostic breast cancer data set

   This data set has been used extensively in previous works [58, 59]. The breast cancer data set is obtained from the University of Wisconsin Hospitals, Madison [57]. This data set is collected in 1992 which shall be referred as Cancer-1992. It contains 9 integer-valued features such as clump thickness, uniformity of cell size, uniformity of cell shape, bland chromatin, etc. The values for each feature are range between 1 and 10. There are 699 samples with 65.5% benign and 34.5% malignant cases. There are 16 samples with missing attribute values. In this study, 0-value is used to replace any missing values. Another breast cancer data set which is collected in 1995 is also used in the research which shall be referred as Cancer-1995. It is composed of 30 real-valued input features computed from a digitalized image of cell nucleus such as radius, texture, smoothness, mean, standard error, etc. to determine whether the cell is malignant or benign. The data set contains 357 benign and 212 malignant samples.

3. Cardiotocography data set

   This data set has been used previously by [10]. It contains the measurement of fetal heart rate (FHR) and uterine contraction features e.g. minimum FHR histogram, percentage of time with abnormal long term variability, etc. on cardiotocograms classified by expert obstetricians. The data set contains 21 input features which can be classified into 10 types of morphologic patterns or 3 fetal states. The data set has unbalanced class distribution.

4. Chess data set
   The chess data set contains sequences of chess-description for chess end game. This data set has been previously used by [55, 56]. The data set consists of 36 categorical-input features to classify if the White can win or cannot win. The class distribution is 52% win and 48% cannot win. The data set uses a string to represent the board-description e.g. f, l, n, etc. therefore these are converted into integer values e.g. f=1, l=2, n=3, etc.

3.6 The proposed multi-sensor activity recognition framework

This section describes the proposed multi-sensor activity recognition framework (See Figure 3.9). The framework is consisted of nine subsystems including 1) sensor acquisition, 2) pre-processing, 3) segmentation, 4) feature extraction, 5) feature selection, 6) classification model construction, 7) classification, 8) classifier combination model construction, and 9) classifier combination.

3.6.1 Sensor acquisition

The framework uses seven different sensors input including accelerometer, temperature sensor, altimeter, heart rate sensor, gyroscope, barometer, and light sensor. The sensor data are wirelessly transmitted from the watch worn on the dominant wrist of the elderly person. The accelerometer and gyroscope data Acc, $Gyro$ are consisted of accelerations and rotations from three axes i.e. X, Y, and Z. The barometer data is consisted of barometric temperature and barometric pressure. The sensor data X received at any time t is:

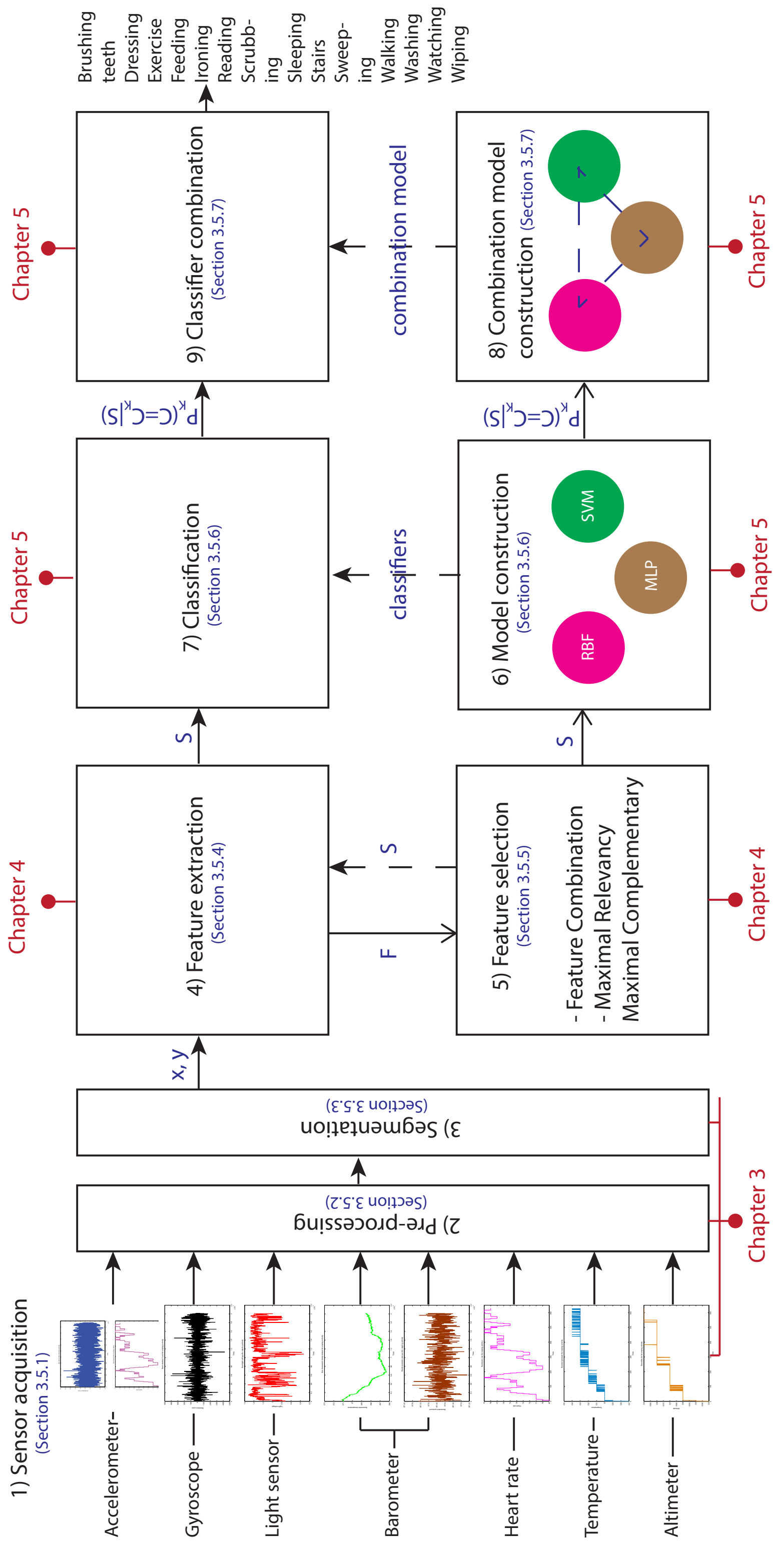

Figure 3.9: The proposed multi-sensor activity recognition framework for assisted living.

$$X_t = \{Acc^{Ub}, Temp^{Ub}, Alt^{Ub}, HR^{Ub}, Gyro^{Ub}, Baro^{Ub}, Light^{Ub}\}$$

$$Acc^{Ub} = \{Acc_x, Acc_y, Acc_z\}^{Ub}$$

$$Gyro^{Ub} = \{Gyro_x, Gyro_y, Gyro_z\}^{Ub}$$

$$Baro^{Ub} = \{Baro_Temp, Baro_Pressure\}^{Ub}$$

where Ub denotes the sensors sampling rate. For example, in this research, Ub of the accelerometer and gyroscope are set at 33 Hz, while the other sensors are set at 1 Hz. Figure 3.10 shows an example of raw data collected from the sensors.

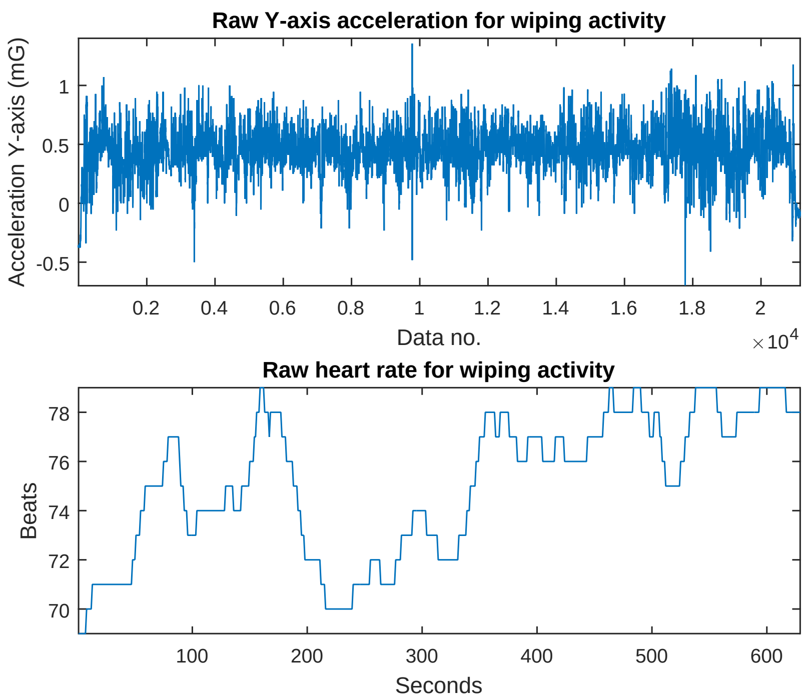

Figure 3.10: An example of acceleration and heart rate data collected from wiping floor activity.

3.6.2 Pre-processing

The sensor data is often noisy e.g. a sudden spike, especially for accelerometer and gyroscope data which may lead to the construction of a poor classification model. Therefore, to smooth the graphs and remove the outlier, the acceleration and rotation data are pre-processed using the Weighted Moving Average (WMA) technique. The WMA assigns different weights on data at different points, specifically higher weights are given to more recent data. For any set of n sensor data, the pre-processed data at time t can be calculated as:

$$X_t = \sum_{i=1}^{n} w_i X_{t-i+1}$$

$$\sum_{i=1}^{n} w_i = 1$$

In this research, two weight orders are used i.e. $w_t = 0.8$ and $w_{t-1} = 0.2$. Figure 3.11 shows the sensor data before and after applying WMA.

3.6.3 Segmentation

In order to prepare the input from the sensor data, the Sliding-window technique is used. This technique is commonly used for separating time series data into the input vector without losing information. An experiment on the different window length including 64, 128, 256 time frames, is carried out where it is decided to use a window of 3.88 seconds (128 time frames). All sensors are divided into 128-window length with 50% overlapping. For a window size l, the segmented data are:

$$X = \begin{bmatrix} 1 & 2 & \cdots & l \\ \frac{l}{2}+1 & \frac{l}{2}+2 & \cdots & \frac{l}{2}+l \\ \vdots & \vdots & \ddots & \vdots \\ \frac{i-1}{2}l+1 & \frac{i-1}{2}l+2 & \cdots & \frac{i-1}{2}l+l \end{bmatrix}$$

where:

$$0 \leq i \leq \frac{2 \times Alldata}{l} - 1$$

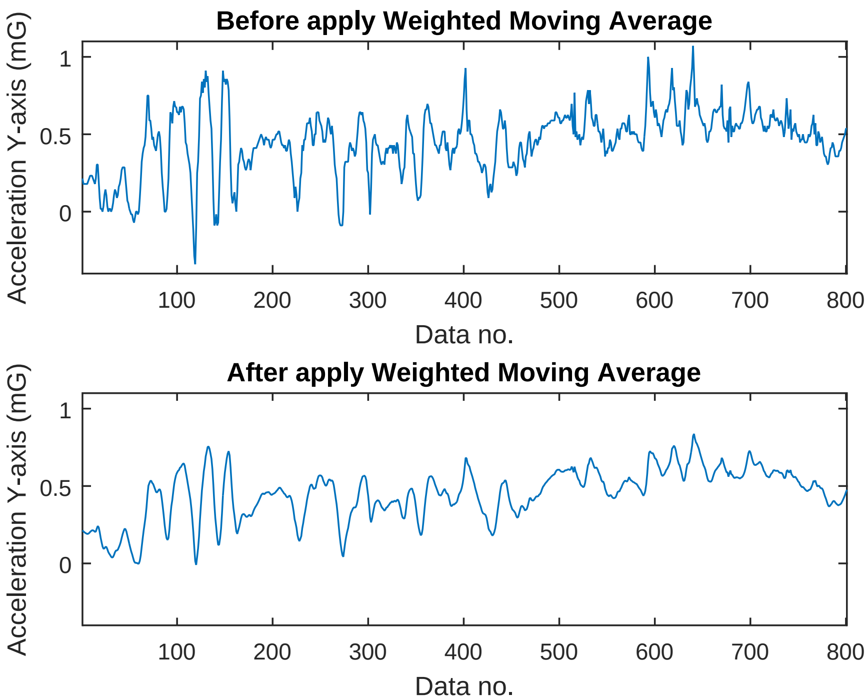

Figure 3.11: An example of acceleration data before and after applying the Weighted Moving Average.

Figure 3.12 shows an example of the sensor data before and after the segmentation using 128 window length and 50% data overlapping.

3.6.4 Feature extraction

After the segmentation, the multi-sensor input i can be represented as $X = \{(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_n, y_n)\}$ where y_i is the activity associated with data x_i. The input X is then passed to the feature extraction system which calculates information from the input in both time and frequency domain. These information are referred as features f. For example, f_1 is the mean of X. In total, a set of N features, F, is obtained where $F = \{f_1, f_2, \ldots, f_N\}$. The study of the features which are extracted from the multi-sensor input is presented in Chapter 4.

3.6.5 Feature selection

The strategy used in the proposed multi-sensor activity recognition is to extract as much information from the sensors as possible, then apply feature selection algorithm to select the optimum set of features S which explain the studied activities. The feature selection process is carried out offline to determine the selected feature set S where $S \subseteq F$. The feature set F obtained from the previous stage. During online process, the feature set S will be extracted from the multi-sensor input X.

In this research novel feature selection techniques which uses the concept of feature complementary which to the best of the knowledge have not been explored before. Normally, feature selection technique employs concepts of feature relevancy and feature redundancy to select a subset of features. Instead of using these concept, it is believed that by exploring the relationship of how a feature complements other features, a more suitable subset of features can be selected. Two feature selection algorithms are proposed which are Feature Combination (FC), and Maximal Relevancy and Maximal Complementary (MRMC). FC emphasises on the performances of a combination of features rather than single feature. It uses Clamping and forward selection to find the best combination of feature for each data set and monitor the network accuracy along so that over-

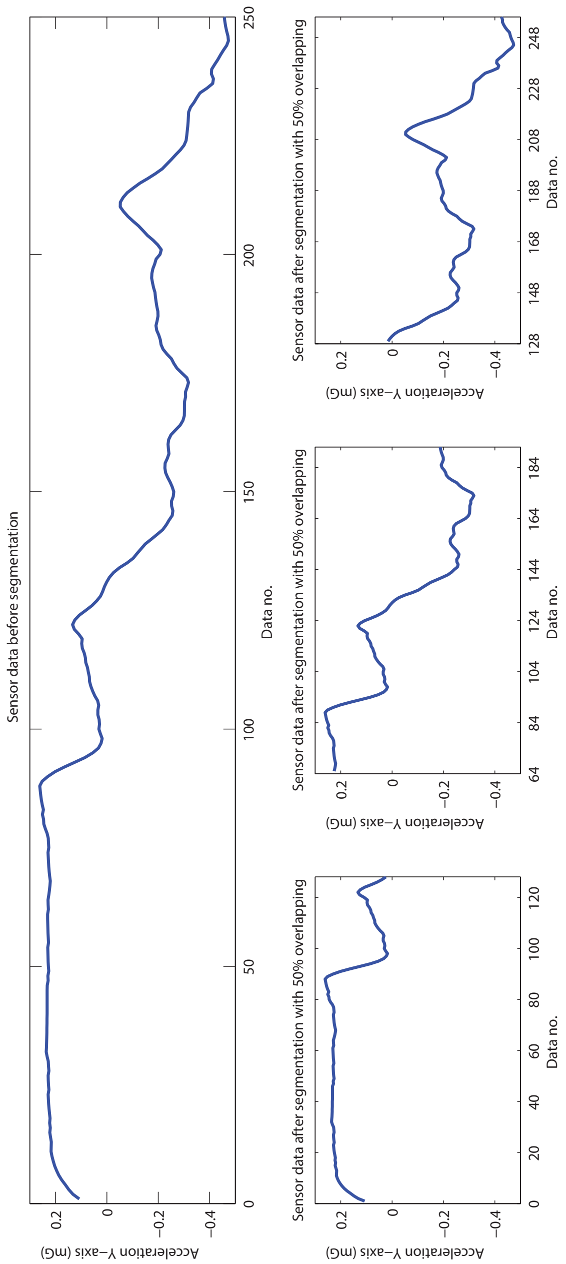

Figure 3.12: An example of acceleration data before and after apply segmentation using 128 window length and 50% overlapping.

lapped features are not selected. MRMC selects the feature based on the criteria of maximum relevance and maximum complementary. Clamping technique is employed to measure the feature relevance. A new measurement to calculate the complementary of the feature to the already selected feature set is introduced. The feature is selected based on the criteria of maximum relevance and maximum complementary. The main difference between the proposed technique and the other algorithms are that the complementary measurement is used instead of the redundancy measurement. Feature redundancy can be detected through the complementary measurement such that the redundant feature should give low complementary score. The proposed feature selection techniques are further investigated in Chapter 4.

3.6.6 Classification model construction and classification

This research investigates three classification algorithms namely RBF, MLP, and SVM. Given the input $X = (x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_m, y_m)$ where x_i contains the selected features S selected from previous stage and y_i is the activity associated with x_i where $y \in \{c_1, c_2, \ldots, c_K\}$ for K activities. The input X is passed to the classification algorithm which learns the input using different techniques to produce the decision boundaries. For example, SVM maps input into a high dimensional space using kernel functions such as linear, Gaussian, etc. and finds the decision boundary that separates two classes with the maximum margin. MLP uses the concept of connectionist where inputs and outputs are connected with weights. It contains 3 layers i.e. input, hidden, and output layer. MLP finds the optimum associated weights by trying to minimise the classification error function. It uses backpropagation technique to learn and adjust the weight. RBF is similar to MLP, however it uses the radial basis function, which is a function that depends on the distant from some point to the centre, as the activation function. The construction of classification models is done offline. In online stage, after unknown input $\hat{x}$ is applied to the classification model, the probability of that input belongs to each class c_i, $P(C = c_i|x)$ are returned. The classification result is the class c_i that has the maximum probability, $\max_K P(C = c_i|\hat{x})$. The study on these classification algorithms for activity recognition is presented in Chapter

5.

3.6.7 Combination model construction and classifier combination

From the study and literature review, it is found that there is no best classifier which is suitable for all data sets. With this in mind, the idea is to combine several classifiers in order to improve the classification accuracy. Given the input received from previous stage from a model, $M_i = \{P(C = c_1|x), P(C = c_2|x), ldots, P(C = c_K|x)\}$, is a set of the probabilities that input x belongs to class c_1 to c_K. The aim of classifier combination is to combine these probabilities together using weight and combiner functions. Given m classification models, the combination model can be expressed as $Com = w_1M_1 \otimes w_2M_2 \otimes w_iM_i \ldots w_mM_m$ where w_i is the combination weight calculated from a weight function for classifier M_i and $\otimes$ is the combiner function. The classifier combination model is carried out offline. In this research, the use of Genetic Algorithm (GA) to find the optimum weights for classifiers combination is investigated. Also, the combination model based on GA is proposed where GA is used to find the optimum combination between classifiers, weight functions, and combiner functions. These combinations are represented in a three-dimensional chromosome. For each bit in the chromosome, value 0 indicates absent and 1 indicates presence of each incident i.e. classifiers, weight functions, and combiner. For example, if there are 3 classifiers, 3 weight functions, and 3 classifier combiners, and the combination model M uses the first and second classifiers with the third weight function, and the first combiner function, then the combination chromosome can be represented as:

$$M = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

The GA tries to find the combination that minimise the mean combination error. The proposed technique also adds the classifier combination selection criteria such that the model with less number of classifiers is preferred. The details of the study on classifier combination are presented in Chapter 5. During online stage, the combination model produces the probabilities of the input belongs to each class and the final output is the class with the maximum probability.

3.7 Assessing classifier performance

Usually in activity recognition research, the performance of different classifier is assessed by calculating how well it performs in recognising target activities. The measurements commonly used for evaluate activity recognition algorithms [143] are such as accuracy, confusion matrix, F-score, true positive rate, false positive rate, true negative rate, false negative rate, precision, recall, area under ROC curve, and other methods such as one defined by [182]. The results of these measurements must be examined carefully as sometimes it can be deceitful e.g. increase in overall performance but decrease in particular classes. Moreover, other techniques such as cross-validation maybe needed to ensure non-overfitting model achieve a better result.

3.7.1 Cross-validation

In this study cross-validation technique is used to evaluate the performance of the proposed algorithms. Normally, for cross-validation, the dataset is separated into K sets. This is called K-fold cross-validation. Firstly, the dataset is separated equally into K sets. For each K time, keep one of the K sets out as the validation set, and one of the K for testing set, while the remaining K-2 sets are used as training set. Throughout this research, 10-fold cross-validation is used, otherwise stated.

3.7.2 Standard quantitative measurements

As mention earlier, there are many measurements used for assessing classifier performance. An informative table which can be used for calculating other measurements is called the Confusion Matrix. The confusion matrix is composed of information regarding the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Example of a confusion matrix is Table 3.9.

Table 3.9: Confusion matrix

		Predict	
		+	-
Actual	+	TP	FN
	-	FP	TN

1. Accuracy

   Accuracy is often used as main measurement in activity recognition. It measures the percentage that the algorithm can correctly detect the samples. It can be easily calculated from the confusion matrix as follow:

$$\begin{aligned} Accuracy &= \frac{\text{Number of correctly classified samples}}{\text{Total number of samples}} \\ &= \frac{(|TN| + |TP|)}{(|FN| + |FP| + |TN| + |TP|)} \end{aligned}$$

2. Error rate

   If in our application we consider all error to have the same effect, error rate can be calculated as:

$$\text{Error rate} = \frac{(|FN| + |FP|)}{(|FN| + |FP| + |TN| + |TP|)}$$

3. Misclassification type

   In a multi-class classification problem, the confusion matrix can be used to pinpoint what types of misclassification occur i.e. if there are any classes that often confused.

4. ROC curve

   Confusion matrix can also be used to draw the ROC curve. ROC curve shows a hit rate versus false alarm rate which is $\frac{|TP|}{|TP|+|FN|}$ VS $\frac{|FP|}{|FP|+|TN|}$ [113]. According to ROC curve, it can be seen that by increasing hit rate, false alarm also increases. Using this information, we can decide a point on this curve to suit our application.

5. Precision and recall
   Precision and recall are measurements used for evaluating the correctness of a classifier. Recall measure emphasises on finding true class accuracy while precision measure how correctly of that positive prediction.

$$Recall = \frac{|TP|}{|FN| + |TP|}$$

$$Precision = \frac{|TP|}{|FP| + |TP|}$$

6. F-score
   F-score measurement is seen as an extended measurement of accuracy. While accuracy ignores the false positive results which means that it cannot differentiate if the classifier is being discriminative to a particular class, F-score does not.

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

3.7.3 Statistical test

In this research, statistical tests are employed to test the hypotheses e.g. to compare the performances between algorithms. All statistical tests are carried out using 95% confidence interval. This section presents the general work flow used to decide the appropriate statistical tests. Firstly, the data is tested whether it has normal distribution using the Shapiro-Wilk test. If the data has normal distribution, then we determine if the data are related. For example, if the data set is the classification accuracies using three sensors on different algorithms, then the data are related. If the data is related, then we look at the number of variables. For the data set with two related variables, the paired T-Test is used. For the data set with more than two variables, the Analysis of Variable (ANOVA) with repeated measures is used. If the variables are independent, then the T-Test is used if there are two variables or the ANOVA if there are more than two variables. If the data is not normal distribution, then a non-parametric test should be used. Similarly to the parametric tests, first we determine if the variables are related. For testing two related variables, the Wilcoxon test is used. For more than two related variables, the Friedman test is used. For independent variables, the Mann-Whitney U is used if there is only two variables, otherwise the Kruskal-Wallis H should be used. Figure 3.13 summarises the flow how the statistical tests should be selected.

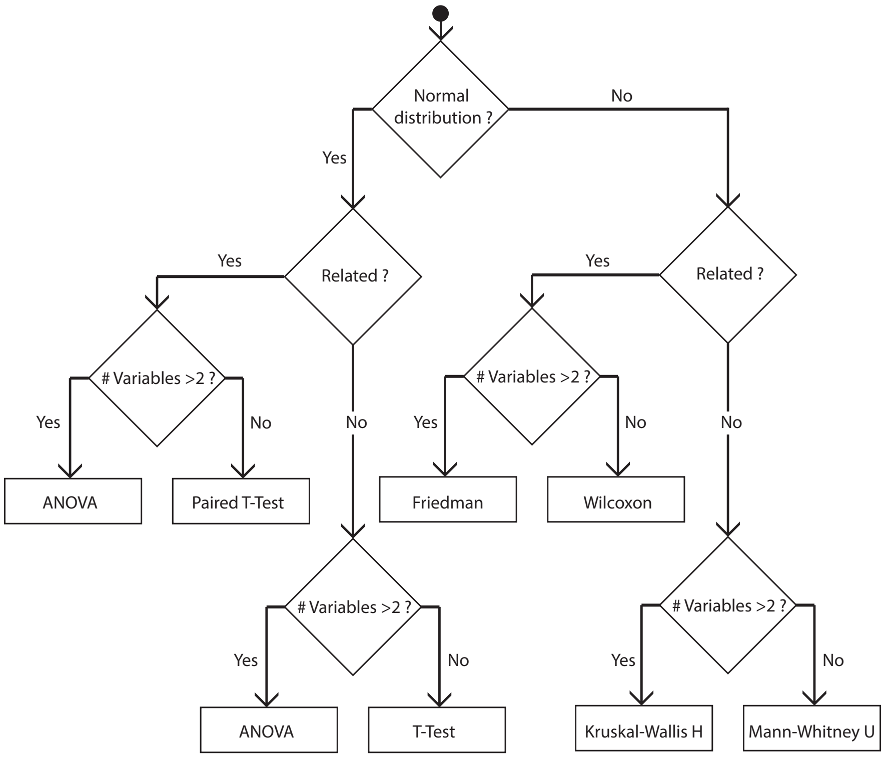

Figure 3.13: The flow chart shows how the statistical tests are chosen in this research.

3.8 Comparison challenges

It is difficult to compare between different activity recognition systems. This is due to variability in system components such as sensor, number of sensors, approach, recognised activity, number of participants, and data set. This section discusses how these factors affect the performance of the activity recognition system.

1. Number of sensor
   According from the literatures, usually the more sensors used in the system, the higher classification accuracies. This is because more information is given to the classifier. Also, sensor type and sensor location can influence the accuracy of the system. Certain sensor types and locations may be more difficult to model human activity than others.

2. Approach
   There are two main approaches in activity recognition which are object based and wearable sensor based. The object-based approach normally obtains higher classification accuracy. This also linked with the number of sensor factor as the object-based approach normally uses a vast number of sensors deployed in the environment.

3. Activity
   The system which recognises complex activities e.g. ADL will normally have lower classification accuracy then the system which recognises simple activities e.g. postures. This also linked with the approach used for activity recognition. For example, to use the wearable sensor-based approach to recognise detailed activities will be more complex than using the object-based approach. The number of the recognised activities is also important as it is more difficult to recognise a more number of activities.

4. Participant
   The number of participants in the data set can influence the classification accuracy. The activity recognition model developed based on a limited number of participants i.e. one or two persons will be less generalise and reliable comparing to the model which is based on a larger group of participants. Especially if the model is based on a single person, the expected classification accuracy will be very high as it is a dependent model.

5. Data set
   In reality, the activities data set is unbalanced. However, in term of classification accuracy, the model developed from an unbalanced data set could

have a higher classification accuracy comparing to balanced data set. However, the result obtained from the unbalanced data set is misleading. For example, if the unbalanced data set contains two classes i.e. normal and abnormal with 98% and 2% class distribution, respectively. The model can achieve classification accuracy of 98% without needing to recognise abnormal activity. This result is misleading as the model should be able to detect both classes.

6. Evaluation method
   There are many techniques which are used to evaluate the activity recognition models performances e.g. cross validation, hold out, subject-dependent, subject-independent, etc. as well as different classification measurements e.g. accuracy, specificity, precision, recall, error, F-score, etc. Due to these varieties, the comparison between different activity recognition systems can be a challenging task. In order to fairly compare the systems, they should use similar evaluation techniques and measurements.

3.9 Comparison strategy used in this research

It can be seen from the previous section that different activity recognition can be varies from several factors. This makes it difficult to compare between different activity recognition systems fairly. This section explains the strategies that will be employed in this research to compare results against other activity recognition systems. The following performance measures that can be computed from the confusion matrix will be used to evaluate recognition algorithms: overall accuracy, accuracy per activity, F-Measure per activity, and the confusion matrix itself. These standard performance measures have been explained in Section 3.7.2. The results of the experiments are also compared against other activity recognition studies throughout the thesis. System architectures such as sensor, approach, activity, participant, data set and evaluation techniques of each study will also be taken into account when perform comparison.

3.10 Summary

This chapter presents the architecture of the proposed multi-sensor activity recognition framework. The justifications of the choices of sensors and sensor location are identified. Also, the details of wearable sensor development and set up are described. A detailed description of the proposed framework are presented which covers sub-systems including sensor acquisition, pre-processing, feature calculation, feature selection, classification, and classifier fusion. In this research, feature selection and classification are mainly focused as they are the main key in a successful activity recognition. The research aims to evaluate different feature selection algorithms and to propose novel feature selection algorithms. An extensive experiments are carried out and the details are presented in Chapter 4. In this research, novel classifier fusion techniques are also proposed which are presented in Chapter 5. Three activity data sets have been collected using the developed sensor platform. The chapter also presents how these data sets have been acquired, along with their descriptions i.e. participants, data distributions, and activities' descriptions. Finally, the chapter presents assessment measurements that are used for evaluation in this research. The challenges in comparison of different activity recognition systems have been identified and the strategies employed to overcome these challenges.

Chapter 4

Features and Feature Selection Study

This chapter investigates the feature selection based on the architecture proposed in Chapter 3. It proposes two feature selection approaches i.e. Feature Combination (Section 4.2), and Maximal Relevance Maximal Complementary (Section 4.3). The extensive experimental studies are conducted to demonstrate the proposed methods. Some parts of the work in this chapter have been published in [1, 2, 3, 8, 9].

4.1 Feasibility study

4.1.1 Study hypothesis and objectives

This research proposes multi-sensor AR based on wrist-worn sensors in order to achieve a practical and high accuracy classification solution. Based on the literature review, it is hypothesised that a wrist worn sensor can be used to recognise human activities. In this study, a feasibility study of the use of wrist-worn sensor to detect human activities is carried out with two main aims. The first aim is to evaluate the feasibility and identify limitations of the approach. The second aim is to investigate different features and classification algorithms studied in the literatures and classifiers for the design of multi-sensor AR in later stage. The objectives of this study are:

1. To study the feasibility of using wrist-worn sensor for human AR.
2. To investigate and identify features suitable for AR.
3. To investigate popular classification techniques namely DT and NN.

4.1.2 Experiment design and data set

This study is carried out using Young Activity data set which is a small data set based on seven young participants (See the details in Section 3.5.2) performing five basic activities including running, walking, standing, sitting, and lying down. WEKA software is used for Correlation-based Feature Selection and classification. The study is carried out using 5-fold cross validation and the results are averaged over 10 runs. A paired t-test with 95% confidence level is used to test the statistical difference between results.

4.1.3 Methodology

4.1.3.1 Pre-processing

The raw sensor data contains noise and consequently signal pre-processing is required. Weighted moving average is used to filter the outlier data (See detail in Section 3.6.2). An example of raw and pre-processed accelerometer data are illustrated in Figure 4.1. For each sample i_{th}, the norm is calculated using $A_i = \sqrt{x^2 + y^2 + z^2}$. Norm represents acceleration size which is used to calculate feature.

The processed data is divided into windows of 128 samples with 50% overlapping (See detail in Section 3.6.3). The size of the window is selected based on [144] where the classifier performance does not increase with window size larger than 128 frames. In total, 1,070 patterns are used in this study and the distribution of each class are sitting 21%, standing 26%, walking 18%, lie down 27%, and running 9%.

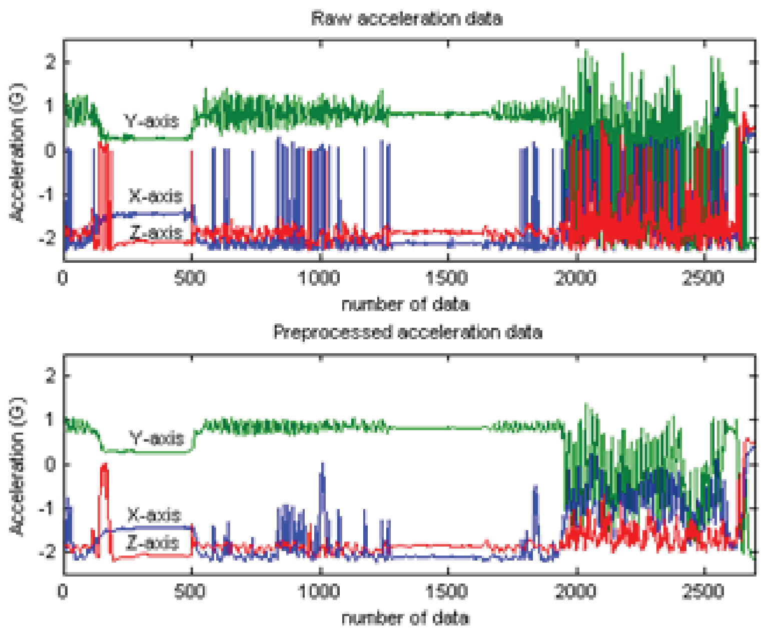

Figure 4.1: An example of triaxial accelerometer data on various activities. Top: Raw sensor data. Bottom: pre-processed data

4.1.3.2 Feature extraction

In this study, a number of features from both time and frequency domains are investigated. Seven features are selected from a survey literature [140]. The selection is based on the two highest test accuracies from each activity scenario. For example, difference and minimum achieved the best results in three activities scenario. A further 8 features which are normally used in accelerometer-based AR [106, 116, 129, 130, 155, 157] are selected. In total, this study investigates 13 features as shown in Table 4.1. For the spectral energy feature, the energy of signal between 0.3 Hz and 6 Hz are calculated as these frequencies include most of the information found in daily activities signals [144]. To calculate coefficient sum feature, the summation of the signal coefficients from 0.5 Hz to 3 Hz are used as it can discriminate between activities like running and walking [140]. For each pattern, 13 features are extracted and analysed using Matlab and the detailed information on these features are described in [140].

Table 4.1: A list of features calculated from Young activity data set

Description	Domain		Total Features
	Time	Frequency	
Features from [140]	Minimum Difference x Difference y Difference z	Spectral energy Spectral entropy Coefficient sum	7
7 Commonly used features	Mean Standard deviation Variance Correlation x, y Correlation x, z Correlation y, z	Spectral energy Spectral entropy	8

4.1.3.3 Feature extraction

Three different algorithms are used for feature selection namely DT C4.5, feed-forward backpropagation NN (ANN) and Correlation-based Feature Selection (CFS) . For C4.5 and ANN, classifications are carried out using each feature. For C4.5, the confidence level which determines the amount of tree pruning is set to 0.25. Two hidden neurons are used for each ANN classifier and learning rate is at 0.3 which was selected based on experimental results. The features are ranked based on their accuracy. The classification is carried out using 60% training and 40% testing data and the accuracy is averaged over 10 runs. The classification results and rankings are shown in Table 4.2.

For CFS technique, each feature is evaluated on the correlation between feature and classes. Using Best-First search, features which are highly correlated to the classes but low correlated to other features are selected. After applying the CFS technique using WEKA, 6 features are selected namely mean, minimum, correlation x, y, difference x, difference y and key coefficient sum.

Using ranking results from both C4.5 and ANN, 24 different feature sets are created. For example, set A contains a feature ranked 1, set B contains features ranked 1 and 2, and so on. Including the result from CFS, there are 24 sets in total ranging from 1 up to 13 features and the list of the feature sets are shown in Table 4.3. Sets C to M are built from the results from C4.5 ranking and sets N to W are built from ANN ranking results. Note that, C4.5 and ANN produce the same rank to minimum and mean, hence only 2 sets i.e. sets A and B are added. Set M includes all features and set X is the result from CFS evaluation.

Table 4.2: Feature ranking using C4.5 and ANN

Feature	C4.5		ANN	
	Accuracy (%)	Rank	Accuracy (%)	Rank
Minimum	79.87	1	65.75	1
Mean	74.64	2	63.64	2
Key coefficient sum	67.31	3	45.23	7
Energy	65.07	4	31.73	11
Entropy	60.05	5	26.62	13
Difference x	54.77	6	48.08	4
Difference y	54.63	7	50.78	3
Standard deviation	53.63	8	46.14	6
Variance	53.63	9	38.90	8
Difference z	52.81	10	47.75	5
Correlation x, y	38.29	11	37.17	9
Correlation y, z	33.41	12	34.00	10
Correlation x, z	29.54	13	28.37	12

4.1.3.4 Classification

Two classifiers: DT C4.5 and Feed-forward ANN are used in this study. An investigation on the performances of C4.5 and ANN in both feature selection and classification processes is carried out. The associated algorithms are discussed as follows:

1. Decision Tree

   DT [113] is a hierarchical model that recursively separates the input space into class regions. It composes of decision nodes and leafs in which each node m has a test function $f_m(x)$. Given a node, a test function is applied to the input and depending on the output one of the branches is taken. This process is repeated until the one of the leaves is reached.

   The learning algorithm of the DT is greedy where it locally finds the best attribute to split the data and keep repeating until it unable to separate further. Its aim is to find the smallest tree possible and in order to achieve that it finds the best attribute that would make the data after the split as pure as possible. The purity is measured by a function called Entropy. For K classes, the entropy at node m is calculated as:

   $$E_m = -\sum_{i=1}^{K} \frac{N_m^i}{N_m} \log_b \frac{N_m^i}{N_m}$$

Table 4.3: A list of feature sets used created from rankings generated by C4.5 (Set A to M) ANN (Set N to W), and CFS (Set X)

Set	Features
A	Minimum
B	Minimum, Mean
C	Minimum, Mean, Key coefficient sum
D	Minimum, Mean, Key coefficient sum, Energy
E	Minimum, Mean, Key coefficient sum, Energy, Entropy
F	Minimum, Mean, Key coefficient sum, Energy, Entropy, Difference x
G	Minimum, Mean, Key coefficient sum, Energy, Entropy, Difference x, Difference y
H	Minimum, Mean, Key coefficient sum, Energy, Entropy, Difference x, Difference y, Standard deviation
I	Minimum, Mean, Key coefficient sum, Energy, Entropy, Difference x, Difference y, Standard deviation, Variance
J	Minimum, Mean, Key coefficient sum, Energy, Entropy, Difference x, Difference y, Standard deviation, Variance, Difference z
K	Minimum, Mean, Key coefficient sum, Energy, Entropy, Difference x, Difference y, Standard deviation, Variance, Difference z, Correlation x, y
L	Minimum, Mean, Key coefficient sum, Energy, Entropy, Difference x, Difference y, Standard deviation, Variance, Difference z, Correlation x, y, Correlation y, z
M	Minimum, Mean, Key coefficient sum, Energy, Entropy, Difference x, Difference y, Standard deviation, Variance, Difference z, Correlation x, y, Correlation y, z, Correlation x, z
N	Minimum, Mean, Difference y
O	Minimum, Mean, Difference y, Difference x
P	Minimum, Mean, Difference y, Difference x, Difference z
Q	Minimum, Mean, Difference y, Difference x, Difference z, Standard deviation
R	Minimum, Mean, Difference y, Difference x, Difference z, Standard deviation, Key coefficient sum
S	Minimum, Mean, Difference y, Difference x, Difference z, Standard deviation, Key coefficient sum, Variance
T	Minimum, Mean, Difference y, Difference x, Difference z, Standard deviation, Key coefficient sum, Variance, Correlation x, y
U	Minimum, Mean, Difference y, Difference x, Difference z, Standard deviation, Key coefficient sum, Variance, Correlation x, y, Correlation y, z
V	Minimum, Mean, Difference y, Difference x, Difference z, Standard deviation, Key coefficient sum, Variance, Correlation x, y, Correlation y, z, Energy
W	Minimum, Mean, Difference y, Difference x, Difference z, Standard deviation, Key coefficient sum, Variance, Correlation x, y, Correlation y, z, Energy, Correlation x, z
X	Mean, Minimum, Correlation x, y, Difference x, Difference y, Key coefficient sum

where N_m^i is the number of data that belongs to class i at node m, N_m is the number of data at node m, and b is the log base usually is 2 or e. The entropy $\acute{E_m}$ after the split by an attribute A which has n values:

$$\acute{E_m} = -\sum_{j=1}^{n} \frac{N_{mj}}{N_m} \sum_{i=1}^{K} \frac{N_{mj}^i}{N_{mj}} \log_b \frac{N_{mj}^i}{N_{mj}}$$

where N_{mj} is the number of data at node m that has value j, and N_{mj}^i is the number of data that belongs to class i at node m and has value j.

The DT searches for the attribute that would create the largest reduction of entropy after the split. To avoid overfitting in a DT, post-pruning is usually performed where a subtree that causes overfitting is deleted.

2. Artificial Neural Network (ANN)

   ANN utilises the concept of a nervous system consisting of several input nodes (dendrites) that are connected (through synapses) to several output nodes (axons). The basic processing unit in ANN is perceptron x_i which is associated with a connection weight W_i. The output of the network is calculated from an activation function, usually a sigmoid function i.e. hyperbolic tangent, of the weighted sum of n perceptrons that linked to the output plus a bias weight:

$$y = f(\sum_{i=1}^{n} W_i x_i + W_0)$$

   Adjusting the weight to minimise the error of the output, any relationship between inputs and outputs could be modelled. For AR, a feed forward MLP (see Figure 4.2) is often used as it can implement nonlinear discriminants. An MLP with one hidden layer can be used to approximate nonlinear function.

In this study, the classification is conducted using WEKA software. For the ANN classifier, a feed-forward backpropagation algorithm is used where different numbers of hidden node are trained and tested for all 24 feature sets. The numbers

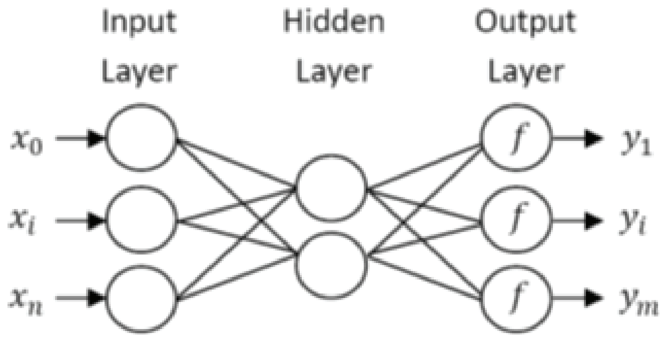

Figure 4.2: A plot between accuracy and number of features.

of hidden neurons ranging from 2 to 30 nodes with the increment of 2 are trained and tested and the learning rate of 0.3 is used. In the DT C4.5, different values of confidence factor used for tree pruning are tested. The optimum neural network models and DT models of each feature set are later compared. All of the tests are carried out using 5-fold cross validation and the results are averaged over 10 runs. A paired t-test with 95% confidence level is used for results comparison.

4.1.4 Results

4.1.4.1 Feature selection

The classification results of each feature (see Table 4.2) show that Minimum is the best feature achieving accuracy of 79.87% using C4.5 and 65.75% using ANN. Also, when observing the area under ROC curve (AUC), classification using minimum has a larger area and thus a better average performance. Classification using mean also gives a comparable result. However, C4.5 using correlation between x and z gives the worst result of only 29.54% accuracy. In the case of ANN, Entropy gives the worst result. When observing the histograms and scatter plots using the Entropy; and Correlation x, z features, it is apparent that the classes are highly overlapped, especially for Correlation x, z feature. Consequently, it is difficult for the classifier to find the decision boundaries using these features, resulting in poor accuracy. When inspecting the F-score of each feature, it is found that Difference y is the best feature for discriminating running from other activities.

It also achieved very good result in discriminating walking activity. Minimum and mean of norm acceleration can separate sitting from other activities quite well and are also useful for discriminating sitting from standing.

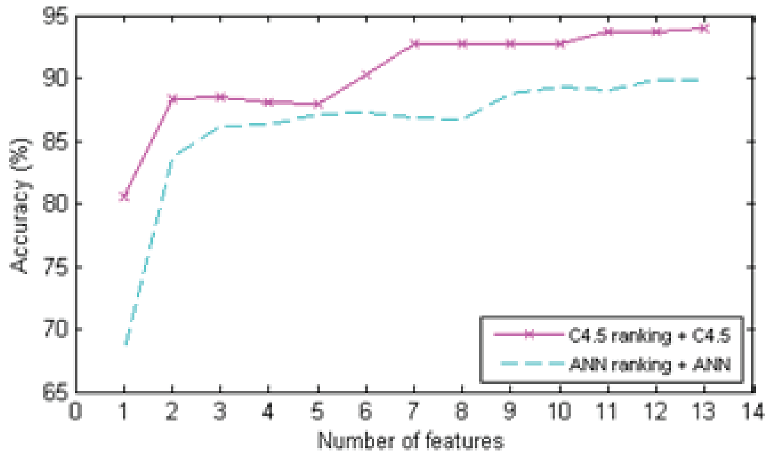

Figure 4.3: A plot between accuracy and number of features.

Using the ranking results, 23 sets (Set A to W) have been created. C4.5 with 0.25 confidence level is used on feature set A to M and ANN with 0.3 learning rate is used on feature set N to X. The number of hidden nodes is between 4 and 7 depending on the number of features used. From Figure 4.3, it is observed that by adding new features, the overall accuracy also increased. The accuracy significantly improves by increasing the number of features from 1 to 2. Using features from ANN ranking, there is no statistically significant improvement after combining more than 3 features i.e. minimum, mean, and difference y. In the case of C4.5, the improvement in accuracy after combining 8 features i.e. minimum, mean, key coefficient, energy, entropy difference x, difference y and standard deviation is not statistically significant.

The classification result using only one feature is disappointing. Although, using the best feature i.e. minimum, the best accuracy is 72.23% classified by

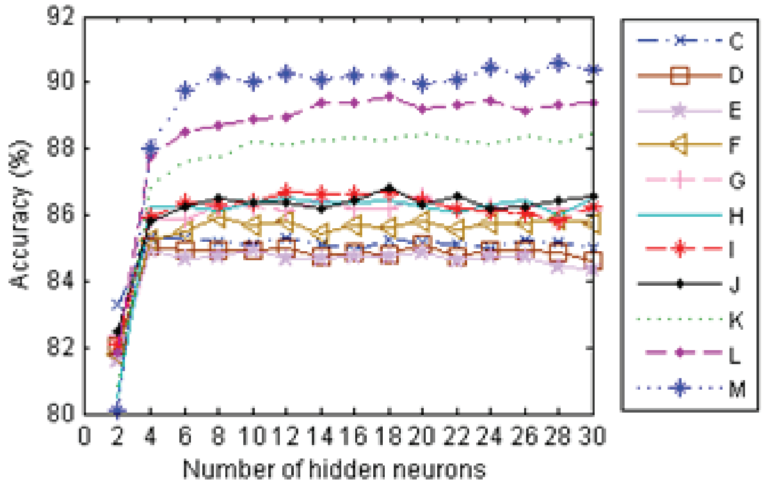

Figure 4.4: Accuracy of ANN on feature set C to L and M.

ANN with 12 hidden neurons and 80.59% classified by C4.5 with 0.25 confidence level. From Table 4.3, feature set A to M is built based on the result of C4.5 for feature ranking. The classification results of using ANN on some of these feature sets are depicted in Figure 4.4.

Different numbers of hidden neurons ranging from 2 to 30 are tested in order to find the optimum model. The accuracy of ANN significantly improves when the hidden neuron increased from 2 to 4. Set E which contains 5 features gives the worse accuracy of 84.95%. Sets G, H, I and J produce similar accuracies with average of 86.65%. Better accuracies are obtained when using more than 11 features. Classification using feature set M which contains all features selected from ANN ranking, is superior to other sets. The optimum number of hidden neurons for each feature set is selected based on the highest accuracy achieved. Table 4.4 shows the ratio of the number of hidden neurons per number of features and classification results. For features from C4.5 ranking, apart from sets A and B, the average number of hidden neurons per feature is 2.10. ANN with 28 hidden neurons achieved the best accuracy of 90.57% using 13 features from set M.

From Table 4.3, sets A, B, M to W are created according to the feature ranking using ANN. Again, different configurations of ANN are tested on these feature sets and some results are depicted in Figure 4.5. The accuracy of ANN improves

Table 4.4: ANN classification result using features from C4.5 ranking and ANN ranking

Feature from C4.5 Ranking			Feature from ANN Ranking		
Set	Confidence level	Accuracy (%)	Set	Number of hidden neurons per feature	Accuracy (%)
A	12.00	72.23	A	12.00	72.23
B	6.00	84.36	B	6.00	84.36
C	2.00	85.32	N	3.33	86.64
D	5.00	85.10	O	3.00	86.42
E	0.80	84.95	P	4.40	87.36
F	1.33	85.91	Q	1.00	87.55
G	2.86	86.57	R	2.29	87.07
H	1.50	86.51	S	1.00	87.08
I	1.33	86.70	T	2.22	88.79
J	1.80	86.83	U	3.00	89.77
K	2.73	88.43	V	2.36	89.49
L	1.50	89.55	**W**	**2.50**	**90.75**
M	**2.15**	**90.57**	M	2.15	90.57

significantly when more than 6 hidden neurons are used. The classification results of using set P, Q, R, and S (using 5-10features, respectively) are similar at nearly 87.27% on average. The accuracy of ANN using set T is slightly better, however there is no statistical difference. ANN using all features except entropy (set W) is statistically better than other sets obtaining the highest accuracy of 90.75% (see Table 4.4). The average number of hidden neurons per feature is 2.49 (set M to W) which is 16% higher than those sets obtained from C4.5 ranking (set C to M).

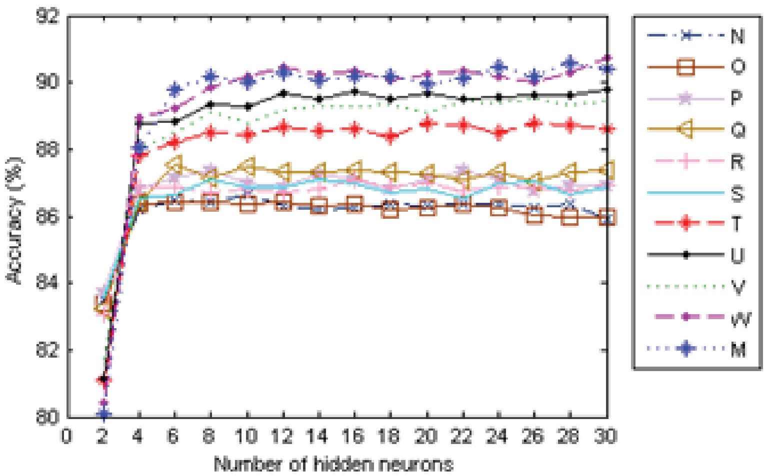

Figure 4.5: Accuracy of ANN on feature set N to W and M.

When comparing ANN classification results, it appears that using feature sets from ANN ranking (Set M to W) produce statistically better accuracy. The AUC also exhibited similar results. Figure 4.6 shows ANN classification on features from CFS ranking (set X containing 6 features) achieved 88.87% accuracy which is statistically better than features from C4.5 ranking (set F containing 6 features). However, there is no statistical difference between ANN ranking (set Q containing 6 features).

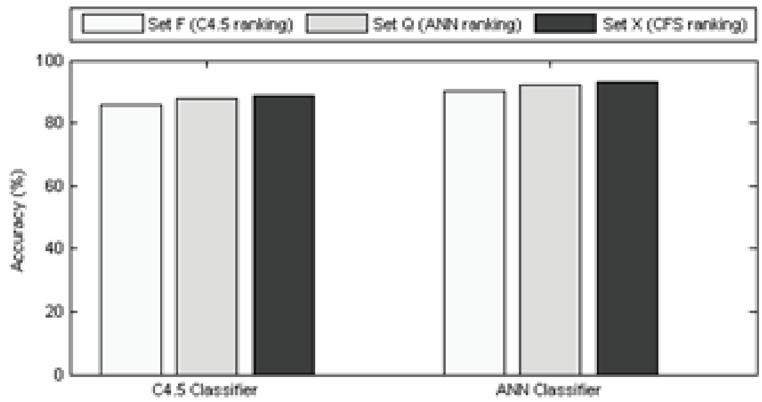

Figure 4.6: Comparison of classification results using 6 features from C4.5 ranking, ANN ranking and CFS evaluation.

Classifications using C4.5 on these feature sets has also been carried out. Different configurations on confidence level, which is used for tree pruning, from 0.15 to 0.95 using increment of 0.1 are tested in order to find optimal models. The confidence level is used in Weka DT classifier where lower confidence level means higher pruning. The results of classification accuracy and AUC show that there is no statistical difference when the confidence level changed, specifically when more than 0.55 confidence level is used. The optimal C4.5 models for each feature set are selected based on the confidence level that achieved highest accuracy. Table 4.5 shows the configurations and accuracy achieved for each feature set. The DT using feature set W using 11 features achieved 94.17% accuracy which is the highest among other sets.

Figure 4.7 illustrates C4.5 classification results on different feature sets. The results show that using features from ANN ranking gives higher accuracies. Over-

Table 4.5: C4.5 Configurations and classification results using features from C4.5 ranking and ANN ranking

Feature from C4.5 Ranking			Feature from ANN Ranking		
Set	Confidence level	Accuracy (%)	Set	Number of hidden neurons per feature	Accuracy (%)
A	0.25	80.59	A	0.25	80.59
B	0.25	88.63	B	0.25	88.63
C	0.15	88.94	N	0.45	91.31
D	0.15	88.70	O	0.35	92.17
E	0.15	88.68	P	0.25	92.12
F	0.25	90.33	Q	0.45	92.02
G	0.25	92.69	R	0.35	92.75
H	0.25	92.71	S	0.35	92.75
I	0.25	92.71	T	0.35	93.79
J	0.25	92.70	U	0.25	93.73
K	0.35	93.69	V	0.25	93.74
L	0.25	93.69	**W**	**0.15**	**94.17**
M	**0.15**	**94.11**	M	0.15	94.11

all, the accuracies increase when number of feature increases. However, there are slightly decreases in accuracies in some of the feature sets from C4.5 ranking e.g. after energy and entropy are added in set D and E. For C4.5 ranking features, the classification accuracies significantly improve when using more than 7 features. For ANN ranking features, there is no statistical difference in accuracy after using more than 9 features.

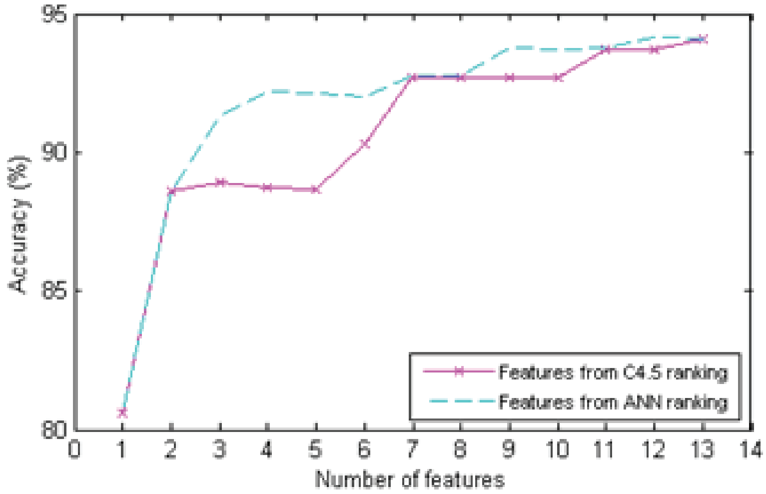

Figure 4.7: Accuracies of C4.5 classification using features from C4.5 ranking and ANN ranking.

C4.5 classification using features from CFS ranking (Set X containing 6 features) gives 93.03% accuracy which is better than using features from C4.5 ranking (Set F containing 6 features). However, there is no statistical difference comparing to using features from ANN ranking (Set Q containing 6 features) (see Figure 4.6).

4.1.4.2 Classification

The best configurations of C4.5 and ANN models on each feature sets are used in order to compare the performances of these two classifiers. Table 4.6 shows different classification results of ANN and C4.5 over feature set W indicating that in general C4.5 outperforms ANN. C4.5 classifiers which use features from ANN ranking perform better than ones which use features from C4.5 ranking (see Figure 4.8).

Table 4.6: Classification results using ANN and C4.5 on Set W

Classifier	TPR	FPR	Precision	Recall	F-score	AUC
ANN	0.907	0.031	0.906	0.907	0.906	**0.976**
C4.5	**0.941**	**0.019**	**0.941**	**0.941**	**0.941**	0.970

Table 4.7, Table 4.8, and Table 4.9 show examples of confusion matrix of different classification models. It reveals that classifiers are often confused between stand and lie down activity. When visually inspecting some of the features e.g. correlation x, z, standard deviation and minimum, which the classifiers use for separating these two classes, it is found that lying down and standing activities exhibit similar values. Lying down activity is sometimes misclassified as sitting or walking. The confusion matrix also reveals that C4.5 and ANN can classify run, sit and walk very well, however ANN classifier has more problem classifying lie down and stand activities.

Table 4.7: Confusion matrix of C4.5 on set W

Actual	Predict				
	Lie down	Run	Sit	Stand	Walk
Lie down	259	0	4	18	3
Run	0	90	0	0	1
Sit	7	0	217	1	0
Stand	21	0	4	249	0
Walk	2	1	0	1	192

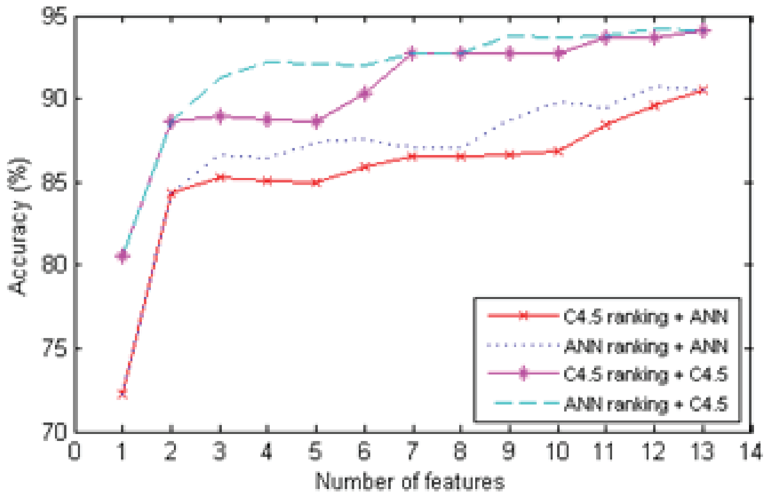

Figure 4.8: Classification results of C4.5 and ANN on different feature selection schemes.

Table 4.8: Confusion matrix of ANN on set W

Actual	Predict				
	Lie down	Run	Sit	Stand	Walk
Lie down	239	0	6	31	8
Run	0	91	0	0	0
Sit	6	0	217	2	0
Stand	37	0	5	232	0
Walk	2	1	1	1	191

Table 4.9: Confusion matrix of C4.5 on set X

Actual	Predict				
	Lie down	Run	Sit	Stand	Walk
Lie down	251	0	2	23	8
Run	0	90	0	0	1
Sit	9	0	214	2	0
Stand	15	0	5	254	0
Walk	2	1	0	0	193

4.1.4.3 Discussion

The study investigates 13 different features and the results suggest that Difference y, Minimum and Mean are the best features for classifying running, walking, sitting and standing activities. This is confirmed by the result of classification on set N, which is the combination of these three features, achieving high accuracy of 91.31%. The results also suggest that Difference y is the best feature for classifying running. As the difference in y-axis acceleration is higher when running, thus its data is distributed further from other classes.

The results of the accuracy on the different number of features show that using more features improves overall accuracy consequently using more features enhances the classifiers performance. Using only 13 features in the study, the results show no statistical difference when more than 7 features are used. Possibly using a larger set of features could diminish the classifiers performance.

In the study on feature selection, the results suggest that using ANN ranking produces a better set of features comparing to C4.5 ranking. Similar to ANN ranking, CFS method also produces a better feature set. The results are consistent even though different classifiers are used. Our results suggest that DT C4.5 classifier performs better than feed-forward backpropagation ANN. These results are also consistent despite different sets of feature used. The results are similar to [155] where similar activities are studied. The results suggest that high classification result can be obtained using ANN ranking feature selection and C4.5 classifier. ANN outperforms DT [124], however the experiment is carried out on unsupervised data, while in our work supervised data is used. Further experiments need to be carried out in order to investigate the effect of using the model of this work with unsupervised data.

4.1.5 Conclusion remarks

This study demonstrates that a single wrist worn accelerometer can be used to identify activities of a user however, only simple activities can be detected. This is because the data obtained from the accelerometer mounted on the wrist cannot provide enough information for complicated or high level activities. Also, other activities involving the use of hand e.g. eating, washing dishes, reading, etc. using only a wrist worn accelerometer to predict those activities can be complicated. Other sensors should be used to provide additional context information to supplement and reduce ambiguity from an accelerometer.

4.2 Feature selection study: Feature Combination

4.2.1 Study hypothesis and objectives

The aim of feature selection is to identify the smallest subset of input features which explains the output classes. This process is important especially to the classification problems which have a large number of input features. For example, a multi-sensor activity classification system normally contains a large number of input features generated from different sensors. Feature selection can help reduce the size of feature space which leads to reduction in computational cost and complexity in the classification system. In real world problems where input features contain irrelevant and redundant features, feature selection can help identify relevance feature set leads to improvement in classification performances. Feature selection techniques mainly focus on the relevancy of the features and classes, and redundancy between features. However, using these two criteria, features with high relevancy and redundancy may be selected. On the other hand, feature complementary concept considers if a feature complement the already selected feature set. It is hypothesised that using the feature complementary concept can identify the optimum feature set which leads to better classification accuracy comparing to other techniques which do not employ this concept. In this study, the aims are to investigate different feature selection techniques and propose a new method suitable for multi-sensor AR. The objectives of this study are as follow:

1. To investigate different feature selection techniques for multi-sensor activity recognition.
2. To propose a feature selection technique which uses feature complementary concept to select relevance features.

3. To validate the proposed technique on the Multi-sensor activity data set.

4. To compare the results with other well established feature selection algorithms.

4.2.2 Experiment design

The study is separated into two experiments according to two data sets: Multi-sensor activity data set and Wearable-sensor activity data set. In this study, the proposed feature selection algorithm called Feature Combination (FC) is compared against three popular feature selection techniques which are Maximal Relevant Minimal Redundant (MRMR), Normalized Mutual Information Feature Selection (NMIFS), and Clamping. The description and formulas of these algorithms are presented in Section 4.3.2.

In the first experiment, FC is compared against the Clamping method using the Multi-sensor activity data set. In the second experiment, FC is compared with MRMR, NMIFS, and Clamping using the Wearable-sensor activity data set.

4.2.3 Methodology

4.2.3.1 Feature generation and transformation

Experimentation in Section 4.1 demonstrates that 13 different features from both time and frequency domains are useful in human activity classification. In this work these features are also computed from the collected sensor data. However, the features are not only calculated from the acceleration magnitude $\sqrt{x^2 + y^2 + z^2}$, but also from raw accelerations of X, Y, Z axis, temperature and altitude as well. In addition, other features including maximum, RMS, and main axis are also calculated. A list of features is displayed in Table 4.10 containing 12 features from X-axis acceleration, 12 features from Y-axis acceleration, 12 features from Z-axis acceleration, 10 features from acceleration magnitude, 8 features from temperature, 8 features from altitude and 1 feature from acceleration. The calculated features are then transformed into [0 1] range. Scaling helps avoiding features with larger numeric ranges dominating features with smaller

Table 4.10: Number of features calculated from each sensor data

Sensor data	Time-domain features	Frequency-domain features
Acceleration X-axis, Acceleration Y-axis, Acceleration Z-axis, Acceleration magnitude, Temperature, Altitude	Mean, minimum, maximum, standard deviation, variance, range, root-mean-square, correlation, difference, main axis	Spectral energy, spectral entropy, key coefficient
Total Number of features	45	18

numeric ranges. It also reduces numerical difficulties during calculation [115]. In the MLP which uses the gradient descent method i.e. backpropagation, scaling can help in faster convergence.

4.2.3.2 Feature selection algorithms

In the feature selection study, two different approaches used for feature ranking which are based on MI and NN are investigated.

1. MI based feature selection MI is based on information theory proposed by [15]. It measures the dependency between two variables. MI value is zero if and only if the variables are independent. Given continuous variables f_i and f_j, the MI can be calculated as:

$$MI(f_i; f_j) = \int\int p(f_i, f_j) \log \frac{p(f_i, f_j)}{p(f_i)p(f_j)} \mathrm{d}f_i \mathrm{d}f_j$$

   In practice, it is difficult to calculate MI of the continuous values and often the variables are discretised using bins. The MI of discrete variables can be calculated as:

$$MI(f_i; f_j) = \sum_i \sum_j p(f_i, f_j) \log \frac{p(f_i, f_j)}{p(f_i)p(f_j)}$$

   where $p(f_i, f_j)$ is the joint probability of features i and j, and $p(f_i)$ is the probability of feature i. There are many feature ranking algorithms based on MI [17, 62, 63, 65]. MRMR is one of the most popular feature selection algorithms. Many algorithms have been based on MRMR. For example, NMIFS enhances MRMR by using entropy of the variables to normalize the MI values when calculating the redundancy between variables. MRMR is enhanced by using the Kernel Canonical Correlation Analysis as inputs rather than the actual features [7].

In this study, some of the commonly used feature selection algorithms based on MI which are MRMR and NMIFS are investigate.

(a) MRMR

The MRMR algorithm [62] ranks the features based on the minimal redundancy and maximal relevance criterion. It calculates the MI between two features to measure the redundancy and the MI between a feature and the outputs to measure the relevance. Using MRMR concept and greedy selection, a set of feature rankings S can be obtained as follow:

(A) Given $S = \{\}$ where S is a set of selected features and $F = \{f_1, f_2, f_i, f_j ..., f_N\}$ where F is a set of N features, selects the feature f_s in F which has the maximum mutual information between itself and output C where $C = \{c_1, c_3, ..., c_K\}$ and $f_s = \max_{f_i \in F} MI(f_i; C)$. Updates S and F.

$$S = S \cup \{f_s\} \tag{4.1}$$

$$F = F \setminus \{f_s\} \tag{4.2}$$

(B) Select feature f_s in F which satisfies the following condition:

$$f_s = \max_{f_i \in F}\{MI(f_i; C) - \frac{1}{|f_i|}\sum_{f_j \in S} MI(f_i; f_j)\}$$

Update S and F using (4.1) and (4.2).

Repeat Step (B) until the desired number of features is obtained.

(b) NMIFS

The NMIFS algorithm [63] is an enhancement of the MRMR algorithm. A Normalized mutual information (NMI) between two features are used instead:

$$NMI(i;j) = \frac{MI(i;j)}{\min\{H(i), H(j)\}}$$

where $H()$ is the entropy function. Similar steps as MRMR are carried out, however the condition in Step (B) is changed to:

$$f_s = \max_{f_i \in F}\{MI(f_i;C) - \frac{1}{|s|}\sum_{f_j \in S} NMI(f_i;f_j)\}$$

2. Neural network based feature selection

Some studies have proposed to use NN for feature selection [18, 19, 58]. For example, NNFS [19] selects features based on weights associated with that features. The weights associated with unimportant features would have values close to zero. NNFS adds a penalty term to the cross-entropy error function in order to distinguish redundant network connection. Clamping technique proposed by [18] ranks the features based on the effect to classification accuracy from clamping features. In this study, the performance of the proposed algorithm with the Clamping algorithm is compared.

The Clamping technique [18] is used to obtain the feature ranking where each feature is clamped to a fixed value (mean of each feature $\bar{x}$ is used) and the impact of the clamped network generalisation performance, $g(X|x_i = \bar{x})$ to the network generalisation performance, $g(X)$ is calculated using:

$$Im_i = 1 - \frac{g(X|x_i = \bar{x})}{g(X)}$$

Clamping the most important feature highly affects generalisation performance while the redundant features show no adverse effect. For an N data set, the rankings can be combined using the Borda Count technique which is a kind of plurality voting where each vote is associated with a specified ranking point. A given feature is associated with a point that is related to its importance. The most important feature is associated with the highest possible ranking value e.g. the total number of features. For example, in a feature space size n, feature X is the most important, hence it is associated with n points. Feature Y is the second most important thus, n-1 points is associated, and so on. Finally, the final feature ranks can be obtained by sorting the summation of the points of each feature.

4.2.4 The proposed Feature Combination

The feature ranking using the Clamping technique can only considers the performance of a single feature. Feature selection based on this ranking may discard the features which are useless in itself but help improve classification performance when combined with other features. Also, the features with high ranking may be overlapped with other high ranking features. To overcome this, a Feature Combination technique is proposed which emphasises on the performances of a combination of features rather than single feature. The idea is to use forward selection to find the best combination of features for a data set. A feature is added to the lists by its importance and difference in accuracy is calculated along each addition. By monitoring the accuracy difference, the feature which is highly overlapped with already added features will not be included into the list. This technique also allows the weaker feature which is not overlapped with existed features to be selected.

Starting from an empty list, a feature is added according to its ranking. For any current feature list using p features, mean of accuracy ($\overline{M\_Acc_p}$) of validation set is calculated and compared with mean of accuracy ($\overline{M\_Acc_{p-1}}$) of the previous feature list i.e. using p-1 features. If ($\overline{M\_Acc_p}$) is less than or equal to ($\overline{M\_Acc_{p-1}}$), then the recently added feature is removed from the list. This process is carried out until all features have been tested. For an N data set, results are combined using majority voting resulted in a new feature ranks. Figure 4.9 describes the pseudo code of the feature selection.

```
PROCEDURE Feature Combination

TRAIN Network using All feature
COMPUTE Network generalized performance
FOR each Feature in All feature
      SET Input to All features
      SET Feature in Input to mean Feature
      TEST Network using Input
      COMPUTE Clamped network generalized performance
      COMPUTE Impact of clamped network generalized performance   (Im)
END LOOP
SORT Im By DESC.
SET Rank to Im

INITIALIZE List
FOR each Feature in Rank
      ADD Feature to List
      TRAIN Network using List
      TEST Network using List
      COMPUTE Accuracy_current
      IF Accuracy_current<=Accuracy_previous
            REMOVE Feature from List
      ELSE
            SET Accuracy_previous to Accuracy_current
      END
END LOOP
```

Figure 4.9: Pseudo code of Feature Combination.

4.2.5 Data sets

In this study, two activity data sets are used: Multi-sensor activity data set and Wearable-sensor activity data set. All experiments are carried out using 10-fold cross-validation where 8 folds are used for training, 1 fold for validation and 1 fold for testing. The size of the training, validation and testing data of each fold used for different data set are shown in Table 4.11.

Table 4.11: Characteristics and data partition per fold for different data sets used in the FC study

Data set	# Features	# Classes	Data type	# Sample	# Training	# Validation	# Testing
Multi-sensor activity data set	63	9	Real	17,488	5,760	720	720
Wearable-sensor activity data set	141	12	Real	39,328	20160	2520	2520

4.2.6 Experimental results

4.2.6.1 Experiment 1: Multi-sensor activity data set

The sensor data is pre-processed using WMA and segmented at 3.88 seconds with 50% data overlapping resulting in a total of 17,843 patterns. It is noted that the number of walking upstairs and walking downstairs classes are relatively low. In the data collections, normally a participant was asked to perform an activity for a limited time i.e. 10 minutes. However, due to the physical restriction because of participants' ages, using stairs (walking up/down) activity were performed without a set time limit as to reduce risk of falling. On an average, a participant used 5 seconds to climb up the 6-step stairs. The data from walking upstairs and downstairs classes only constitute to 2% of all data set which is clearly imbalanced. This will affect classification performance where most techniques assume samples are distributed evenly among different classes. Also, an imbalanced data set poses other problems such as difficulty in establishing accurate decision boundary, error in interpreting classification results, and data from minority class tend to be treated as noise [16].

In this work, it is decided to remove data from walking downstairs and walking upstairs classes as the numbers of samples are too low to be able to discover true classes boundaries especially in our case which shows highly overlapped classes. Also, according to the interview with the participants, it is found that majority of them live in a bungalow or on ground floor while participants who live on 2-floored houses only use stairs couple times a day (to access their bedrooms). The under-sampling technique is used to obtain a new data set with the balanced number of samples from each class. All data from the smallest class i.e. dressing class are preserved. The same data size is obtained from the other 8 classes. In total, the new balanced data set contained $805 \times 9 = 7245$ patterns. This study used 10-fold cross validation which the data are randomly divided into 10 folds, one of 10 folds is used as the validation data, one of 10 folds is used as the test data and the remaining 8 folds are used as the training data in turn, the mean of the classification rates by using these 10 test data sets is used as the final classification rate.

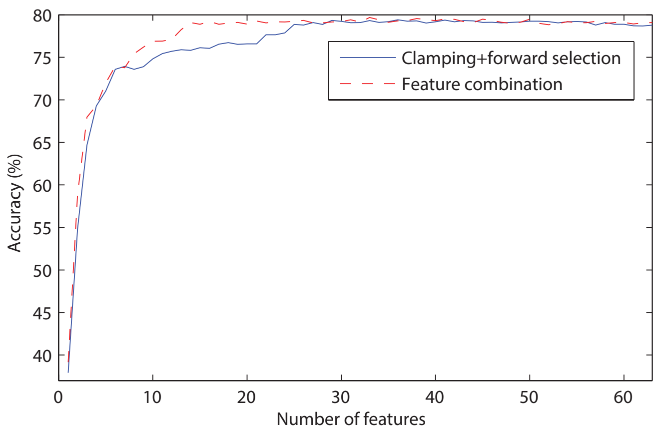

Figure 4.10: Classification accuracy between using Feature Combination and Clamping feature selection techniques

Features are then generated from the raw sensor data as described in Section 4.2.3.1. Next, a feature selection is carried out using NN and the proposed Feature Combination techniques in order to select the optimum feature subset. The NN with resilient backpropagation and 20 hidden neurons is used. The feature ranking procedure using 10-fold cross validation and 10 runs is carried out. Figure 4.10 shows that using the proposed Feature combination method can achieved higher accuracy. The final subset is obtained by observing the truncation point of the mean accuracy of all data sets. 16 features are selected as listed in Table 4.12.

Table 4.12: The selected features by Feature Combination using Multi-sensor activity data set

Sensor	Selected features		
Accelerometer	RMS Y axis	RMS X axis	Maximum Y axis
	Minimum Y axis	Difference Z axis	Maximum Z axis
	Key Coefficient Y axis	Correlation X, Y	Minimum Z axis
	Minimum X axis	Maximum norm acc.	Difference Y axis
Temperature sensor	Mean temperature	Key Coefficient temperature	Min temperature
Altimeter	Entropy altitude		

4.2.6.2 Experiment 2: Wearable-sensor activity data set

Firstly, features are ranked using the specified techniques mentioned in Section 4.2.4. The results from different runs are combined using the Borda count. Feature selection is performed using NN. A multilayer perceptron with one hidden layer is used where the hidden node is set to $\alpha\times$ number of input. Experiments are carried out to determine the appropriate value of alpha and the number of epoch where trade-off between accuracy and training time are considered.

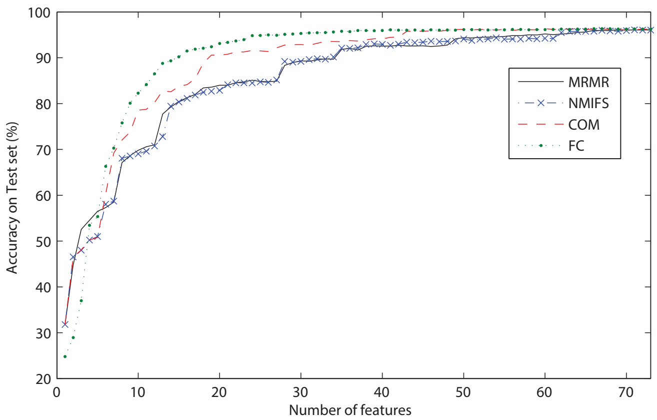

Figure 4.11: The mean classification accuracy obtained using MRMR, NMIFS, Clamping, and FC

The result of averaged accuracy on test sets is shown in Figure 4.11. From the graph, it can be seen that FC achieves the highest accuracy. The hypothesis if the accuracy difference is significant is tested. First, the data normality is tested using Shapiro-Wilk and the results indicate that these are not normal distribution. Thus, the Related-Samples Wilcoxon Signed Rank Test is applied and the result indicates that the accuracy of FC is significant higher than other techniques ($p<0.05$). COM is significant higher than MRMR and NMIFS ($p<0.05$). The difference in accuracies of MRMR and NMIFS are not statistically significant ($p = 0.315$). To sum up, the performance of the feature selection techniques can be expressed as $FC >^* COM >^* MRMR = NMIFS$ where $>^*$ indicates significantly better and $=$ indicates no significant difference at 95% confidence interval.

MRMR and NMIFS produce similar accuracy and also select similar set of features. The reason is that these two techniques are based on MI. This is evident in which MRMR and NMIFS produce similar ranking. When investigate why these two techniques cannot achieve higher accuracy, it is found that majority of the features selected at the beginning are from accelerometer and gyroscope only. Although features extracted from these two sensors contain valuable information, when using the forward selection strategy this would lead to a selection of redundant features. MRMR and NMIFS only selects features from accelerometer, gyroscope and light sensor.

On the other hand, Clamping ranking selects features from a variety of sensors such as accelerometer, gyroscope, heart rate sensor, barometer, light, and altimeter (see Table 4.13). When Clamping ranking is combined with MRMR and NMIFS (COM), it can be seen that the result has considerably improved. Besides accelerometer, gyroscope and light sensor, COM also selects features from barometer which means that this sensor provides valuable information for activity classification. Features selected from Clamping and FC are similar as FC is modified from Clamping technique. However, FC searches for only the subset of features which are complementing each other and reduce redundant features. FC clearly achieved better accuracy comparing to the other three techniques. However, according to the graph, the accuracies at the beginning are lower. Thus, in the case of data set with small number of features (fewer than 5), using MRMR should produce a better result. The truncation at 24 features is selected where the accuracy starts to remain constant. A list of selected features is shown in Table 4.13.

Table 4.13: The selected features by Feature Combination using Wearable-sensor activity data set

Sensor	Data	MRMR	NMIFS	Clamping	COM	FC
Accelerometer	X - axis	-	-	RMS, mean	RMS	RMS, mean
	Y - axis	RMS, max, median, mode, key coefficient, mean, min	Max, median, mean, mode, min	RMS, max, median, key coefficient, mode, mean	RMS, median, mean, min, mode	Max, median, mean, min, mode, RMS
	Z - axis	Min, median, mode, mean	max	RMS, mean	Mean, median, min, mode	RMS, mean
	$\sqrt{x^2+y^2+z^2}$	Intensity, max, median, mean, RMS	Intensity, RMS, max, mean	Correlation X, Z, max,RMS	Max, intensity, RMS, median, mean	Correlation X, Z, max, RMS
Temperature	-	-	-	-	-	-
Altimeter	Altitude	-	-	Min	-	Min
Heart rate monitor	Heart rate	-	-	-	-	Min
Light	Light intensity	Max	Max	Max, min	Max, RMS, mean, median	Max, min
Barometer	Temperature	-	-	Max, median, RMS, mean	Median, Max	Max, median, RMS
	Pressure	-	-	Max, median	Max	Max, median
Gyroscope	X - axis	STD, RMS	STD , mode	-	STD	STD
	Y - axis	-	-	-	-	-
	Z - axis	Std, RMS, intensity	Min, median, mode, mean	-	-	-
	$\sqrt{x^2+y^2+z^2}$	RMS, mean, median, std	RMS, mean, median	Correlation X, Y	RMS	Correlation X, Y

4.2.7 Discussion

The objective of this study is to compare the performance of 4 feature selection techniques. Our results suggest that FC is the most appropriate technique for our application. FC can select a more diversity set of features comparing to other techniques. It monitors the performance of a subset of features along the selection to make sure that redundant features are not selected. However, according to the FC algorithm, redundant features may still be selected at earlier stage and we suggest that post checking should be added to remove any redundant feature after selection. MRMR and NMIFS only measure the redundancy between 2 variables. The results of the experiment show that this measurement is not enough to detect the overlapped features. MRMR and NMIFS select features with high relevancy to classes and low redundancy with

The result of this study implies that the technique which can select a subset of features with the lowest feature redundancy is the most optimum technique.

4.2.8 Conclusion remarks

FC can select the optimum set of features comparing to MRMR, NMIFS, COM and FC as it can select features from diverse sensors which helps reduce feature redundancy. This technique can be improved by adding a post feature check to remove redundant features which may be selected in the early stage of selection process.

4.3 Feature selection study: Maximal Relevance Maximal Complementary

4.3.1 Study hypothesis and objectives

From previous section, it can be seen that using the concept of feature complementary helps improve classification accuracy. However, there are some limitations with FC techniques. Firstly, since the algorithm employs a forward selection technique, there is a possibility that the good features are eliminated in earlier stages. Secondly, redundant features can get selected in very early stages as FC performs forward selection and does not do any comparison between other features except the adjacent feature. In this study, the aim is to propose a new feature selection technique that overcomes the mentioned limitations. The proposed feature selection technique introduces relevancy and complementary measurements which are used for features ranking. It is hypothesised that using the proposed feature selection technique, a optimum feature set will be ranked and selected, comparing to other techniques which do not use the feature complementary concept. The objectives of this study are as follow:

1. To propose a feature selection technique which uses feature complementary concept that can select relevance features.

2. To validate the proposed technique using well-defined problems, benchmark data sets, and real world data sets.

3. To compare the results with other well established feature selection algorithms.

4.3.2 Feature selection algorithms

In this experiment, two different approaches used for feature ranking which are based on mutual information i.e. MRMR and NMIFS and NN i.e. Clamping are studied. Their descriptions and formulas are presented in Section 4.2.3.2. There are still some limitations on these three techniques. Clamping technique provides robust ranking even in noisy data. However, it only considers the relationship between one feature and the classes. It does not consider any relationship between the features. MRMR and NMIFS do consider the relationship between features. However, the relationship between only two features are measured. Among these three techniques, none considers how a feature would complement to the already selected features. In this experiment, a new feature selection technique which considers the relationship between feature and the class as well as the relationship among a group of features is proposed.

4.3.3 The proposed Maximal Relevance Maximal Complementary Feature Selection

The proposed feature selection method is based on the criteria of maximum relevance and maximum complementary (MRMC) of the feature. In our method, NN is employed for the calculation of the relevance and complementary score. NN is based on the concept of connectionist where several input nodes are connected with associated weights to several outputs nodes. A network with one hidden layer which is used. Given input of N features $F = \{f_1, f_2, ..., f_i, ..., f_N\}$ to predict output of K classes $C = \{c_1, c_2, ..., c_K\}$. Figure 4.12 depicts the NN architecture where b_1 is a bias input and weights $W = \{w_{11}, w_{12}, ..., w_{Nj}\}$ where w_{11} represents a weight connect from f_1 to hidden node 1, and j is the number of hidden nodes. The weights and bias are generated randomly from a univariate distribution. The network output node $\hat{y}_i$ can be calculated from the summation function [53]:

$$\hat{y}_i = g(\sum_{i=1}^{N} W^T f_i + b_1)$$

Where $g(z)$ is a sigmoid activation function. In this study, a logistic function $g(z) = \frac{1}{1+e^{-z}}$ is used. The network tries to minimize the following cost function:

$$J(W) = -\frac{1}{N}\{\sum_{i=1}^{N}\sum_{k=1}^{K} y_i^{(k)} \log(\hat{y}_i)^{(k)} + (1 - y_i^{(k)}) \log(1 - \hat{y}_i)^{(k)}\}$$

where $y_i^{(k)}$ is the predicted output for class k using pattern i. First, the two measurements i.e. the relevance and complementary used for calculating feature's score are introduced.

1. Relevancy score

   The relevancy score measures how much the feature is important to the network. By removing the feature node in the network then calculating the network's performance, the relevancy of the feature can be obtained such that if the clamped feature is important, the network performance will significantly affected. First, the base network is constructed using all the features F and its performance is used as the base line. Next, the feature f_i is removed from the network. In order to remove the feature without disrupting the whole network, a static value is used. In this study, a mean value of the feature is used ($f_i = \bar{f}_i$) as has been used successfully in [18]. This network is referred as the relevancy network. After the feature is removed, the network performance is re-calculated and evaluated with the base line performance. Figure 4.12 shows the architecture and concept of the base line network and the network with the removed feature.

   Given a set of feature F, the relevance of the feature Rel_{f_i} is calculated by:

$$Rel_{f_i} = 1 - \frac{P'(F|f_i = \bar{f}_i)}{P(F)} \tag{4.3}$$

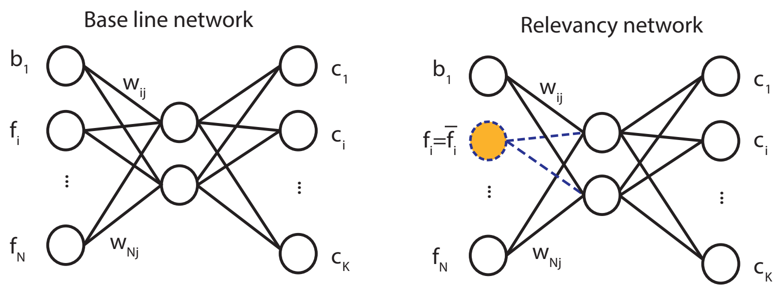

Figure 4.12: The architecture of the network with all features and the relevancy network

where P is the generalised performance of the NN using feature set F and P' is the generalised performance of the NN using feature set F where feature f_i values are substituted by mean value of f_i. Note that the value of P and P' is always between 0 and 1. The higher score of relevancy means the feature is more important. The score reflects how much effect if the feature is not used in the network. For example, $Rel_i = 0.7$ means that the absent of the feature f_i will lower the network's performance by 70%.

The relevance measurement only considers the relationship between a single input and the class. It does not consider the relationship between features i.e. redundancy and complementary. We enhance the Clamping method by introducing another measurement to measure complementary of the features to the already selected feature set. Also, unlike other techniques which consider redundancy measurement, MRMC considers feature complementary.

2. Complementary score

The complementary score measures how much the feature complements the already selected features set. It also takes feature redundancy into account such that if the feature is redundant to the already selected features, the score should be low as it does not bring additional information to the classification. Firstly, the base line performance is obtained by constructing a network using all selected features S and calculating its performance. Next, a new feature f_i is added to the network. This network is referred as the complementary network. The architecture and concept of the base line network and the network with new feature is shown in Figure 4.13.

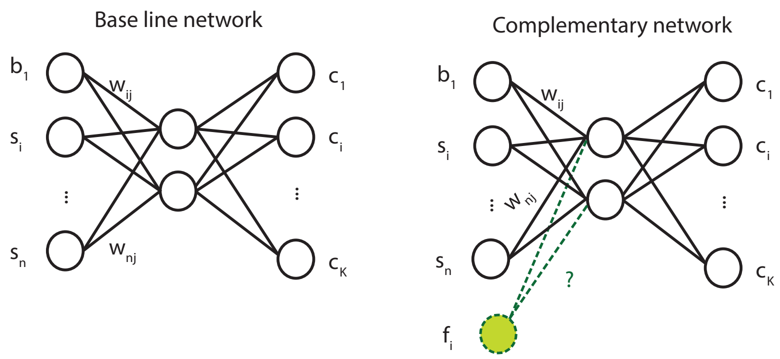

Figure 4.13: The architecture of the network with selected features and the complementary network

From Figure 4.13, it can be seen that the weights for the feature f_i needs to be obtained as they are not existed in the base line network. In our algorithm, we modify the construction of the complementary network such that it partly uses the weights and biases from the base line network. We assume that the baseline network has already identified the correct weights for the already selected features. Thus, by using the same weights and bias would help the network converges faster. This also reduces the possibility of the complementary network obtaining poor performance resulting from random initial weights. As the input and hidden nodes of the baseline network and the complementary network are different, the number of weights and biases are also different. The other weights and biases that are missing are generated randomly using the standard normal distribution with mean 0 variance 1 scaled by the number of input nodes for bias and weights in the first layer and the number of hidden nodes in the second layer.

Given an already selected feature set S, the complementary of feature f_i to S can be calculated as:

$$Com_{f_i} = \frac{P(S \cup f_i)}{P(S)} - 1 \tag{4.4}$$

where $P(S \cup f_i)$ is the generalised performance of the complementary network and $P(S)$ is the generalised performance of the baseline network. The values of P is always between 0 and 1. The complementary score reflects how much the new feature f_i contributes to the base line network. For example $Com_{f_i} = 0.1$ means by adding feature f_i, the performance of the network is improved by 10%.

3. Maximum relevance and maximum complementary score

   The proposed algorithm ranks features based on the maximum relevance and maximum complementary score. After the relevancy and complementary score are obtained, the relevance-complementary score (RC) can be calculated as:

$$RC_{f_i} = Rel_{f_i} + Com_{f_i} \tag{4.5}$$

   The feature is then selected based on the maximum RC score. From the algorithm, it can be seen that the complementary measurement can reduce the chance of selecting overlapping or redundant features. For example, given three features f_1, f_2, f_3 where $f_3 = f_1$ to represent overlapped feature and suppose their relevance scores are expressed as $f_1 = f_3 > f_2$. If Clamping technique is used, the feature ranking will be f_1, f_3, f_2. However, by combing the complementary with relevancy, the ranking will be f_1, f_2, f_3. As the complementary score of f_3 should be zero, the RC score for f_2 will then be higher than f_3.

The steps of MRMC algorithm are summarised in Figure 4.14 which are explained in detail below:

Step 1 : Normalize features value to [0 1] range. This step makes sure that features with larger values do not overwhelm features with smaller values. Set $S = \{\}$ and F contains all features.

Step 2 : Calculate the relevance score of all features f_i in F using (4.3). Note that the network is constructed using training data, then the generalised performance is calculated using validation data.

Step 3 : Select the first feature which has the maximum relevance score $f_s = \max_{f_i \in F} Rel(f_i)$.

Step 4 : Update S and F using equations (4.1) and (4.2).

Step 5 : Check if the size of feature set F is more than 1. If Yes, go to Step 6. Otherwise, update S using $S = S \cup F$. Terminate the algorithm.

Step 6 : Calculate the complementary score for all features f_i in F using (4.4).

Step 7 : Calculate the RC score using (4.5).

Step 8 : Select feature f_s which has the maximum RC score $f_s = \{\max_{f_i \in F} RC(f_i)\}$. Go back to Step 4.

4.3.4 Data sets

The experiments are carried out using two well-defined problems studied in [65] and four benchmark classification data sets including iris, breast cancer, cardiotocography, and chess which are obtained from the UCI Machine Learning Repository [64] available at http://archive.ics.uci.edu/ml. The proposed algorithm is also evaluated using a real world data set which we have collected from wearable sensors used for predicting human activities. All experiments except the first and second experiments are carried out using 5-fold cross-validation where 3 folds are used for training, 1 fold for validation and 1 fold for testing. The reason that we used 5 fold here is to reduce experimental time due to large data size. The size of the training, validation and testing data of each fold used for different data set are shown in Table 4.14.

4.3.5 Experiment setup

For the calculation of MI of MRMR and NMIFS, the input features are discretised using bin 10. For Clamping and MRMC, the number of hidden nodes is set to 2

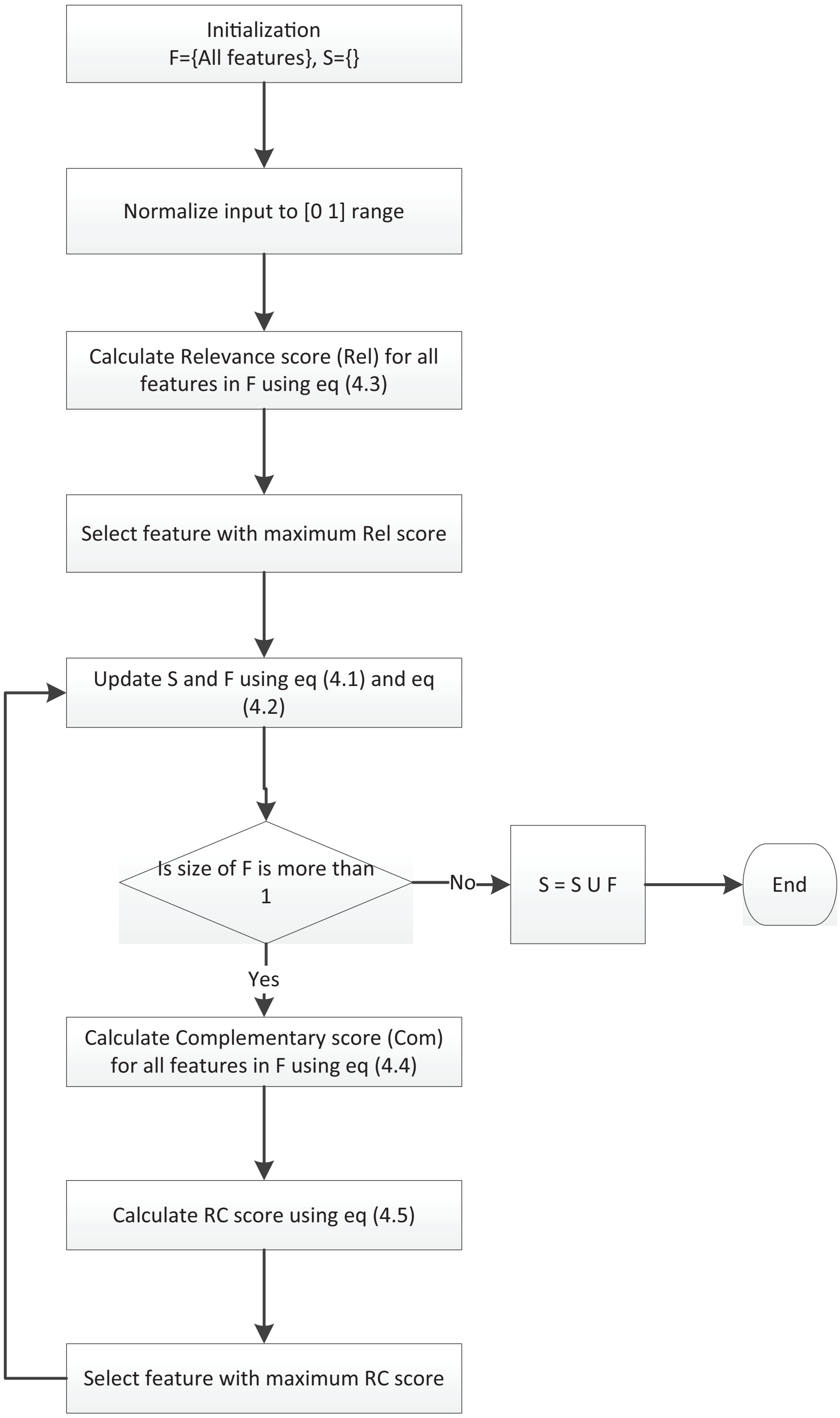

Figure 4.14: A flow chart of MRMC feature selection algorithm

Table 4.14: Characteristics and data partition per fold for different data sets used in the MRMC study

Data set	# Features	# Classes	Data type	# Sample	# Training	# Validation	# Testing
Nonlinear AND	14	2	Real	500	500	-	-
Nonlinear AND with partly overlapped features	17	2	Real	500	500	-	-
Iris	4	3	Real	150	90	30	30
Cancer-1992	9	2	Integer	699	288	96	96
Cancer-1995	30	2	Real	569	252	84	84
Cardiotocography-fetal	21	3	Real	2126	315	105	105
Cardiotocography-morp	21	10	Real	2126	300	100	100
Chess	36	2	Categorial	3196	1830	610	610
Wearable-sensor activity	141	12	Real	39328	15120	5040	5040

$\times$ number of input nodes and the number of epoch is 300 regardless the network converges or not.

For the real world problems, the feature selection methods are evaluated using NN. The number of hidden nodes is set to 2 $\times$ number of inputs and the number of epoch is set to 300. For each size of input, 10 models are constructed and the best one is selected using validation data. The test data is then applied to obtain the classification results. The validation data is also used to determine the size of features. The number of features is selected at the point where there is no significant improvement when more features are added. The performance of the four algorithms are compared using statistical tests at 95% confidence interval.

4.3.6 Experimental results

This section presents evaluation results of MRMC against other feature selection methods as presented in Section 4.2.3.2. The results are reported on each data set.

4.3.6.1 Experiment 1: Nonlinear AND problem

In the first experiment, a well-defined problem which the correct features are known is studied. A nonlinear AND problem which have been previously studied in [63, 65] is used. There are 14 features in this problem. The first five features f_1 to f_5 are generated randomly from an exponential distribution with mean 10. These features represent irrelevant features. The next six features f_6 to f_{11} are relevant features generated randomly from a uniform distribution range [-1 1]. The next three features f_{12} to f_{14} are redundant features (fully overlapped features) where f_{12}, f_{13}, f_{14} are identical to f_9, f_{10}, f_{11}, respectively. The class label is determined by:

$$f(x) = \begin{cases} C_1 & \text{If} \quad f_6 * f_7 * f_8 > 0 \quad \text{AND} \quad f_9 + f_{10} + f_{11} > 0 \\ C_2 & \text{If} \quad f_6 * f_7 * f_8 < 0 \quad \text{AND} \quad f_9 + f_{10} + f_{11} < 0 \end{cases} \tag{4.6}$$

According to this problem, the optimal feature set is $\{f_6, f_7, f_8, [f_9 \text{ or } f_{12}], [f_{10} \text{ or } f_{13}], [f_{11} \text{ or } f_{14}]\}$. The set of 500 data samples is generated randomly from a uniform distribution. The class label for each data sample is determined using equation (4.6). Feature selection algorithms which described in Section 4.2.3.2 and Section 4.3.3 are applied on the data set. For Clamping and MRMC which require validation data set, the 500 training data set is used. Table 4.15 presents the ranking results using these algorithms.

Table 4.15: Feature rankings using different feature selection methods (Nonlinear AND)

Algorithm	Feature rankings													
MRMR	f_{11}	f_9	f_1	f_2	f_4	f_3	f_{10}	f_5	f_6	f_8	f_7	f_{14}	f_{12}	f_{13}
NMIFS	f_{11}	f_9	f_{10}	f_6	f_8	f_7	f_3	f_4	f_2	f_1	f_5	f_{14}	f_{12}	f_{13}
Clamping	f_8	f_7	f_6	f_9	f_{11}	f_{12}	f_{14}	f_{10}	f_{13}	f_4	f_5	f_1	f_2	f_3
MRMC	f_8	f_7	f_6	f_9	f_{11}	f_{10}	f_{14}	f_{12}	f_{13}	f_4	f_5	f_1	f_3	f_2

From Table 4.15, it can be seen that only NMIFS and MRMC can identify the correct set of features. The first important feature ranked by MRMR and NMIFS is f_{11} and by Clamping and MRMC is f_8. This is expected as MRMR and NMIFS selects the first feature using maximum MI. Similarly, Clamping and MRMR use the same measurement to select the first feature. MRMR cannot detect the irrelevant feature where it ranks f_1 as the third important feature. Clamping correctly select the first five features, however it fails to detect that f_{12} is the redundancy of f_9 and f_{14} is the redundancy of f_{11}. According to this result, it can be seen that NMIFS gives the emphasis on detecting redundancy where it puts redundant features f_{12}, f_{13}, f_{14} at the end of the rank. On the contrary, MRMC gives emphasis on complementary where all irrelevant features are put at the end.

4.3.6.2 Experiment 2: A nonlinear AND problem with partly overlapped features

This experiment aims to show the superior ability of MRMC over the other three algorithms where it can select the correct features set from the data set which contains irrelevant, complete overlapped and partly overlapped features.

We use the same data set as generated in experiment 1. However, we introduce another three features f_{15} to f_{17} which will represent partly overlapped features. Feature f_{15} is set to $f_{15} = f_6 * f_7$ which overlaps the feature f_6 and f_7. Feature f_{16} is set to $f_{16} = f_9 + f_{10}$ which overlaps the feature f_9 and f_{10}. Feature f_{17} is set to $f_{17} = f_8 * f_{11}$ which overlaps the feature f_8 and f_{11} but has no relationship to the classes. From this example, it can be seen that the relevant features are f_6 to f_{16}. Feature f_{15} is the overlap of feature f_6 and f_7. However, it is better to select f_{15} and treat f_6 and f_7 as redundant as f_{15} contains information from f_6 and f_7, therefore by selecting f_{15} the feature space would be smaller. The same reason also applies for selecting f_{16} over f_9 and f_{10}. The optimal subset of this data set is $\{f_8, [f_{11} \text{ or } f_{14}], f_{15}, f_{16}\}$.

Table 4.16: Feature rankings using different feature selection methods (Modified nonlinear AND)

Algorithm	Feature rankings																
MRMR	f_{16}	f_{11}	f_1	f_2	f_4	f_5	f_3	f_{15}	f_9	f_8	f_{10}	f_6	f_7	f_{17}	f_{14}	f_{12}	f_{13}
NMIFS	f_{16}	f_{11}	f_6	f_9	f_8	f_7	f_{10}	f_3	f_4	f_2	f_1	f_{14}	f_5	f_{15}	f_{12}	f_{17}	f_{13}
Clamping	f_{15}	f_8	f_{14}	f_{11}	f_9	f_{12}	f_4	f_{10}	f_{13}	f_6	f_{16}	f_1	f_{17}	f_3	f_5	f_7	f_2
MRMC	f_8	f_{15}	f_{16}	f_{11}	f_{14}	f_6	f_{10}	f_4	f_{12}	f_9	f_{13}	f_1	f_7	f_3	f_{17}	f_5	f_2

The result from Table 4.16 shows that only MRMC can produce the correct feature set. Only two features f_{16}, f_{11}) are selected correctly by MRMR. The next five features selected by MRMR are irrelevant features. Clamping can select the first three features (f_{15}, f_8, f_{14}) correctly. However, the fourth feature (f_{11}) is the redundant of the third feature (f_{14}). This is because Clamping has no method of detecting overlap or redundant features. NMIFS can identify the first two features correctly. However, it selects f_6 and f_7 instead of f_{15} which makes the size of feature set larger. It also fails to detect that f_9 is the redundant feature of f_{16}.

4.3.6.3 Experiment 3: Iris data set

This data set has been widely used in classification literatures [60, 61]. The data set contains three classes of the type of Iris plant: Setosa, Versicolor, and Verginica. There are 50 samples per class. One class is linearly separable from the others. Two classes are not linearly separable. There are four features in this data set including sepal length (cm), sepal width (cm), petal length (cm), and petal width (cm). Different feature selection algorithms are applied on the data set and the mean classification accuracy of the test set is presented in Figure 4.15.

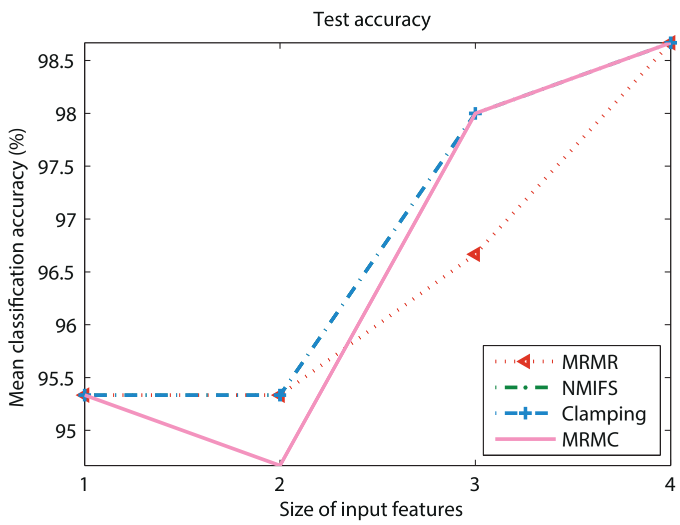

Figure 4.15: Mean classification accuracy of test data on different FS algorithms (Iris data set)

From Figure 4.17, all algorithms select the first feature correctly. MRMC does not correctly select the second feature in all folds and MRMR does not correctly select the third feature, therefore classification accuracy is slightly affected. The size of the feature set for each algorithms is determined using the validation data. The paired T-test is used to compare the accuracy between different size of features. The number of features is selected when no significant improvement is detected when adding more features. The size of features, test classification accuracy and standard deviation are shown in Table 4.17.

The performances of each feature selection techniques are compared empirically. First, the data is tested for normality using Shapiro-Wilk. The result

Table 4.17: Feature sets selected by different feature selection algorithms and mean test accuracy (Iris data set)

Algorithm	No. of features	Mean test accuracy (%)	Standard Deviation
MRMR	1	95.333333	1.8257419
NMIFS	1	95.333333	1.8257419
Clamping	1	95.333333	2.9814240
MRMC	1	95.333333	2.9814240

indicates that the variables are not normal distribution, therefore the Friedman Test is used. The results show that there is no statistical significance in classification accuracy between different feature selection algorithms (p=1.00).

4.3.6.4 Experiment 4: Wisconsin diagnostic breast cancer data set

This data set has been used extensively in previous works [58, 59]. The breast cancer data set is obtained from the University of Wisconsin Hospitals, Madison [57]. This data set is collected in 1992 and we shall refer this data set Cancer-1992. It contains 9 integer-valued features such as clump thickness, uniformity of cell size, uniformity of cell shape, bland chromatin, etc. The values for each feature is range between 1 and 10. There are 699 samples with 65.5% benign and 34.5% malignant cases. There are 16 samples which contain some missing values. For example, clump thickness value is missing in sample 1. The mean classification accuracy on test data are shown in Figure 4.16.

From Figure 4.16, the accuracy of Clamping and MRMC at the beginning are lower than MRMR and NMIFS. MRMR, NMIFS and MRMC reach similar accuracy when 3 features are used. Clamping reaches the highest accuracy when 5 features are used. The accuracies of MRMR and MRMC fluctuate slightly after 3 features. The number of features used for each algorithm is shown in Table 4.18.

Table 4.18: Feature sets selected by different feature selection algorithms and mean test accuracy (Breast cancer 1992 data set)

Algorithm	# Selected features	Mean test accuracy (%)	Standard Deviation
MRMR	3	96.666667	2.8905077
NMIFS	4	95.625000	2.0036858
Clamping	8	95.625000	1.1410887
MRMC	2	95.833333	1.6470196

Based on the mean test accuracy, the algorithms' performances can be ex-

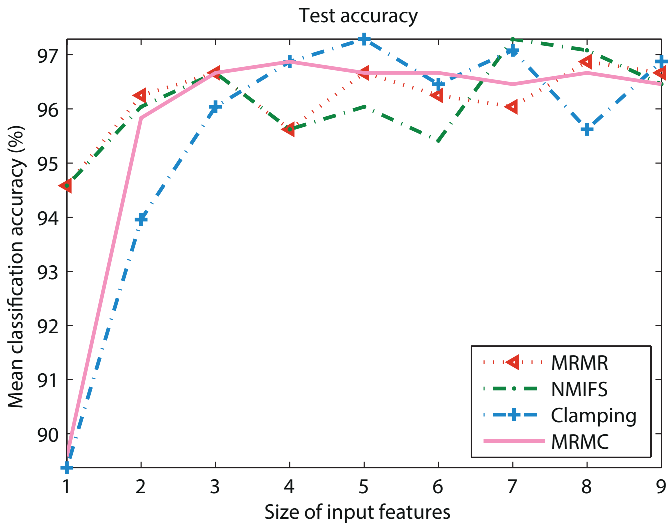

Figure 4.16: Mean classification accuracy of test data on different FS algorithms (Breast cancer data 1992 set)

pressed as Clamping<NMIFS<MRMC<MRMR. The performances of the four algorithms are compared using statistical tests. The data normality test shows that the data is not normal distribution therefore we use the Friedman test to detect the differences between algorithms. The result indicates that there is no statistical significance between four algorithms (Chi Square(3)=1.826, p=0.609). When we look at the number of features used in each algorithm, it can be seen that MRMC uses the smallest number of features. Hence, MRMC is the most optimum algorithm for this data set.

We also evaluate the proposed algorithm on another breast cancer data set which is collected in 1995. It is composed of 30 real-valued input features computed from a digitalised image of cell nucleus such as radius, texture, smoothness, mean, standard error, etc. to determine whether the cell is malignant or benign. The data set contains 357 benign and 212 malignant samples. In this study, a balanced sampling is used where an equal number of positive and negative classes are randomly selected using a uniform distribution. The size of training, validation, and testing data for each fold is shown in Table 4.14. The mean classification accuracy of the test data set for all four algorithms are shown in Fig 4.17.

From Figure 4.17, the first feature selected by Clamping and MRMC has lower

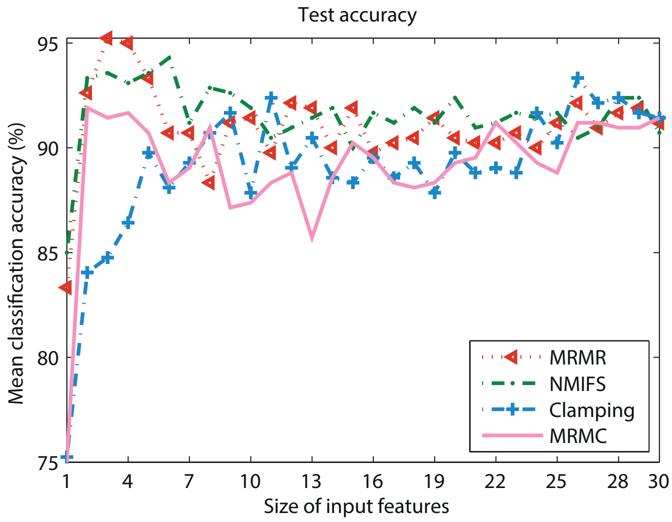

Figure 4.17: Mean classification accuracy of test data on different FS algorithms (Breast cancer 1995 data set)

accuracy then the feature selected by MRMR and NMIFS. However, using two selected features by MRMC, the accuracy significantly improves. MRMR and NMIFS provide similar performances on this data set.

The number of features for each algorithm is selected based on validation accuracy. The number is selected when there is no statistically significance when adding more features. The data normality is tested and appropriate test e.g. paired T-test or Wilcoxon Signed Ranks test is applied. The results are shown in Table 4.19.

Table 4.19: Feature sets selected by different feature selection algorithms and mean test accuracy (Breast cancer data 1995 set)

Algorithm	# Selected features	Mean test accuracy (%)	Standard Deviation
MRMR	4	95.000000	2.7147034
NMIFS	1	85.000000	13.0573044
Clamping	2	84.047619	8.7319623
MRMC	2	91.904762	2.7147034

The test accuracy for each algorithm are shown in Table 4.19. Based on the test accuracy, the algorithms' performances can be expressed as $Clamping < NMIFS < MRMC < MRMR$ where $A < B$ indicates that the algorithm B is better than the algorithm A. The normality test shows that the data have normal distribution. The within-subjects ANOVA is applied to compare the performance of each feature selection algorithms. Since the Mauchly result is significant, the Greenhouse-Geisser test is reported. The results indicate that there is no statistical significant between each algorithm($F(1.322, 5.288) = 2.273, p = 0.192$). From Table 4.19, it can be seen that NMIFS uses the lowest number of features. Therefore, it can be concluded that NMIFS is the optimum method on this data set.

4.3.6.5 Experiment 5: Cardiotocography data set

This data set has been used previously in [10]. It contains the measurement of fetal heart rate (FHR) and uterine contraction features e.g. minimum FHR histogram, percentage of time with abnormal long term variability, etc. on cardiotocograms classified by expert obstetricians. The data set contains 21 input features which can be classified into 10 types of morphologic patterns or 3 fetal states. The data set has unbalanced class distribution. In this study, the balanced sampling is used to obtain the equal number samples per class.

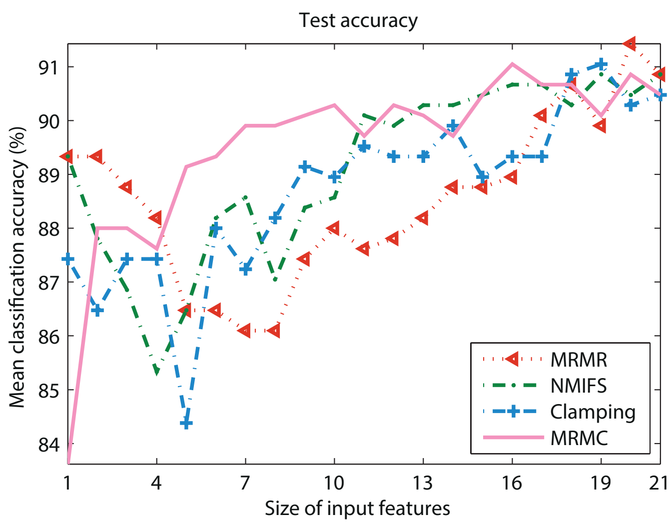

Figure 4.18: Mean classification accuracy of test data on different FS algorithms (Cardiotocography-Fetal data set)

The average classification accuracy of 3-class fetal states and 10-class morphologic patterns are shown in Figure 4.18 and Figure 4.19, respectively. From Figure

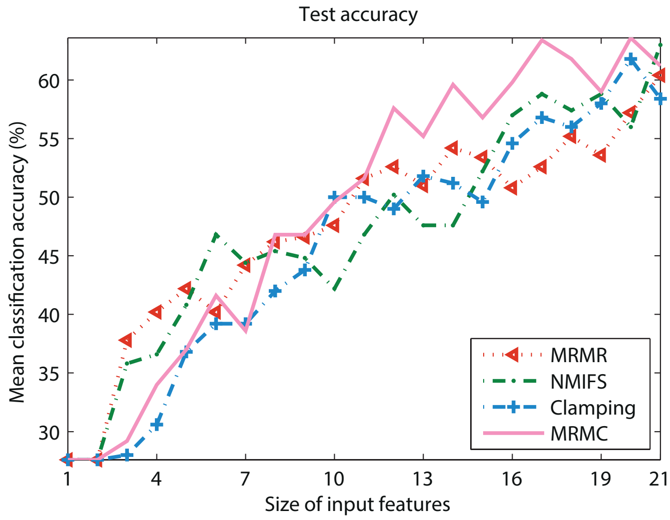

Figure 4.19: Mean classification accuracy of test data on different FS algorithms (Cardiotocography-Morp data set)

4.18, the classification accuracy of MRMC starts at the lowest but continues improving as the number of features is increasing. The classification accuracy of MRMR and NMIFS are high when using one feature. The classification accuracy of NMIFS falls to the lowest point when 5 features are used. The performance of MRMC is better than the other 3 algorithms when 6 to 10 features are used. From Figure 4.19, all feature selection algorithms produce similar accuracy trend. The classification accuracy improves when more features are used. The performance of MRMC is superior to the other 3 algorithms when 12 to 17 features are used.

Table 4.20: Feature sets selected by different feature selection algorithms and mean test accuracy (Cardiotocography-Fetal data set)

Algorithm	# Selected features	Mean test accuracy (%)	Standard Deviation
MRMR	18	90.666667	1.7036708
NMIFS	15	90.476190	2.0203051
Clamping	21	90.476190	1.5058465
MRMC	4	87.619048	3.5634832

Table 4.20 shows the number of features selected by each algorithm, the mean classification accuracy on test data and the standard deviation on Cardiotocography data set for classifying 3 fetal states. Based on the test classification

accuracy, the performance of each algorithms can be expressed as $MRMC < Clamping = NMIFS < MRMR$. The normality test shows the data is normal distributed. The ANOVA test is applied to test the null hypothesis that classification accuracy for all algorithms is the same. Since the Mauchly result is significant, the Greenhouse-Geisser test is reported. The results indicate no statistical significance between accuracy obtained by different feature selection algorithms ($F(1.045, 4.18) = 6.711, p = 0.058$). From Table 4.20, it can be seen that MRMC selects the lowest number of features. Therefore, it can be concluded that MRMC is the optimum method on this data set.

Table 4.21: Feature sets selected by different feature selection algorithms and mean test accuracy (Cardiotocography-Morp data set)

Algorithm	# Selected features	Mean test accuracy (%)	Standard Deviation
MRMR	21	91.428571	1.5058465
NMIFS	16	90.666667	1.2417528
Clamping	21	87.428571	4.1184282
MRMC	15	83.619048	6.8146834

Table 4.21 shows the results of different feature selection methods on classifying 10 morphologic patterns of cardiotocography data set. Based on the mean classification accuracy on the test data, the performance of the algorithms can be expressed as $MRMC < NMIFS < Clamping < MRMR$. The Shapiro-Wilk is applied to test data normality. The result shows that the data is normal distribution. The ANOVA is applied to test the performance of different feature selection algorithms. The results show that there is no statistical significant in accuracy between four algorithms at the 5% level ($F(3, 12) = 0.278, p = 0.840$). Among the four algorithms, it can be seen that MRMC uses the lowest number of features. Therefore, it can be concluded that MRMC is the optimum feature selection method for this data set.

4.3.6.6 Experiment 6: Chess data set

The chess data set contains sequences of chess-description for chess end game. This data set has been previously used in [55, 56]. The data set consists of 36 categorical-input features to classify if the white can win or cannot win. The class distribution is 52% win and 48% cannot win. Equal class distribution is used and the number of training, validation, and testing data are shown in Table 4.14. The data set uses a string to represent the board-description e.g. f, l, n, etc. which we convert these into integer values e.g. f=1, l=2, n=3, etc. The mean classification accuracy of the test data set is shown in Figure 4.20. The classification result of each algorithm using the number of features determined by validation data is presented in Table 4.22.

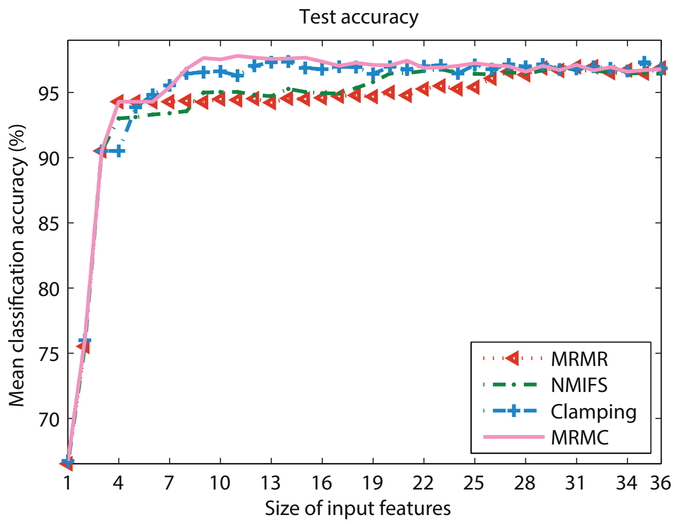

Figure 4.20: Mean classification accuracy of test data on different FS algorithms (Chess data set)

From Figure 4.20, the performances of all algorithms are increased when using more number of features. When we observe the feature selected by each algorithm, it is found that the first three features selected are the same. Generally, the performances of Clamping and MRMC are better than MRMR and NMIFS in this data set. MRMC performance is better than Clamping when 8 to 21 features are used. All algorithms reach similar accuracy when 29 and more features are used.

Based on the mean classification accuracy of test data, the algorithms' performances can be expressed as $NMIFS < MRMR < Clamping < MRMC$. We first test the data normality of test accuracy using different number of features for different algorithms as shown in Table 4.22. The normality test indicates that the data is not normal distribution. Therefore, the Friedman test is used to test

Table 4.22: Feature sets selected by different feature selection algorithms and mean test accuracy (Chess data set)

Algorithm	# Selected features	Mean test accuracy (%)	Standard Deviation
MRMR	27	96.590164	1.2234825
NMIFS	20	96.459016	1.2835134
Clamping	14	97.377049	0.9628967
MRMC	10	97.540984	0.8439041

the different in performances between four algorithms. The result reveals that there is no statistical significant difference between the algorithms ($p = 0.054$). Based on the number of features used in each algorithm, MRMC uses the lowest number while MRMR uses the highest number of features. Therefore, we can conclude that MRMC is the most optimum algorithm for this data set.

4.3.6.7 Experiment 7: Wearable-sensor activity data set

We collected raw sensor data of accelerometer, gyroscope, heart rate monitor, light, temperature, altimeter, and barometer from 12 elderly people performing 12 activities of daily livings including walking, feeding, exercising, reading, watching TV, washing dishes, sleeping, ironing, feeding, scrubbing, wiping, and brushing teeth. The participants wore the sensors one their wrists and heart rate monitor on their chests. The data set consists of 141 real-valued input features. The classification accuracies of the test data set for all algorithms are shown in Figure 4.21. The size of the feature set of each algorithm are determined using validation data and the results are presented in Table 4.23.

From Figure 4.21, MRMC accuracy is better than the other algorithms when four or more features are used. Differences in accuracies are noticeable when 10 and 34 features are used. The accuracy of Clamping is lower than other algorithms when few features are used. However, it achieves comparable accuracy as MRMR and NMIFS when more than 14 features are used. MRMR and NMIFS achieve the same accuracy when few features are used. However, NMIFS performance drops when 3 and 25 features are used.

Based on the mean classification accuracy on test data, the algorithms' performances can be expressed as $MRMR < Clamping < NMIFS < MRMC$. The algorithms' performances are compared statistically. The data normality

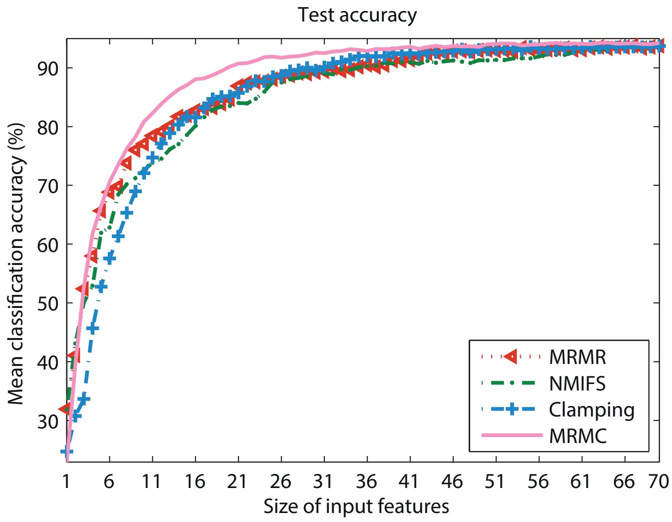

Figure 4.21: Mean classification accuracy of test data on different FS algorithms (Wearable-sensor activity data set)

Table 4.23: Feature sets selected by different feature selection algorithms and mean test accuracy (Wearable-sensor activity data set)

Algorithm	# Selected features	Mean test accuracy (%)	Standard Deviation
MRMR	64	93.313500	0.4858261
NMIFS	66	93.662700	0.5337766
Clamping	62	93.611120	0.8777444
MRMC	50	94.027800	0.6026319

test reveals that the data is normal, thus ANOVA test is applied. The Sphericity test is significant therefore the Greenhouse-Geisser test is reported. The results indicate that there is no significant difference between the four algorithms ($F(1.474, 5.895) = 1.417, p = 0.301$). Based on the number of features used in each algorithm, it can be seen that MRMC only uses 50 features while the other use over 60 features. Therefore, we can conclude that MRMC is the most optimum algorithm for this data set.

4.3.7 Discussion

The summary of the experiments is presented in Table 4.24. The optimal feature selection algorithms of each data set is based on the statistical results and the number of features. Based on 8 experiments, MRMC is the optimum feature selection algorithm in general. It is able to obtain high classification result using the minimum number of features. NMIFS is the next best feature selection algorithm.

Table 4.24: Optimum feature selection algorithm on each data set

Data set	Optimum feature selection algorithm
Nonlinear AND	NMIFS, MRMC
Nonlinear AND with partly overlapped features	MRMC
Iris	MRMR, NMIFS, Clamping, MRMC
Cancer-1992	MRMC
Cancer-1995	NMIFS
Cardiotocography-fetal	MRMC
Cardiotocography-morp	MRMC
Chess	MRMC
Multi-sensor AR	MRMC

The results from experiments 1 and 2 show that MRMC is capable of detecting completely overlapped and partial overlapped features. In other experiments, the result also shows that MRMC can be used on various data type i.e. categorical, real, and integer values. The performance of MRMC is not as good as NMIFS in breast cancer-1995 data set. This is due to the fact that the first feature selected by MRMC normally results in a low classification accuracy. The differences in accuracy between NMIFS and MRMC are about 10% when one feature is selected. When looking at other data sets, the differences in accuracies are about 5% or less. This implies that when using one feature, if an algorithm obtains a significantly higher accuracy than MRMC then that algorithm would be more optimum for that data set, provided that the number of input features is small.

For the cardiotocography data set, MRMC shows that it is the optimum algorithm among the four algorithms. It achieves good accuracy while using the smallest number of features. Experiment 6 demonstrates that MRMC also works well with categorical data. In experiment 7, the proposed algorithm is evaluated with the data set with a large number of inputs. The result shows that MRMC is much superior. In general, it uses fewer than 10 features comparing to other algorithms while achieving the highest accuracy. When comparing MRMC with Clamping, it can be seen that by introducing a complementary measurement, the performance of the algorithm is better. For example, in breast cancer 1995 data set, using the same number of features, MRMC can obtain higher accuracy.

From this study, it can be seen that using Clamping to detect the most important feature may not give the correct result. This affects the performance of MRMC as it uses the same criteria to select the first feature. As forward search is used, the performance of the feature selection algorithm depends on the first selected feature. Therefore, in case of the feature selection of a small set of features, using MRMC may not guarantee good results. However, when the number of features is increased, MRMC is demonstrated to be superior to the other three algorithms. This is due to the fact that although the first feature selected by Clamping algorithm may not always be the most important but it is somewhat important i.e. the second or third most important feature, and by using complementary measurement, the correct subset of features can later be identified.

4.3.8 Conclusion remarks

In this study, a new feature selection algorithm based on Maximum Relevance Maximum Complementary using NN has been proposed. The proposed methods are evaluated on well-defined problems and real world data sets containing small to larger set of features (N=4 to 100+). The study is carried out using 5-fold cross validation. The algorithms performances are evaluated empirically using statistical tests at 95% confidence interval. The results show that in general

MRMC provides a good performance comparing to the other three algorithms. Also, the complementary measurement introduced improves the performance of Clamping algorithm. The study indicates that for the problem with small set of features, the performance of MRMC is affected by the selection of the first feature. Future research will be focusing on the identification of the first feature in order to improve the performance of the algorithm. Also, sometimes there is more than one important features with equal scores, in order to correctly identify the feature, the next important feature needs to be considered.

4.4 Sensor contribution study

4.4.1 Study hypothesis and objectives

In this research, multiple sensors are used for AR. The aim of this study is to understand the importance of different sensor in AR model. Two techniques i.e. MI and Clamping are used to analyse the features generated from the sensors. MI is used to measure the importance of each feature to the activity classification while Clamping is used for measure the importance of each feature within the model to the activity classification. Based from literatures, it is hypothesised that accelerometer is the most important sensor for recognising the interested activities. This is because these activities are mainly use movement on the wrist and acceleroemeter is capable for capturing movement information. Also, it is hypothesised that specific sensor i.e. light will be important for specific activity with different lighting condition i.e. sleeping.

4.4.2 Methodology

We use two techniques to investigate the importance of a particular sensor i.e. MI and Clamping [18].

1. Mutual information (MI)

   MI is based on information theory. It is used for defining the dependency between variables. Given two variables, x, y, the MI can be calculated as [15]:

$$I(x;y) = \int\int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \mathrm{d}x\mathrm{d}y$$

2. Clamping

   MLP is constructed using several sensors based on the feature selection process. Features of each sensor are substituted using their mean values. If the sensor is important in the network, removing it would result in lower network performance. Assuming all features within a sensor give equal significance, the contribution of a particular sensor could be calculated as:

$$con_{(S)} = 1 - \frac{g(F|S=\bar{S})}{g(F)}$$

   where F is a set of features, S is the set of features of a particular sensor, $g(F|S=\bar{S})$ is the performance of the network where S is substituted of with their mean values, and $g(F)$ is the generalised performance of the network.

4.4.3 Experiment design and data set

The study of sensor contribution is separated into two experiments. The first experiment is carried out to understand the importance of a sensor and feature to the activity classification using MI and to understand the importance of the sensor within the model using Clamping. The second experiment is carried out to understand the effect of the absent of a sensor within the model using Clamping. To study the contribution of all sensors, the Wearable-sensor data set is used in this study.

4.4.4 Experimental results and discussions

4.4.4.1 Experiment 1: Sensor contribution using MI and Clamping

For each feature, MI between feature and class is calculated. Figure 4.22 and the Shapiro-Wilk tests reveal that MI is not normal distributed ($P \leq 0.05$). Thus, it is appropriate to analyse the data using non-parametric statistics e.g. median, quartile, etc.

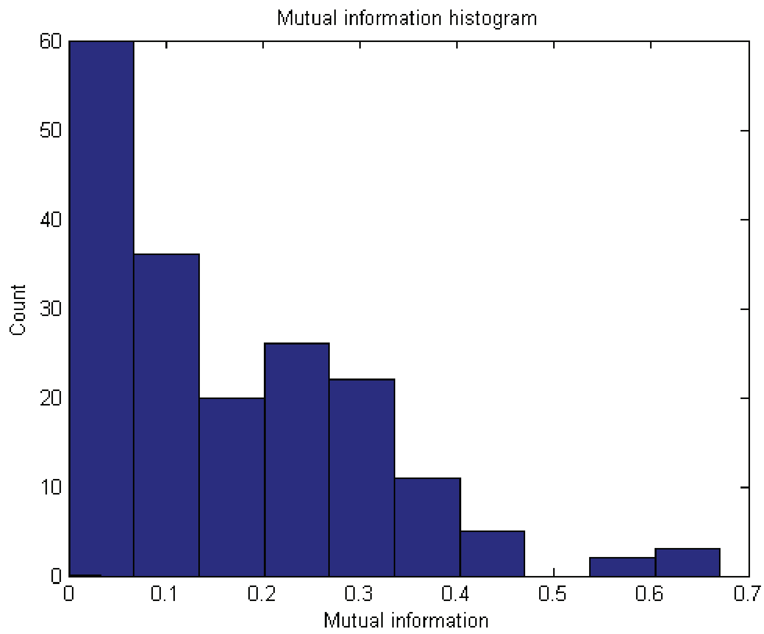

Figure 4.22: A histogram of MI of features

Figure 4.23 to Figure 4.29 show the plot of MI of each feature. The plot of MI of each feature shows that accelerometer sensor contains the most information about the activities. 33.96% of accelerometer features have more than third quartile of MI. Altimeter and temperature sensors are the least important sensors. The result shows that accelerometer, gyroscope, barometer and light are among the most important sensors containing useful information in classifying 12 activities. Accelerometer and gyroscopes produce the top ten MI (See Table 4.25). MI of some of the features calculated from these sensors are in the 3^{rd} quartile or higher (See Table 4.26). Also, it can be seen in Table 4.27 that the time domain features provide more useful information than the frequency domain features. Maximum, RMS, mean, median, STD, mode, minimum, intensity are the most important features, respectively.

Table 4.25: Top ten features

Source	Feature
Acceleration Y axis	Max, mean, median, min, mode, RMS
Norm gyro	RMS, mean
Acceleration Z axis	Min, mode

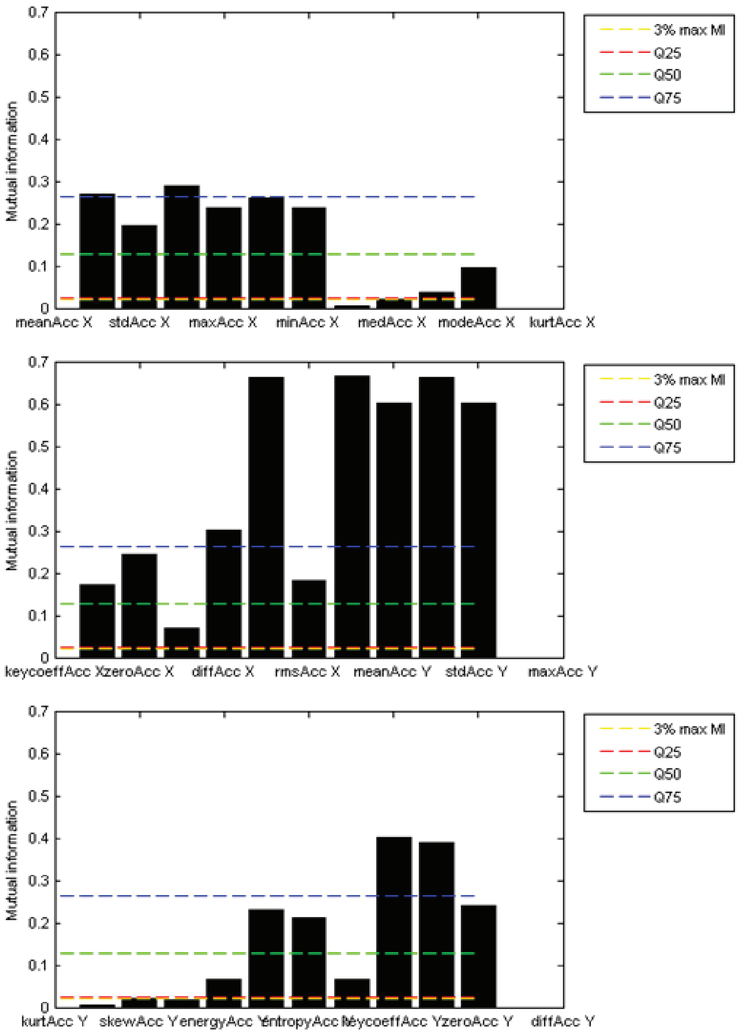

Figure 4.23: A histogram of MI of features (cont.)

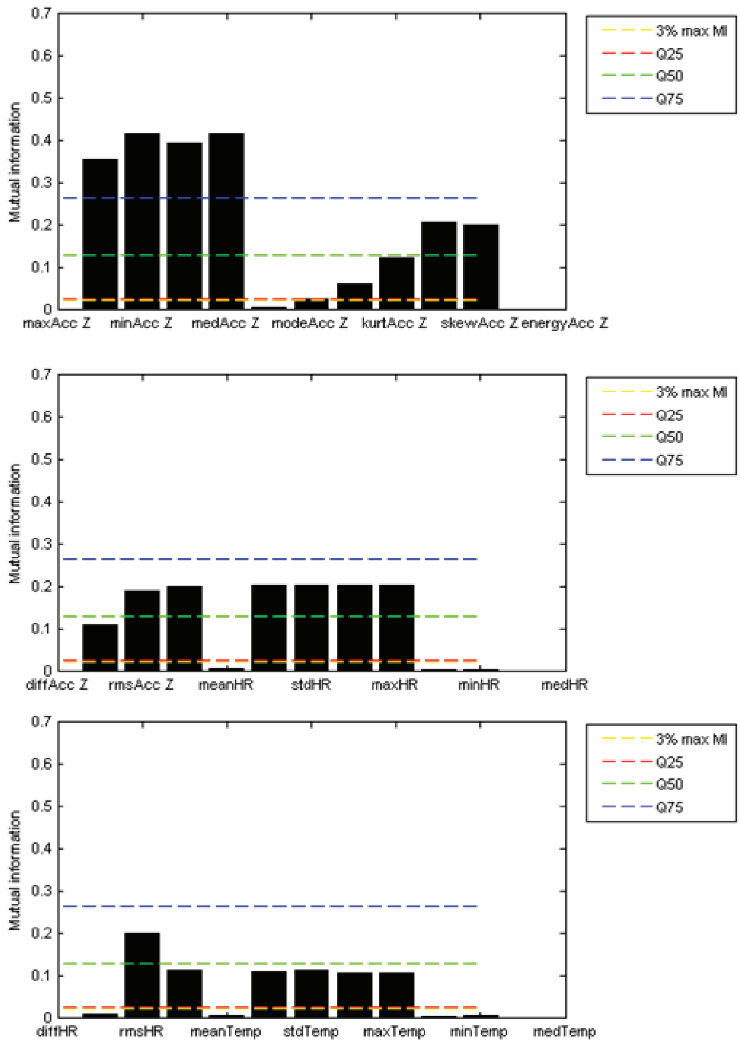

Figure 4.24: A histogram of MI of features (cont.)

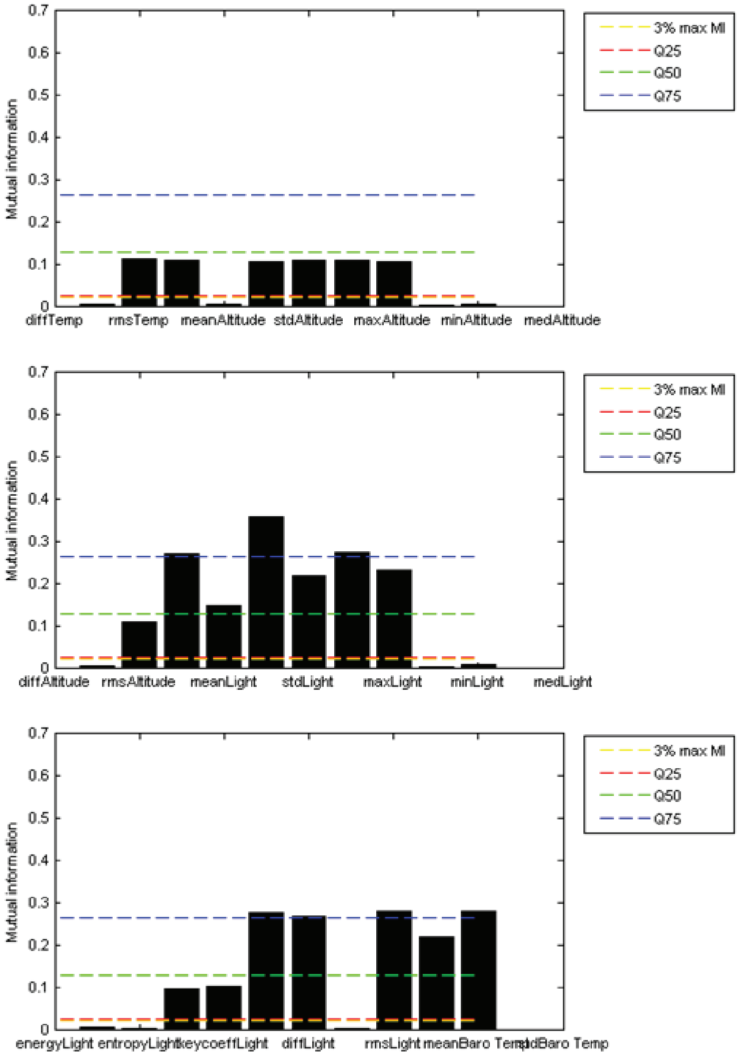

Figure 4.25: A histogram of MI of features (cont.)

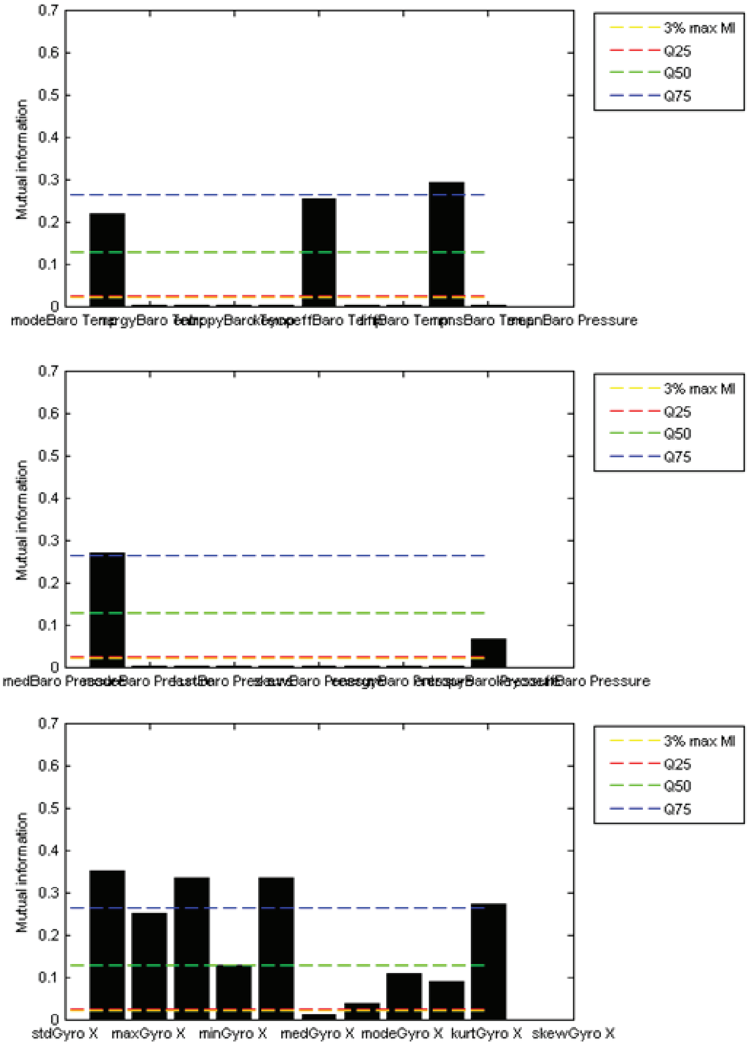

Figure 4.26: A histogram of MI of features (cont.)

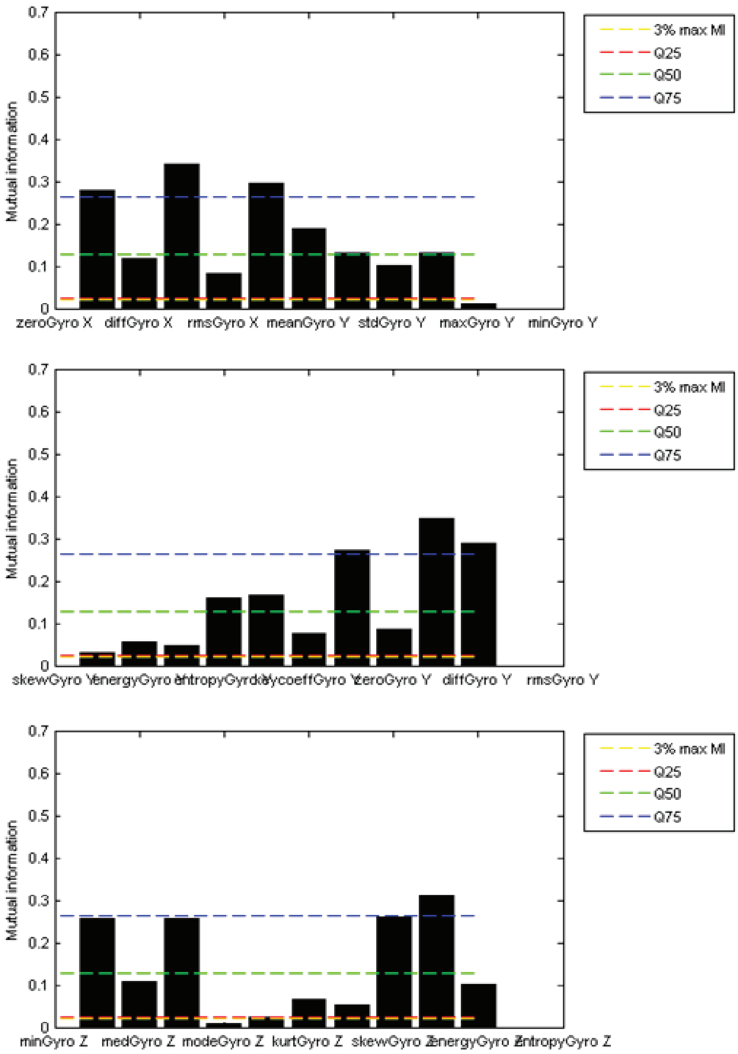

Figure 4.27: A histogram of MI of features (cont.)

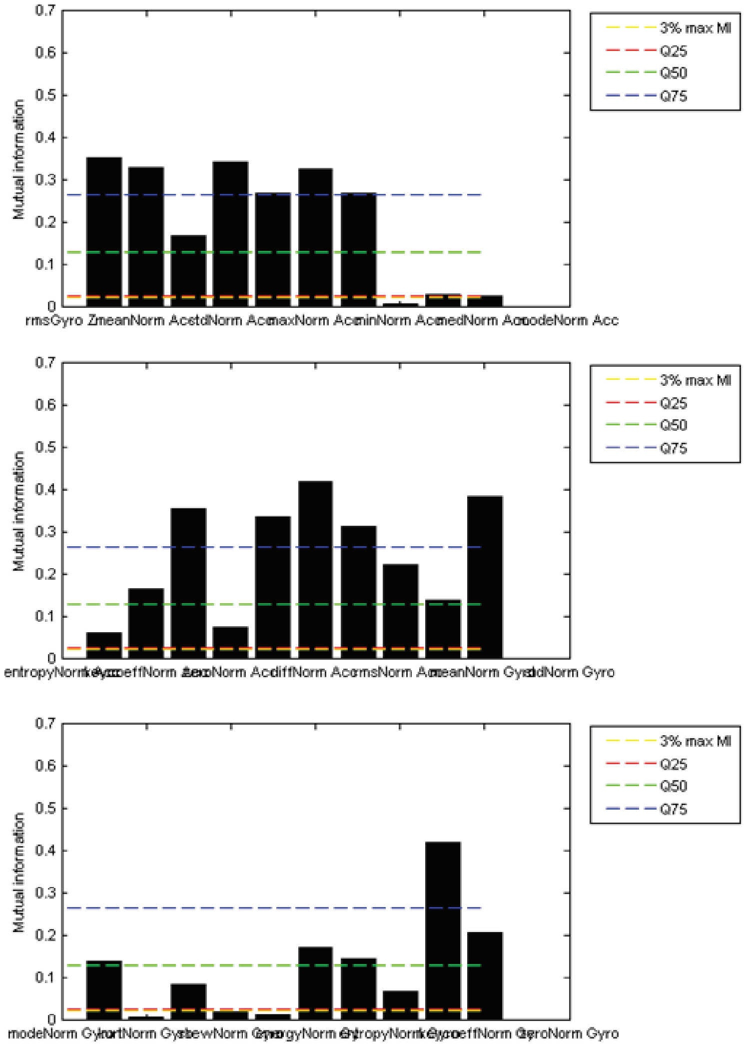

Figure 4.28: A histogram of MI of features (cont.)

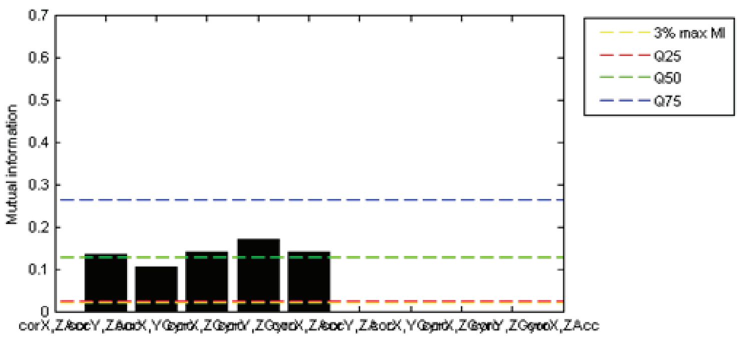

Figure 4.29: A histogram of MI of features (cont.)

Table 4.26: The percentage of features in different quartile of MI

Sensor	% > 0.75 Quartile	%0.5-0.75 Quartile	%0.25-0.5 Quartile	% < 0.25 Quartile
Accelerometer	33.96	24.53	16.98	24.53
Gyroscope	26.42	13.21	30.19	30.19
Barometer	25	75	0	0
Light	11.11	55.56	11.11	22.22
Heart rate	0	83.33	16.67	0
Altimeter	0	0	66.67	33.33
Temperature	0	0	66.67	33.33

Table 4.27: Features with MI over 75% Quartile

Sensor	Data	Feature
Accelerometer	X - axis	Max, RMS
	Y - axis	Mean, max, min, median, mode, RMS
	Z - axis	Mean, max, min, median, mode
	$\sqrt{X^2+Y^2+Z^2}$	Mean, max, median, intensity, RMS
Light	Light intensity	Max
Barometer	Temperature	Max
	Pressure	Max
Gyroscope	X - axis	STD, min, mode, intensity, RMS
	Y - axis	STD
	Z - axis	STD, max, intensity, RMS
	$\sqrt{X^2+Y^2+Z^2}$	Mean, STD, median, RMS

The feature selection is performed using FC. The truncation at 24 features is selected as the accuracy started to remain constant. Features from accelerometer, altimeter, heart rate monitor, light and barometer are selected. Also, 16 features are used to conform to previous study. Next, the contributions of sensors in our model (with 24 features) are investigated. The result shows that accelerometer is the most important sensor in the model (See Figure 4.30). This is followed by altimeter, heart rate monitor (HR), barometer, gyroscope, and light respectively. The top three features with the highest importance in the model are mean acceleration on Z-axis, maximum barometer pressure, and minimum altitude, respectively.

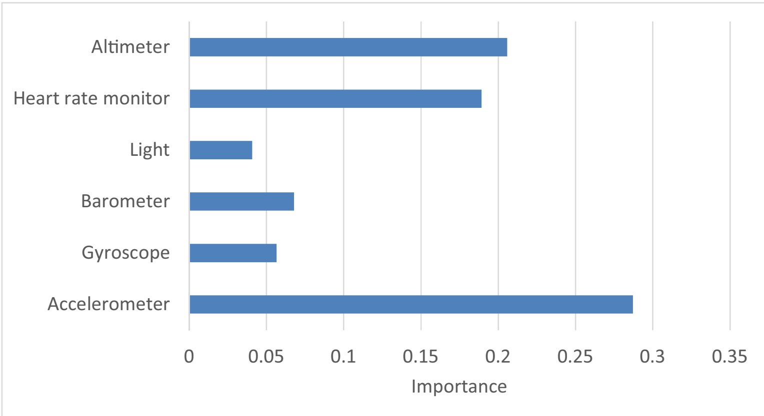

Figure 4.30: The contribution of each sensor in AR model with 24 features.

4.4.4.2 Experiment 1: Discussion

The result of the study indicates that accelerometer is the most important sensor for AR. This confirms that accelerometer has ability of measuring human activity quantitatively, fast reaction to changes in movement and reflects type of activity well [150]. It is found that the new sensors introduced including gyroscope, barometer and light contain useful information about human activities. Similar to accelerometer, gyroscope can reflect changes in activity well. We also observe that data obtained from gyroscope are similar to those from accelerometer. Barometer and light can be used to differentiate activities such as using stairs and sleeping.

Interestingly, although gyroscope, barometer and light are shown to be very important sensors on their own, this is not the case when they are combined together. In our model of 24 feature selected using FC, only 2 gyroscope features are selected. Also, its contribution to the network is not as high as other sensors. This may be explained that although gyroscope is a good sensor on its own, when it is used with accelerometer, many of its features become redundant. This is possibly due to the feature calculated from gyroscope data are similar to accelerometer data. It is suggested that other features such as angle, roll, and pitch should be calculated. The result also indicates that heart rate has significant contribution to the model. Using heart rate in the model increases the accuracy by 1.75%. The statistical tests show that the improvement is significant (p<0.05). This may be due to the fact that majority of activities studied in [143] are exercise related activities e.g. cycling, running, rowing, etc. Although, the authors reported that heart rate help improve exercise activities, due to the similarity in these activities and large number of classes, the overall improvement is not as high as they expected. On the other hand, our study contains activities which are rather different e.g. walking, sleeping, exercise, large difference in heart rate between these activities are expected and thus resulting in heart rate having a significant impact in our model.

4.4.4.3 Experiment 2: Sensor contribution within the model using the absent of a sensor concept

In this section, we study how each sensor within the model helps with classification. We perform experiments to understand how the loss of a particular sensor affects the classification accuracy and to which activity. To control the experiment, top features (based on MI) of each sensor are selected to use in the classification. The selected features are maximum acceleration Y-axis, maximum heart rate, maximum barometric pressure, maximum light intensity, RMS gyro magnitude, minimum temperature, and minimum altitude.

Firstly, we generated a classification model (called base model) which uses all sensors. We constructed the next model by removing one sensor. For example, model 1 used all sensor except accelerometer. Model 2 used all sensor except heart rate sensor. In total, 8 models are built. The notation of the model is given by M followed by the name of the removed sensor e.g. M_{Acc} represents model which does not use accelerometer. The classification is performed using MLP and the number of hidden nodes is twice the number of input. Table 4.28 shows mean accuracy of the model when a particular sensor is not used. The test of normality indicates that model M_{Light} is not normal distribution, thus we employ the Wilcoxon Signed Ranks to test the effect of the loss of a sensor. The statistical results indicate that there is a statistical significant different between the base model and all the other models (p<0.05). Based on the reduced accuracy, the contribution of the sensor can be ranked from the highest to the lowest as accelerometer, gyroscope, light sensor, barometer, heart rate sensor, temperature sensor, and altimeter, respectively. We examine the F-score of each class of each model (See Table 4.29). The model which does not include accelerometer has an effect on several activities including brushing teeth, feeding, ironing, reading, scrubbing, walking, and wiping. The effect on the absent of light sensor is on sleeping, stairs, and washing dishes activities. The model without a gyroscope sensor has effects on exercise and watching TV activity.

Table 4.28: The effect of the loss of a particular sensor

Model	Missing sensor	Accuracy (%)	Std. Deviation
Base model	None	65.1913	1.4354
M_{Acc}	Accelerometer	50.0933	1.4140
M_{HR}	Heart rate sensor	62.0873	1.2548
M_{Baro}	Barometer	60.7004	1.2010
M_{Light}	Light sensor	57.6663	1.1589
M_{Gyro}	Gyroscope	55.8540	1.4780
M_{Temp}	Temperature sensor	62.2528	1.1885
M_{Alt}	Altimeter	62.8056	1.1016

Table 4.29: F-score of models developed for sensor contribution study.

Model	Brush	Exercise	Feed	Iron	Read	Scrub	Sleep	Stairs	Walk	Wash	Watch	Wipe
Base model	0.6771	0.5818	0.5506	0.5856	0.5549	0.7140	0.7382	0.7144	0.7809	0.5191	0.7088	0.6683
M_{Acc}	0.5036	0.4438	0.4239	0.3715	0.4271	0.5025	0.6579	0.6437	0.3858	0.4307	0.6048	0.5325
M_{HR}	0.6493	0.5382	0.5393	0.5797	0.5122	0.6826	0.6995	0.6995	0.7725	0.4459	0.6652	0.6229
M_{Baro}	0.6406	0.5494	0.5397	0.5483	0.4824	0.6771	0.6456	0.6639	0.7596	0.4715	0.6500	0.6211
M_{Light}	0.5688	0.5639	0.4673	0.5640	0.5062	0.6843	0.5994	0.5354	0.7428	0.3973	0.6193	0.6035
M_{Gyro}	0.5995	0.3807	0.4841	0.5147	0.4879	0.5878	0.6676	0.6402	0.7286	0.4489	0.4838	0.6304
M_{Temp}	0.6544	0.5410	0.5405	0.5644	0.5197	0.6968	0.7204	0.6885	0.7541	0.4542	0.6816	0.6155
M_{Alt}	0.6583	0.5541	0.5417	0.5624	0.5157	0.7094	0.7033	0.6885	0.7645	0.4883	0.6814	0.6359

4.4.4.4 Experiment 2: Discussion

In this study we develop several models to investigate the absent of a particular sensor. It is found that each sensor has a significant contribution toward the classification accuracy in general. This means that each sensor has given specific information which is useful for activity classification. The results also show that accelerometer is the most important sensor since the classification accuracy has significantly dropped when the sensor is not used. However, missing this sensor does not strongly affect the detection of sleeping. This is due to the fact that this activity is not involved in much movement. On the other hand, missing the light sensor has significantly affected sleeping detection. This suggests the model uses information from the light sensor to detect sleeping activity. Similarly, stairs activity is also affected by missing light intensity information. When observing the plot of the maximum light intensity of these two classes, it is found that, unlike other classes, the light intensity data from sleeping and stairs activities are rather clustered. Therefore, missing the light sensor affects the classification of these two classes. The absent of gyroscope has effects on exercise and watching TV activities. This shows that although the M_{Gyro} model contains accelerometer feature, it is not enough to detect these activities. RMS of gyro magnitude significantly helps classify exercise and watching TV activities. Although the results demonstrate that each of the seven sensors are important, these models are constructed based on only one feature from each sensor. It is possible that when a model is developed with more features, information from a particular sensor could be substituted by the other features from other sensor as well. In fact, in the proposed model, temperature sensors are not selected.

4.4.5 Conclusion remarks

In general, accelerometer is the most important sensor. It is found that the new added sensors (gyroscope, barometer, light and heart rate monitor) provide valuable information for AR. It is found that gyroscope and accelerometer exhibit similar data and some are overlapped. Heart rate data can be useful when classifying activities which have diversity in heart rate data and may not be useful if contain activities which exhibit similar heart rate e.g. similar exercise activities. We also find that maximum light intensity can be useful for detecting sleeping, stairs, washing dishes activities. The RMS of gyro magnitude can help in classifying exercise and watching TV activities. Although we find that all the sensors provide important information toward classification, when larger features of sensors are available, a particular sensor could be omitted. The results also show that combining several sensor data improves classification accuracy.

4.5 Summary

This chapter presents extensive experiment results on features and feature selection. First, the feasibility of using wrist worn sensor for AR is investigated. It is found that activities can be recognised using only an accelerometer worn on wrist, however only basic activities such as walking, running, sitting, standing, etc. can be detected. Further investigation is carried out where multiple sensors are used. One of the most important tasks in AR using multi-sensor is to select the optimal set of features from a large feature space. In this chapter, two feature selection techniques are proposed. First, FC is proposed which combines Clamping technique and modified forward selection. The experimental results indicate that FC can select a better set of features comparing to other well-known feature selection algorithms i.e. MRMR, NMIFS, and Clamping. However, FC has two limitations due to the use of forward selection. Firstly, redundant features may be selected in the early round of selection and secondly, good features may be eliminated in the early selection stage. Another feature selection technique called MRMC is proposed which uses the concept of feature complementary. The experimental results indicate that MRMC provides comparable results with MRMR, NMIFS, and Clamping when applying on data sets with small numbers of features. The results also show that MRMC outperforms other algorithms when applying on data sets with large numbers of features. It is found that MRMC performance is affected by the selection of the first feature. Future research can be carried out to identify the method to select the first feature in order to improve MRMC performance. In this chapter, the comparison study on FC and MRMC has not been carried out due to limited time strain, which may be of interest for future study. Finally, the chapter presents a study on sensor contribution to AR performance. The experimental results indicate that accelerometer is the most important sensor. The missing of accelerometer has a strong impact on AR accuracy. Also, in general, it is found that multiple sensors contribute to the increased classification accuracy. However, if a large set of features is available, some sensors may be omitted. For future research, investigations on the proposed techniques can be carried on other data sets, and compared with other feature selection techniques.

Chapter 5

Activity classification and classifier combination study

In previous chapter, two feature selection techniques have been proposed and evaluated. The next step is to identify the suitable classification method for AR. In this chapter, experimental studies on the classification algorithms for multi-sensor AR are presented. The chapter also addresses the challenges in classifier combination by proposing two classifier combination techniques i.e. classifier fusion weights using GA and classifier combination model using GA. The proposed techniques are tested and validated using the extensive experimental studies. Some parts of this chapter have been published in [1, 4, 8].

5.1 Activity classification algorithms study

5.1.1 Study hypothesis and objectives

After an optimum feature set is identified, the activity classification can be carried out. There are various classification techniques which can be used for AR. In this study, three classification techniques i.e. NN, SVM, and RBF are selected due to their popularity. Based on literatures, it is hypothesised that SVM will achieve the highest accuracy due to its capability in formulating problems into convex optimization problems which guarantee to achieve the global minimum. The objectives are to evaluate different classification techniques and to identify suitable technique for AR.

5.1.2 Experiment design

The study is separated into two experiments using two activity data sets i.e. Multi-sensor activity data set and Wearable-sensor activity data set. The experiments investigate three popular classification algorithms i.e. MLP, RBF, and SVM.

5.1.3 Methodology

5.1.3.1 Classification algorithms

In this study, three classification algorithms are investigated. A brief description on these techniques is reviewed below.

1. Multi-Layer Perceptron [53]

   MLP exploits the idea of the nervous system in which numerous inputs are connected to numerous outputs. These connections are associated with weights and the outputs are usually calculated from activation functions, such as sigmoid functions, of summation of weighted inputs. MLP is capable of learning any nonlinear functions by adjusting the connection weights to minimise the error of the output. Several works on sensor-based activity classification have been conducted using MLP [152, 155, 175]. Given the input x and output o for i_{th} data. Based on the connectionist concept, the network output and can be calculated as:

   $$o_i = \phi(\sum_i w_i x_i)$$

   where ϕ is the activation or transfer function which normally is a sigmoid function e.g. logistic function, hyperbolic tangent, etc. MLP learns the classification error through the back propagation algorithm and minimises that error by adjusting the weights w_i.

2. RBF [53]

   RBF is a type of neural network which uses the radial basis function as the activation function. For N hidden neurons, the activation function is defined as:

$$f(x) = \sum_{i=1}^{N} w_i \varphi(\|x - c_i\|)$$

   where c_i is the centre vector for neuron i and φ is a kernel function e.g. Gaussian, thin plate spline, etc.

3. SVM [114]

   SVM projects inputs into a higher dimensional space so that non-linear data can be separated. It then searches for hyperplane with a maximal margin to separate the data by solving the following optimisation problem:

$$\min_{w,b,\xi}[\frac{1}{2}w^T w + C\sum_{i=1}^{m} \xi_i]$$

   subject to:

$$o_i(w^T f(x_i) + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

   The slack term ξ_i is used to relax the constraints allowing misclassified examples. The associated cost parameter C is used for penalizing ξ_i. $f()$ is a kernel function which transforms the input x_i into a higher dimensional space. Common kernel functions are such as linear kernel, RBF kernel and polynomial kernel, etc. This study uses RBF kernel function $f(x_i) = exp(-\frac{1}{(2\sigma^2)} \parallel x_i - x_j \parallel^2)$ where σ is the width of the Gaussian kernel. For K-class classification, K binary classifiers are constructed and one-VS-all classification is applied.

5.1.3.2 Data sets

Two activity data sets i.e. Multi-sensor activity data set and Wearable-sensor activity data set are used in the study. All experiments are carried out using 10-fold cross-validation where 8 folds are used for training, 1 fold for validation and 1 fold for testing. The size of the training, validation and testing data of each fold used for different data set are shown in Table 5.1.

Table 5.1: Characteristics and data partition of different data sets used in the activity classification algorithm study

Data set	# Features	# Classes	Data type	# Sample	# Training	# Validation	# Testing
Multi-sensor activity data set	63	9	Real	17,488	5,760	720	720
Wearable-sensor activity data set	141	12	Real	39,328	20160	2520	2520

5.1.4 Experimental results

5.1.4.1 Experiment 1: Multi-sensor activity data set

The neural network used in this experiment test is developed using MatLab Neural Network Toolbox®. The network has one hidden layer and the numbers of hidden nodes are selected based on the minimum error on validation sets. The RBF network used are built using MatLab Neural Network Toolbox®. The RBF parameters, SPREAD, which defines the radius of the RBF neurons are determined from 10-fold cross validation.

The classification results from MLP and RBF are not very good comparing to SVM. The highest accuracy achieved by MLP is 81.52% while RBF only achieves 72.18%. SVM, on the other hand, shows statistically better classification performance. This is because SVM encodes classification into optimization problems allowing it to solve classification as a convex problem. This means global minimum can be guaranteed. On the other hand, MLP and RBF use random weights and gradient which cannot guarantee global minimum. The analysis of confusion matrix shows that MLP, RBF, and SVM make similar misclassifications thus combining these three classifiers will not increase classification accuracy. Therefore, it is decided to use only SVM in the proposed method.

Table 5.2: Mean precision and recall of the proposed method

	Brush teeth	Dress/undress	Feed	Iron	Sleep	Sweep	Walk	Wash dishes	Watch TV
Precision	0.8495	0.7907	0.9041	0.8554	0.9814	0.9573	0.9494	0.8904	0.9606
Recall	0.8556	0.8305	0.9304	0.8551	0.9545	0.9538	0.9334	0.8600	0.9470
F-score	0.8517	0.8091	**0.9166**	0.8546	**0.9676**	**0.9553**	**0.9409**	0.8741	**0.9534**

SVM is applied which is constructed using LIBSVM [114] which is a free library for constructing the SVM model. A radial Gaussian kernel function is used. For SVM parameters, a grid search [6] using $C = 2^0, 2^{0.25}, ..., 2^{7.75}, 2^8$ and $\gamma = 2^1, 2^{1.25}, ..., 2^{3.75}, 2^4$ on validation sets are carried out using 10-fold cross validation with 10 runs. All combinations of C and γ are tested on each data set. The optimum parameters are selected based on the highest mean accuracy which are $C = 2^{2.5}$ and $\gamma = 2^3$.

A test using unseen data sets are carried out. The proposed model achieves mean classification accuracy of 90.23%, standard deviation of 1.179, and standard error mean of 0.1179. When observing classification results of each class, the model also achieves high accuracy between 83.05% and 95.45%. Table 5.2 shows mean precision, recall and F-score of the nine classes. In general, the results show high precision and recall indicating that the model is high performance (Precision = 90.43%±6.37%, Recall = 90.23%±5.06%). The average F-score of the proposed model is 0.9026 and standard deviation is 0.0567. The model performs extremely well in detecting sleeping activity. Activities such as watching TV, sweeping, walking and feeding also have been detected very well. However, the model does not perform well in detecting dressing activity.

Within the 9.77% of mean misclassification, the errors are mostly from dressing (19.27%), ironing (16.47%) and brushing teeth (16.41%) and washing dishes (15.91%) classes. Table 5.3 shows the confusion matrix of the proposed method. The numbers with the underlines show results from the model that achieves the lowest accuracy, the numbers with the bars show results from the highest accuracy, and the mean values are in between those two numbers.

The confusion matrix reveals that the model often confused dressing class with ironing (24.41%) or brushing teeth (23.82%) classes. Ironing activities are also frequently misclassified as dressing (42.45%), washing dishes (23.30%) or brushing teeth (20.36%) activities. Classification of brushing teeth is regularly

Table 5.3: Confusion matrix of the proposed method

Actual	Predict								
	Brush teeth	Dress/undress	Feed	Iron	Sleep	Sweep	Walk	Wash dishes	Watch TV
Brush teeth	$\underline{64}$ 68.45 $\overline{73}$	$\underline{3}$ 3.03 $\overline{1}$	$\underline{5}$ 2.57 $\overline{1}$	$\underline{4}$ 2.44 $\overline{1}$	$\underline{1}$ 0.52 $\overline{0}$	$\underline{0}$ 0 $\overline{0}$	$\underline{0}$ 0.02 $\overline{0}$	$\underline{1}$ 2.37 $\overline{2}$	$\underline{2}$ 0.060 $\overline{2}$
Dress/undress	$\underline{6}$ 3.23 $\overline{4}$	$\underline{57}$ 66.44 $\overline{68}$	$\underline{4}$ 1.23 $\overline{1}$	$\underline{4}$ 3.31 $\overline{1}$	$\underline{0}$ 0.07 $\overline{0}$	$\underline{3}$ 1.63 $\overline{1}$	$\underline{3}$ 1.97 $\overline{2}$	$\underline{2}$ 1.85 $\overline{3}$	$\underline{1}$ 0.27 $\overline{0}$
Feed	$\underline{4}$ 1.96 $\overline{1}$	$\underline{2}$ 0.76 $\overline{2}$	$\underline{72}$ 74.43 $\overline{75}$	$\underline{0}$ 0.74 $\overline{0}$	$\underline{0}$ 0.27 $\overline{0}$	$\underline{0}$ 0.02 $\overline{0}$	$\underline{0}$ 0.08 $\overline{0}$	$\underline{1}$ 0.80 $\overline{0}$	$\underline{1}$ 0.94 $\overline{2}$
Iron	$\underline{5}$ 2.36 $\overline{2}$	$\underline{4}$ 4.92 $\overline{2}$	$\underline{1}$ 0.77 $\overline{1}$	$\underline{65}$ 68.41 $\overline{72}$	$\underline{0}$ 0.10 $\overline{0}$	$\underline{1}$ 0.18 $\overline{0}$	$\underline{0}$ 0.14 $\overline{1}$	$\underline{3}$ 2.70 $\overline{2}$	$\underline{1}$ 0.42 $\overline{0}$
Sleep	$\underline{0}$ 0.72 $\overline{0}$	$\underline{2}$ 1.13 $\overline{1}$	$\underline{1}$ 0.71 $\overline{0}$	$\underline{4}$ 0.29 $\overline{0}$	$\underline{72}$ 76.36 $\overline{79}$	$\underline{0}$ 0.09 $\overline{0}$	$\underline{0}$ 0.18 $\overline{0}$	$\underline{0}$ 0.25 $\overline{0}$	$\underline{1}$ 0.27 $\overline{0}$
Sweep	$\underline{0}$ 0.03 $\overline{0}$	$\underline{0}$ 1.49 $\overline{1}$	$\underline{0}$ 0.11 $\overline{0}$	$\underline{0}$ 0.10 $\overline{1}$	$\underline{0}$ 0.08 $\overline{0}$	$\underline{77}$ 76.30 $\overline{76}$	$\underline{2}$ 1.57 $\overline{1}$	$\underline{1}$ 0.23 $\overline{0}$	$\underline{0}$ 0.09 $\overline{1}$
Walk	$\underline{1}$ 0.24 $\overline{0}$	$\underline{1}$ 3.18 $\overline{2}$	$\underline{0}$ 0.11 $\overline{0}$	$\underline{0}$ 0.14 $\overline{0}$	$\underline{0}$ 0.04 $\overline{0}$	$\underline{3}$ 1.34 $\overline{0}$	$\underline{75}$ 74.67 $\overline{78}$	$\underline{0}$ 0.09 $\overline{0}$	$\underline{0}$ 0.19 $\overline{0}$
Wash dishes	$\underline{4}$ 2.72 $\overline{3}$	$\underline{2}$ 2.82 $\overline{1}$	$\underline{2}$ 1.27 $\overline{1}$	$\underline{1}$ 3.85 $\overline{4}$	$\underline{0}$ 0.12 $\overline{0}$	$\underline{0}$ 0.04 $\overline{0}$	$\underline{0}$ 0.01 $\overline{0}$	$\underline{71}$ 68.80 $\overline{71}$	$\underline{0}$ 0.37 $\overline{0}$
Watch TV	$\underline{0}$ 1.00 $\overline{1}$	$\underline{1}$ 0.46 $\overline{0}$	$\underline{2}$ 1.22 $\overline{2}$	$\underline{1}$ 0.80 $\overline{0}$	$\underline{1}$ 0.26 $\overline{0}$	$\underline{1}$ 0.14 $\overline{0}$	$\underline{0}$ 0.07 $\overline{0}$	$\underline{1}$ 0.29 $\overline{0}$	$\underline{73}$ 75.76 $\overline{77}$

Note: $\underline{n}$ indicates minimum and $\overline{n}$ indicates maximum

confused with dressing (26.23%), feeding (22.25%), ironing (21.13%) and washing dishes (20.52%).

1. H1: The proposed method can achieve more than 90% accuracy
   The result from the Shapiro-Wilk test indicates that the data are normally distributed (SW=0.979, df=99, p=0.107). The result of the null hypothesis testing is $H_0 : \mu \leq 90.00$ which indicates that the accuracy difference is significant at the 5% level on a one-tailed test (T=2.336, df=99, p=0.0296). Therefore, the null hypothesis is rejected in favour of the experimental hypothesis that the mean accuracy of the proposed method is higher than 90% indicating that the method can accurately detect elderly ADL. Particular classes with F-score higher than 90% are hilighted with bold faces in Table 5.2.

2. H2: Classification accuracy can be improved by combining data from tem-

perature sensor and/or altimeter with accelerometer
In order to control the experiment, the number of inputs is set to 16 to comply with the number of selected features used in our proposed method. For configuration A, 16 top accelerometer features based on feature rankings in Section 4.2.4 are selected. We select the best combination of features for both configurations B and C through the experimentations. 16 features in configuration D is selected from the proposed method (see Table 4.12).

Classifications are conducted using SVM on 10-fold cross validation × 10 times = 100 data sets. Optimum SVM parameters C and γ are selected using the grid search for each configuration. The results of classification accuracy using features from only accelerometer, accelerometer with temperature sensor, and accelerometer with altimeter, and combination of these sensors are shown in Table 5.4. The result shows that classification accuracy is increased when temperature or altimeter is combined with accelerometer. The data are tested for normality using the Shapiro-Wilk test which reveals that the data are not normally distributed ($p < 0.001$). Thus, it is appropriate to use non-parametric statistics for hypothesis testing. The Kruskal-Wallis test at 5% significance level is used to test the null hypothesis that the median classification accuracies are the same across all configurations. The result indicates that there is a statistically significant difference in median of accuracies between different configurations ($H(3)=305.730$, $p < 0.001$) with a mean rank of 50.59 for using only accelerometer, 170.49 for using accelerometer with temperature sensor, 265.06 for using accelerometer with altimeter, and 315.87 for using combination of all three sensors.

A further pair-wise comparison between configuration D and others e.g. A VS D, B VS D, C VS D are conducted using the Mann-Whitney U test. The comparison results indicate that there is a statistically significant difference in median accuracy between configuration A and D ($U=0.00$, $p < 0.001$), B and D ($U=660.00$, $p < 0.001$), C and D ($U=2803.50$, $p < 0.001$). The results also indicate that the mean rank of configuration D is significantly higher than other configurations. Therefore, it can be concluded that by combining data from temperature sensor and/or altimeter with accelerom-

Table 5.4: Classification accuracies of using different set of sensors

Configuration	Sensor	Selected features	*Accuracy*
A	Accelerometer	RMS_Y, RMS_X, MAX_Y, MIN_Y, MIN_Z, MIN_X, DIF_Y, MAX_Z, KEY_Y, COR_{XY}, MAX_{NORM}, DIF_Y, $MEAN_Y$, MAX_X, $MEAN_Z$, $RANGE_Z$	82.7694%
B	Accelerometer, Temperature	RMS_Y, RMS_X, MAX_Y, MIN_Y, MIN_Z, MIN_X, DIF_Y, MAX_Z, KEY_Y, $MEAN_{TEMP}$, STD_{TEMP}, MAX_{TEMP}, MIN_{TEMP}, $RANGE_{TEMP}$, ENT_{TEMP}, KEY_{TEMP}	87.5764%
C	Accelerometer, Altimeter	RMS_Y, RMS_X, MAX_Y, MIN_Y, MIN_Z, MIN_X, DIF_Y, MAX_Z, KEY_Y, $MEAN_{ALT}$, STD_{ALT}, MAX_{ALT}, MIN_{ALT}, ENE_{ALT}, ENT_{ALT}, KEY_{ALT}	89.3736%
D	Accelerometer, Temperature, Altimeter	RMS_Y, RMS_X, MAX_Y, MIN_Y, MIN_Z, MIN_X, DIF_Y, MAX_Z, KEY_Y, COR_{XY}, MAX_{NORM}, DIF_Y, ENT_{ALT}, $MEAN_{TEMP}$, MIN_{TEMP}, KEY_{TEMP}	90.2250%

eter, classification accuracy can be improved. The result show that using a combination of accelerometer, temperature sensor and altimeter achieves the highest classification accuracy among other configurations.

5.1.4.2 Experiment 1: Discussion

Different classification models are compared based on MLP, RBF and SVM. The results indicate that SVM is the most powerful classification algorithm. Therefore, we propose a wrist-worn multi-sensors based AR and classification method for detecting elderly ADL using SVM. The proposed method achieves high classification performance of F-score between 0.81 and 0.97 and overall accuracy of 90.23%. This demonstrated that the proposed method performs very well on detecting activities of an elderly person. The method can detect several daily activities including basic ADL such as feeding, brushing teeth, dressing, walking sleeping and I-ADL such as washing dishes, ironing, sweeping floor and watching TV.

The confusion matrix reveals that the proposed method often gets confused among dressing, ironing, brushing teeth and washing dishes activities. Dressing is the most difficult activity to be detected as there is no clear pattern on how this activity should be performed e.g. one participant may undress/dress her top first while the other may do in different sequences. Finding a generalised decision boundary for the dressing activity proved to be challenging. For the other three classes i.e. ironing, brushing teeth and washing dishes, the results could be implied that these classes have some common characteristics. Ironing, brushing teeth and washing dishes are all involved in some kind of repetitive stroke motion e.g. back-and-forth motion. The analysis shows that maximum and minimum acceleration on Y-axis of ironing and washing dishes activities are highly overlapped. Nevertheless, when comparing to previous works [170], our method still achieves a higher classification result on these activities. Also, the proposed method can accurately detect less active activities i.e. sleeping and watching TV (F-score=95.45% and 96.04%, respectively comparing to 93.9%).

The results from the experiment indicate that, in our application, accelerometer is the most valuable sensor for AR. This result supports the previous finding from the literature [152]. The temperature sensor and altimeter when using on their own do not achieve good classification results comparing to the accelerometer. However, the experimental results reveal that by adding information from temperature sensor or altimeter, the classification performance is statistically improved, and that the combination of all three sensors achieves the highest classification accuracy confirming our hypothesis. This result supports the theory that a variable that is completely useless by itself can provide significant performance improvement when taken with others [5]. Features from accelerometer when taken with features from temperature and altitude improve class separability resulting in better classification performances.

Table 5.5 shows a comparison between the proposed method and previous works. The proposed method can achieve comparable or even higher accuracy comparing to previous works considering the sensor locations and number of recognised activities.

5.1.4.3 Experiment 2: Wearable-sensor activity data set

The classification models are developed using classification algorithms as described in Section 5.1.3.1 with 24 selected features as shown in Table 4.12. Also, to demonstrate that the proposed method using more sensors can achieve better accuracy, we construct another model where 16 features from three sensors are

Table 5.5: Accuracy between the proposed method and previous works

Author	Recognised activities	Brush teeth	Dress/ undress	Feed	Sleep	Walk	Wash dish	Iron	Average
Proposed model	9	85.56%	83.05%	93.04%	95.45%	93.34%	86.00%	85.51%	90.23%
Fleury et al. [170]	7	64.30%*	75.00%	97.80%	93.9%	95.00%	-	-	86.20%
Maurer et al. [75]	6	-	-	-	-	>90%	-	-	87.10%
Huang et al. [76]	4	85.00%	-	84.00%	-	-	76.00%	-	82.00%
Hong et al. [166]	-	-	-	-	92.66%	84.36%	-	97.94%	-

* activities include wash hand and teeth are detected

used and classification is based on SVM in previous experiment. From here, we shall refer this model as $SVM16_{3S}$. As the $SVM16_{3S}$ uses only 16 features, we also constructed classification models using truncation point at 16 features. The notation of the model name is given by the algorithm, number of features, and number of sensors. For example, $RBF16_{7S}$ represents the classification model using RBF with 16 features from 7 sensors.

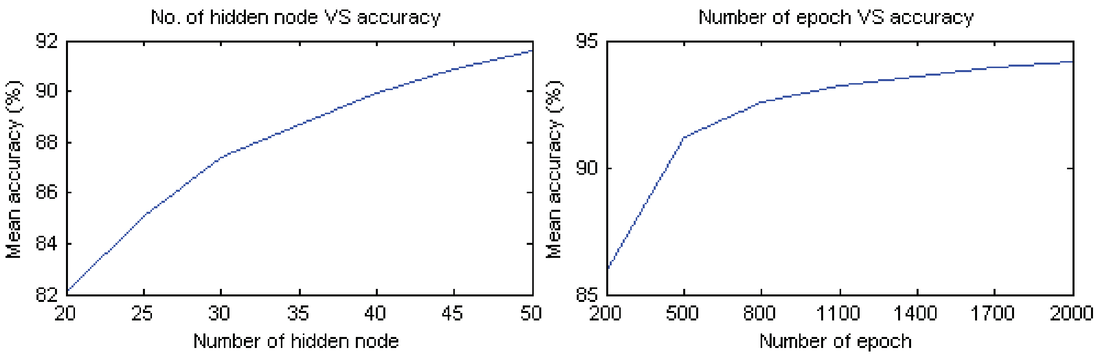

Figure 5.1: A plot between classification accuracy and number of hidden nodes in MLP with 16 features.

Firstly, we carry out the experiments to determine the optimum number of hidden node and epoch used for the neural network model. All experiments are done using validation data set 10-fold cross validation for 10 runs. It can be seen from Figure 5.1 that the more hidden nodes, the higher accuracies. However, using a large number of hidden nodes will increase the complexity with the network model. It is decided to use alpha = 3 where number of hidden nodes = alpha*input. Since we have a trade-off of using lower number of hidden nodes, it is decided to use 2000 epoch for training so that high accuracy can be achieved. The network is set up using the mentioned configuration and trained using the scaled conjugate gradient and the logistic output function. The network is validated using test set with 10-fold cross validation and 10 runs. The network is trained and tested 5 times and then, network with highest accuracy is selected. The mean accuracy obtained is 94.8496 ± 0.4207%. Confusion matrix of the $MLP16_{7S}$ is presented in Table 5.6, and precision, recall, and F-score are presented in 5.7.

Table 5.6: Confusion matrix of the $MLP16_{7S}$

Actual	Predict											
	Brush	Exercise	Feed	Iron	Read	Scrub	Sleep	Stairs	Walk	Wash	Watch	Wipe
Brush	**19810**	101	324	119	85	34	103	121	1	176	44	82
Exercise	84	**19901**	59	228	137	85	7	104	62	95	34	204
Feed	376	81	**19075**	473	116	157	116	102	2	368	92	42
Iron	96	256	290	**19600**	75	187	15	34	23	170	37	217
Read	126	124	206	120	**20072**	55	47	14	24	80	111	21
Scrub	63	112	43	56	28	**20037**	21	91	69	116	72	292
Sleep	195	17	126	61	50	64	**20071**	168	12	95	48	93
Stairs	184	61	87	84	15	206	53	**19877**	212	45	127	97
Walk	2	25	10	25	7	105	5	247	**20472**	16	7	31
Wash	123	111	402	219	65	129	26	49	17	**19740**	38	81
Watch	21	36	58	56	82	64	47	119	7	26	**20464**	20
Wipe	41	174	58	107	22	307	29	102	72	160	26	**19902**

Table 5.7: The precision, recall and F-score of the $MLP16_{7S}$

Activity	Precision	Recall	F-score
Brush	0.9379	0.9433	0.9406
Exercise	0.9477	0.9477	0.9477
Feed	0.9198	0.9083	0.9140
Iron	0.9268	0.9333	0.9301
Read	0.9671	0.9558	0.9614
Scrub	0.9350	0.9541	0.9445
Sleep	0.9772	0.9558	0.9663
Stairs	0.9453	0.9444	0.9448
Walk	0.9761	0.9771	0.9766
Wash	0.9361	0.9400	0.9381
Watch	0.9699	0.9745	0.9722
Wipe	0.9440	0.9477	0.9459

Similar experiments are carried out to determine the appropriate number of hidden nodes and epochs for $MLP24_{7S}$. It is decided to use alpha = 3 and epoch =2000. The model is tested using test set with 10-fold cross validation and 10 runs. The network is trained and tested 5 times and then, network with highest accuracy is selected. The mean accuracy is 96.7349 ± 0.3705%. The classification results including confusion matrix, precision, recall, and F-score of the $MLP24_{7S}$ are presented in Table 5.8 and Table 5.9.

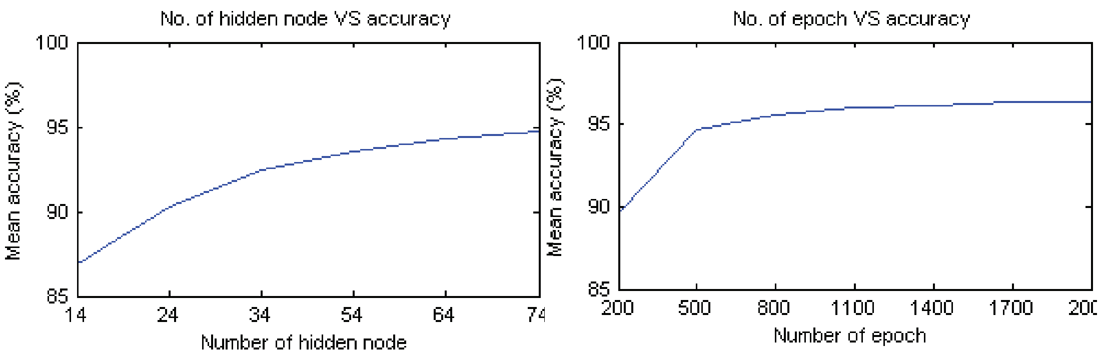

Figure 5.2: A plot between classification accuracy and number of hidden nodes in MLP with 24 features.

Table 5.8: Confusion matrix of the $MLP24_{7S}$

Actual	Predict											
	Brush	Exercise	Feed	Iron	Read	Scrub	Sleep	Stairs	Walk	Wash	Watch	Wipe
Brush	**20230**	67	214	58	46	28	80	95	0	97	26	59
Exercise	95	**20364**	58	111	63	37	14	53	17	76	18	94
Feed	202	44	**19859**	257	107	55	86	85	2	214	54	35
Iron	65	85	199	**20111**	44	84	32	46	23	106	25	180
Read	67	54	109	86	**20482**	34	22	25	13	31	63	14
Scrub	36	37	20	70	19	**20390**	22	62	22	67	38	217
Sleep	85	16	121	44	17	52	**20364**	142	10	59	26	64
Stairs	113	55	63	48	14	86	83	**20293**	130	43	61	59
Walk	5	21	12	26	5	36	10	154	**20634**	17	13	19
Wash	79	70	203	144	44	75	28	57	20	**20172**	20	88
Watch	16	12	32	28	53	27	33	73	8	21	**20677**	20
Wipe	30	88	35	123	25	226	36	61	47	105	28	**20196**

For SVM classification, firstly, a search for optimum C and γ is carried out. A rough grid search is done using 10-fold cross validation using $C = 2^b$ where b is [-5, 15] and $\gamma = 2^c$ where c is [-15, 3]. Next, a fine grid search is carried out using 10-fold cross validation with 10 runs using $C = 2^b$ and $\gamma = 2^c$ where b and c are the selected power from coarse grid search and their values are between [b-1, b+1] and [c-1, c+1]. The parameters with the highest averaged validation accuracy are chosen for the model.

For SVM with 16 features $SVM16_{7S}$, C = 23.75 and gamma = 22.5 are used.

Table 5.9: The precision, recall and F-score of the $MLP24_{7S}$

Activity	Precision	Recall	F-score
Brush	0.9623	0.9633	0.9628
Exercise	0.9737	0.9697	0.9717
Feed	0.9491	0.9457	0.9474
Iron	0.9529	0.9577	0.9553
Read	0.9791	0.9753	0.9772
Scrub	0.9650	0.9710	0.9680
Sleep	0.9786	0.9697	0.9741
Stairs	0.9597	0.9641	0.9619
Walk	0.9860	0.9848	0.9854
Wash	0.9602	0.9606	0.9604
Watch	0.9823	0.9846	0.9835
Wipe	0.9597	0.9617	0.9607

The accuracy averaged over 100 results is 96.9575% ± 0.3485%. The confusion matrix and other results are shown in Table 5.10 and Table 5.11.

Table 5.10: Confusion matrix of the $SVM16_{7S}$

Actual	Predict											
	Brush	Exercise	Feed	Iron	Read	Scrub	Sleep	Stairs	Walk	Wash	Watch	Wipe
Brush	**20286**	20	240	73	62	20	39	78	0	97	34	51
Exercise	49	**20427**	44	84	83	58	7	52	19	64	33	80
Feed	247	78	**19876**	224	110	51	58	67	1	203	40	45
Iron	86	123	160	**20097**	37	105	19	70	4	104	15	180
Read	62	62	195	94	**20397**	28	17	18	11	53	40	23
Scrub	19	66	20	59	17	**20442**	17	65	10	67	38	180
Sleep	119	10	64	19	9	49	**20487**	116	4	37	22	64
Stairs	71	43	90	60	7	112	32	**20422**	82	30	45	54
Walk	0	26	6	11	9	63	12	141	**20642**	13	1	28
Wash	64	67	249	114	51	70	15	36	14	**20249**	15	56
Watch	22	21	15	25	38	21	29	98	4	18	**20690**	19
Wipe	25	66	12	95	29	196	19	68	30	114	28	**20318**

For SVM with 24 features, $C = 24.25$ and $\gamma = 22$ are used. The accuracy averaged over 100 results is 97.2040% ± 0.3103%. The classification results are shown in Table 5.12 and Table 5.13.

For RBF, experiments are carried out to determine the appropriate number of hidden nodes and activation function among Gaussian function, Thin Plate Spline (TPS) function, and r4logr. The experiments are done using 10-fold cross validation. Firstly, we experiment with different activation functions using a fixed hidden node. The result shows that using r4logr function achieves the highest validation accuracy (See Figure 5.3). Next, experimentations using the r4logr function with different hidden nodes show that the accuracy is increased when

Table 5.11: The precision, recall and F-score of the $SVM16_{7S}$

Activity	Precision	Recall	F-score
Brush	0.9637	0.9660	0.9649
Exercise	0.9723	0.9727	0.9725
Feed	0.9478	0.9465	0.9471
Iron	0.9591	0.9570	0.9580
Read	0.9783	0.9713	0.9748
Scrub	0.9636	0.9734	0.9685
Sleep	0.9873	0.9756	0.9814
Stairs	0.9619	0.9703	0.9661
Walk	0.9914	0.9852	0.9883
Wash	0.9620	0.9642	0.9631
Watch	0.9852	0.9852	0.9852
Wipe	0.9630	0.9675	0.9653

Table 5.12: Confusion matrix of the $SVM24_{7S}$

Actual	Predict											
	Brush	Exercise	Feed	Iron	Read	Scrub	Sleep	Stairs	Walk	Wash	Watch	Wipe
Brush	**20246**	29	280	65	53	20	30	85	0	101	47	44
Exercise	40	**20667**	26	57	20	16	1	38	17	59	11	48
Feed	289	35	**19824**	197	142	67	67	63	2	186	78	50
Iron	91	69	162	**20210**	30	56	10	62	8	127	14	161
Read	61	32	154	101	**20463**	14	25	20	5	40	68	17
Scrub	9	23	34	58	6	**20549**	8	29	4	38	40	202
Sleep	65	9	70	24	21	37	**20526**	124	2	28	26	68
Stairs	86	37	96	38	14	55	44	**20498**	99	22	30	29
Walk	0	33	3	8	6	38	8	153	**20670**	6	0	27
Wash	78	28	208	123	54	66	19	34	19	**20278**	18	75
Watch	13	6	19	8	55	6	30	72	6	20	**20742**	23
Wipe	43	52	17	112	25	164	21	48	20	195	29	**20274**

the hidden nodes are increased. It is decided to use 3000 hidden nodes as the accuracy starts stabilise. RBF network is constructed using activation function r4logr where the activation, Z, is calculated as $Z(r) = r^4 \log r$ with 3000 hidden nodes and the linear output function. The model is tested using 10-fold cross validation with 10 runs. The network is trained and tested 5 times and then, network with highest accuracy is selected. The averaged accuracy is $95.3075 \pm 0.4133\%$.

Similar experiments are carried out for RBF with 24 features. The results show that using the TPS activation function achieves the highest accuracy. For the number of hidden nodes, the results indicate that the more hidden nodes, the higher the accuracies. However, it is decided to use 3000 hidden nodes as the accuracy becomes stable after this setting.

RBF network is constructed using activation function $r^2 \log r$ with 3000 hidden

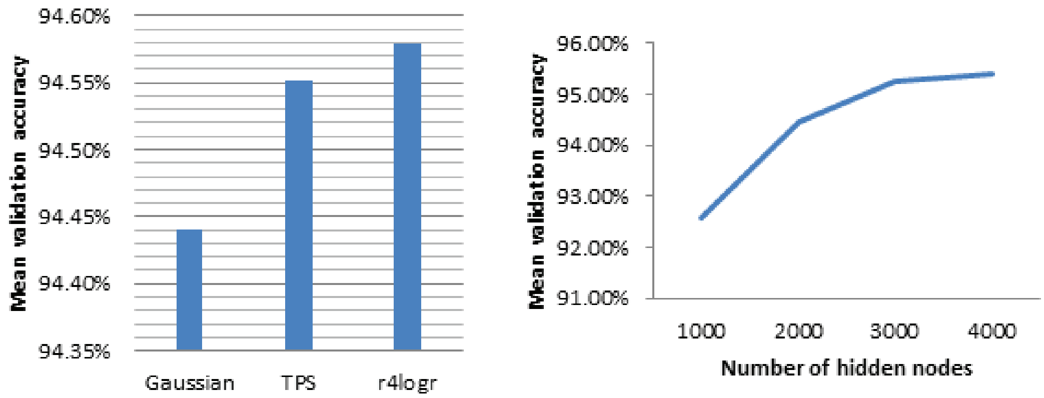

Figure 5.3: A plot between classification accuracy and number of hidden nodes in RBF with 16 features.

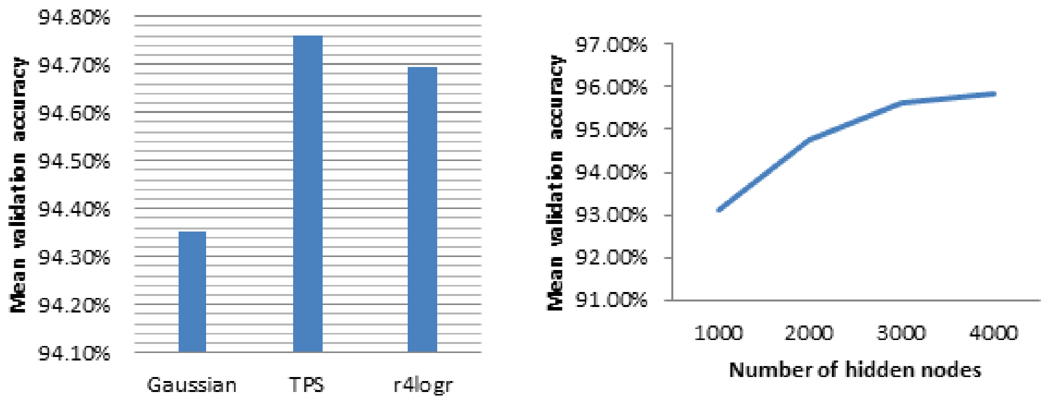

Figure 5.4: A plot between classification accuracy and number of hidden nodes in RBF with 24 features.

Table 5.13: The precision, recall and F-score of the $SVM24_{7S}$

Activity	Precision	Recall	F-score
Brush	0.9631	0.9641	0.9636
Exercise	0.9832	0.9841	0.9837
Feed	0.9488	0.9440	0.9464
Iron	0.9623	0.9624	0.9624
Read	0.9796	0.9744	0.9770
Scrub	0.9744	0.9785	0.9765
Sleep	0.9873	0.9774	0.9824
Stairs	0.9657	0.9739	0.9698
Walk	0.9913	0.9865	0.9889
Wash	0.9610	0.9656	0.9633
Watch	0.9829	0.9877	0.9853
Wipe	0.9646	0.9654	0.9650

Table 5.14: Confusion matrix of the $RBF16_{7S}$

Actual	Predict											
	Brush	Exercise	Feed	Iron	Read	Scrub	Sleep	Stairs	Walk	Wash	Watch	Wipe
Brush	**19850**	79	501	81	63	39	37	105	0	155	32	58
Exercise	137	**19869**	125	171	90	119	14	55	64	208	48	100
Feed	491	51	**19218**	391	106	196	43	73	4	301	67	59
Iron	113	133	260	**19920**	22	141	5	40	25	172	18	151
Read	133	52	445	104	**19974**	33	44	7	10	48	124	26
Scrub	17	28	54	55	22	**20277**	14	21	15	94	90	313
Sleep	175	14	145	22	10	66	**20229**	151	1	67	51	69
Stairs	149	39	175	73	22	239	30	**19950**	138	64	88	81
Walk	17	18	22	39	23	91	9	191	**20455**	40	6	41
Wash	158	73	441	196	68	119	15	20	9	**19799**	16	86
Watch	27	13	33	18	65	61	59	69	5	22	**20608**	20
Wipe	80	82	58	63	17	300	25	55	43	210	41	**20026**

nodes and the linear output function. The model is tested using test set 10-fold cross validation with 10 runs. The network is trained and tested 5 times and then, network with highest accuracy is selected. The averaged accuracy is $95.6734 \pm 0.3744\%$.

The data normality is tested using the Shapiro-Wilk and the results indicate that they have normal distribution ($p \geq 0.05$). Thus, the Paired-sample T-test is used to test the accuracy difference between each model and the result is shown in Table 5.18. The result indicates that the differences between each model are statistically significant where $SVM24_{7S} >^* SVM16_{7S} >^* MLP24_{7S} >^* RBF24_{7S} >^* RBF16_{7S} >^* MLP16_{7S} >^* SVM16_{3S}$ where $>^*$ indicates significantly better at 95% confidence interval. An experiment to test if there is a difference in accuracy when 16 and 24 features are used is carried out. The result indicates that using 24 features achieves statistically higher accuracy than using

Table 5.15: The precision, recall and F-score of the $RBF16_{7S}$

Activity	Precision	Recall	F-score
Brush	0.9299	0.9452	0.9375
Exercise	0.9715	0.9461	0.9587
Feed	0.8948	0.9151	0.9049
Iron	0.9426	0.9486	0.9456
Read	0.9752	0.9511	0.9630
Scrub	0.9352	0.9656	0.9502
Sleep	0.9856	0.9633	0.9743
Stairs	0.9620	0.9478	0.9549
Walk	0.9849	0.9763	0.9806
Wash	0.9348	0.9428	0.9388
Watch	0.9726	0.9813	0.9769
Wipe	0.9523	0.9536	0.9529

Table 5.16: Confusion matrix of the $RBF24_{7S}$

Actual	Predict											
	Brush	Exercise	Feed	Iron	Read	Scrub	Sleep	Stairs	Walk	Wash	Watch	Wipe
Brush	**19714**	13	650	45	87	22	57	106	0	206	37	63
Exercise	119	**20171**	80	129	40	58	10	52	16	200	35	90
Feed	490	8	**19432**	284	82	220	39	62	8	204	106	65
Iron	111	63	253	**20074**	16	72	11	58	30	137	19	156
Read	96	19	383	90	**20048**	21	87	11	4	61	157	23
Scrub	11	18	55	42	11	**20301**	11	23	14	66	138	310
Sleep	125	9	129	55	15	54	**20224**	180	2	45	96	66
Stairs	201	32	172	52	23	83	30	**20056**	173	54	102	70
Walk	13	14	16	39	21	57	8	188	**20524**	31	9	32
Wash	147	52	273	170	47	90	9	29	7	**20022**	24	130
Watch	30	5	37	21	51	25	74	56	4	19	**20663**	15
Wipe	80	41	48	108	29	283	39	76	13	378	37	**19868**

16 features ($p < 0.05$).

The results reveal that SVM is the best classification model among others. In general, the models can classify walking very well. However, they have difficulty in classifying feeding activity. The result shows that in our data set SVM is superior to MLP and RBF. $SVM24_{7S}$ achieves the highest classification accuracy while $MLP16_{7S}$ achieves the lowest accuracy. When observing the F-score for each class, it is found that in general $SVM24_{7S}$ obtains the highest score, especially for exercise activity. $SVM16_{7S}$ achieves slightly better result in classifying brushing teeth and feeding than $SVM24_{7S}$. When observing precision and recall, it can be seen that $SVM16_{7S}$ achieves higher precision in washing dishes and watching TV comparing to $SVM24_{7S}$. While $SVM24_{7S}$ has higher sensitivity in obtaining these classes, $SVM16_{7S}$ makes prediction more accurately.

When examining classification algorithms using 24 features, it is found that

Table 5.17: The precision, recall and F-score of the $RBF24_{7S}$

Activity	Precision	Recall	F-score
Brush	0.9327	0.9388	0.9357
Exercise	0.9866	0.9605	0.9734
Feed	0.9026	0.9253	0.9138
Iron	0.9510	0.9559	0.9534
Read	0.9794	0.9547	0.9669
Scrub	0.9537	0.9667	0.9602
Sleep	0.9818	0.9630	0.9723
Stairs	0.9598	0.9529	0.9563
Walk	0.9870	0.9796	0.9833
Wash	0.9346	0.9534	0.9439
Watch	0.9645	0.9840	0.9741
Wipe	0.9512	0.9461	0.9486

Table 5.18: Test classification accuracy of each model

Model	Mean	Std. Error
$SVM16_{7S}$	96.9575	0.0349
$SVM24_{7S}$	97.2040	0.0310
$MLP16_{7S}$	94.8496	0.0421
$MLP24_{7S}$	96.7349	0.0371
$RBF16_{7S}$	95.3075	0.0413
$RBF24_{7S}$	95.6734	0.0375
$SVM16_{3S}$	85.4238	0.0672

SVM has the highest F-score in most classes except feeding and reading where MLP is better. RBF has the lowest F-score in every class especially in feeding which is substantially lower. However, we found that RBF has comparable or even higher precision with SVM in some classes such as exercising, and reading. MLP has a comparable F-score with SVM in brushing teeth, washing dishes and watching TV. When examining at the models using 16 features (which is not the optimal number of features), SVM has the highest F-score in all classes. The F-score of RBF is higher than that of MLP in most classes except for brushing teeth and feeding.

The statistical results indicate that our models using 7 sensors obtains a significant higher accuracy than the model based on 3 sensors regardless the classification algorithms used. The improvement in accuracy is between 9.43% and 11.78%. Next, the F-score of each class between previous work and our SVM models is compared. The results indicate that the proposed system achieves a higher F-score than $SVM16_{3S}$ model in all 12 activities (See Table 5.19). The F-score of all classes of the $SVM24_{7S}$ are higher than $SVM16_{7S}$ except for brushing teeth, feeding and wiping.

When observing the confusion matrix of $SVM24_{7S}$ (See Table 5.12), it is found that the model often confuses between feeding and brushing teeth, wiping and scrubbing, and walking and using stairs. Ironing and washing sometimes are also confused with feeding. It is observed that these activities have similar motions on the wrist.

To evaluate the trade-off between accuracy and the use of heart rate monitor, a classification without using the features from the heart rate is performed. The heart rate feature is substituted with the next best feature. The classification using MLP obtains 93.1020% ± 0.5850%. Since the data is normal distributed, we applied the Paired Sample test. The result indicates that by removing heart rate feature, the classification accuracy is significantly lowered (T=-28.993, p <0.05).

Table 5.19: F-score comparison between models based on 3 sensors and 7 sensors

Model	Brush	Exercise	Feed	Iron	Read	Scrub	Sleep	Stairs	Walk	Wash	Watch	Wipe
$SVM16_{3S}$	0.7684	0.8670	0.7575	0.8214	0.8496	0.8615	0.9478	0.8771	0.9530	0.8069	0.9398	0.8055
$SVM16_{7S}$	**0.9649**	0.9725	**0.9471**	0.9580	0.9748	0.9685	0.9814	0.9661	0.9883	0.9631	0.9852	**0.9653**
$SVM24_{7S}$	0.9636	**0.9837**	0.9464	**0.9624**	**0.9770**	**0.9765**	**0.9824**	**0.9698**	**0.9889**	**0.9633**	**0.9853**	0.9650

5.1.4.4 Experiment 2: Discussion

Comparing with $SVM16_{3S}$, the results suggest that the addition of heart rate sensor, barometer, gyroscope and light sensor improve classification accuracy. This means that they provide valuable information for classification of the activities studied. The results of the study provide suggestion on possible sensors for other activity classification systems. Also, these sensors except for heart rate monitor are used on a users wrist will allow practical applications of AR for home-based care. The results of the study show that the proposed system achieves statistically better performance.

The results show that combining heart rate with other sensor significantly improves classification accuracy. Nevertheless, the classification accuracy without using heart rate is still high comparing to $SVM16_{3S}$. This suggests that it is possible to use only wrist worn sensors to maintain the practicality and better accuracy can be achieved.

Table 5.20 indicates that the proposed model achieves comparable or in some activities higher than previous studies. Also, the proposed approach only requires sensor worn on wrist and chest. Also, results indicate that even when the heart rate sensor is removed, high accuracy can be achieved. This is an important aspect for a practical application in elder care. The system which is not intrusive or perceived as stigmatisation can be easily accepted by the elderly.

Table 5.20: Accuracy comparison between previous works and the proposed system

	#activity	Sensor location	Brush teeth	Feed	Iron	Sleep	Stairs	Walk	Average
$SVM24_{7S}$	12	Wrist,chest	**96.36**	94.64	96.24	**98.24**	**97.39**	**98.65**	**97.20**
Wang et al. [49]	12	wrists, ankles, chest	-	89.50	-	89.20	90.80	88.20	91.3
Fleury et al. [170]	7	Body, environment	64.30	**97.80**	-	93.90	-	95.00	86.20
Hong et al. [166]	-	Wrist, objects	-	-	**97.94**	92.66	-	84.36	-
Parkka et al. [155]	7	On-body	-	-	87.00	-	79.00	86.00	82–86
Maurer et al. [75]	6	Wrist	-	-	-	-	>90	-	87.10
Trabelsi et al. [48]	12	Chest, thigh, left ankle	-	-	-	95.4	-	98.1	91.4

5.1.5 Conclusion remarks

The results show that in general SVM achieves the highest accuracy followed by MLP and RBF. Classification models which use 24 features are better than ones using 16 features. Also, classification models which generate from seven sensors are better than three sensors. Nevertheless, it is found that different models are better at classifying different activities. Therefore, the classification accuracy can be improved by the combination of these classifiers.

5.2 Classifier fusion weight using Genetic Algorithm study

5.2.1 Study hypothesis and objectives

In this study, GA is used to determine the fusion weights. Studies indicate GA improve the classifier fusion accuracy [92, 93]. For example, classifier combination using 8-10 ensembles generated from different techniques is studied [92]. Weight combination using GA to combine several Bayesian classifiers is investigated [93]. However, there are some limitations on these studies. Firstly, most of them focused only on the fusion of all classifiers. For example, they produce 6 classifiers then used GA to combine them. Based on this, the conclusion that GA could improve classifier combination accuracy is not always true as all possible combinations have not been tested. Secondly, fitness functions such as function which reflects on the classifier combination function e.g. Sum, Min, Max, Product, Ranking, and Weighted Average have not been investigated before. Finally, some of these results are often compared with the mean accuracy of a set of classifiers rather than to the best individual classifier. The mean accuracy is always equal or less than the accuracy of the best individual classifier (equal accuracy is only occurred if and only if all classifiers have the same accuracy). For example, if there are 3 classifiers with 90%, 85%, 95%, the mean accuracy is 90% which is less than the best individual (95%). This weakens the conclusion that the classifier combination is better than a single classifier. In this study we investigate which fusion and weight techniques are optimums for all possible classifier fusions and compare the results with the best classifier. It is hypothesised that using GA, classification accuracy will be at least equal to the highest accuracy obtained by the best classifier, or higher. Also, we hypothesis that using fitness function which reflects the same combiner function will improve the classifier fusion result.

5.2.2 Experiment design and data set

In this study, six classifiers are generated from MLP, RBF, and SVM using 16 features and 24 features selected in feature selection study. Seven classifier combination methods and six fusion weights are investigated. Also, a method based on GA to find the optimum fusion weight is proposed and evaluated against other methods. Wearable-sensor activity data set is used in this study. All experiments are carried out using 10-fold cross-validation where 8 folds are used for training, 1 fold for validation and 1 fold for testing. The size of the training, validation and testing data of each fold used are shown in Table 5.21.

Table 5.21: Characteristics and data partition of the data set used in the study

Data set	# Features	# Classes	Data type	# Sample	# Training	# Validation	# Testing
Wearable-sensor activity data set	141	12	Real	39,328	20160	2520	2520

5.2.3 Methodology

5.2.3.1 Multi-model fusion methods

In this study we experiment with 7 fusion methods which are widely used in the classifier combination context [47, 92, 93, 103]. Given that $pred_i^j$ is the prediction of input x_i using classifier model j, $\hat{P}_{ik}^{(j)}$ is the posterior probability that x_i belongs to class k and w_j is the weight for classifier model j, the prediction of the multi-model fusion can be calculated as:

1. Majority vote (MV)

$$pred_i = mode_J\{pred_i^{(j)}\}$$

2. Product

$$pred_i = \max_K\{\frac{1}{p(C_k)^{J-1}}\prod_{j=1}^{J}(\hat{P}_{ik}^{(j)})^{w_j}\}$$

3. Sum

$$pred_i = \max_K\{\frac{1}{J}\sum_{j=1}^{J}(\hat{P}_{ik}^{(j)})^{w_j}\}$$

4. Min

$$pred_i = \max_K\{\frac{\min_J(\hat{P}_{ik}^{(j)})^{w_j}}{\sum_{k=1}^{K}\min_J(\hat{P}_{ik}^{(j)})^{w_j}}\}$$

5. Max

$$pred_i = \max_K\{\frac{\max_J(\hat{P}_{ik}^{(j)})^{w_j}}{\sum_{k=1}^{K}\max_J(\hat{P}_{ik}^{(j)})^{w_j}}\}$$

6. Ranking

   First the probability $\hat{P}_i^{(j)}$ is converted to ranks where the maximum rank score is K and minimum is 1. Given that $rank_{ik}^{(j)}$ is the ranking score of model j predicting that data x_i belong to class k, the prediction of the multi-classifier can be calculated as:

$$pred_i = \max_K \sum_{j=1}^{J} w_j rank_{ik}^{(j)}$$

7. Weighted average (WA)

   The weight accuracy $\hat{P}_k^{(j)}$ of class k using model j can be calculated as:

$$\hat{P}_k^{(j)} = \sum_{k=1}^{K} Accuracy$$

$$pred_i = \max_K \sum_{j=1}^{J} w_j \hat{P}_{ik}^{(j)}$$

In case of equal scores, the model selects the result based on the best classifier.

5.2.3.2 Fusion weight

Since each classification model may be superior to others, it is common to incorporate weights to the models to reflex this. We study 6 weight functions. Given m training examples and J models, the weight for each classifier model j can be calculated as:

1. Simple average (SA)

$$w_j = \frac{1}{J}$$

2. Variance-covariance (VACO)

   This technique uses the mean square error to calculate the weights. For a classification problem, we propose the modified version below:

$$w_j = \frac{[\sum_{i=1}^{m}(1 - \hat{P}_{iK}^{(j)})]^{-1}}{\sum_{j=1}^{J}[\sum_{i=1}^{m}(1 - \hat{P}_{iK}^{(j)})]^{-1}}$$

where $\hat{P}_{iK}^{(j)}$ is the probability that model j predicts that data x_i belongs to class K, given that the true class is K.

3. Discounted mean square forecast error (DMSFE)

$$w_j = \frac{[\sum_{i=1}^{m} \beta^{m-i+1}(1-\hat{P}_{iK}^{(j)})]^{-1}}{\sum_{j=1}^{J}[\sum_{i=1}^{m} \beta^{m-i+1}(1-\hat{P}_{iK}^{(j)})]^{-1}}$$

where β is often chosen between 0.95, 0.9, 0.85 and 0.80.

4. Unit weight

$$w_j = 1$$

5. Weighted accuracy (WACC)

$$w_j = \frac{Accuracy^{(j)}}{\sum_{j=1}^{J} Accuracy^{(j)}}$$

Note that all calculated weights must be summed to 1 i.e. $\sum_{j=1}^{J} w_j = 1$. This is except for the unit weight function where all the weights are 1.

5.2.4 The proposed Genetic Algorithm based Fusion Weight

In this study, we propose to use GA to find weights for classifiers. GA has been commonly used to solve an optimisation problem [46]. The advantage of GA over other optimisation techniques is that instead of starting at a single point to find the solution, a population of points is created. It mimics natural selection in which the population is modified over time. Individuals are randomly selected as parents to produce children of the next generation.

5.2.4.1 Fitness function

GA is used to find the weights that minimise the mean square of the combination error. The classification error is defined as follow:

$$error_i = \begin{cases} 0 & \text{if } tru_i = pred_i \\ 1 & \text{otherwise} \end{cases}$$

Unlike previous works that used average weight function i.e. $f(w) = w_1x_1 + ... + w_Kx_K$ as a fitness function, we propose to use the fitness function (ff) according to the fusion method. Given the fusion method (fm) as any function described in Section 5.2.3.1, the fitness function is defined as:

$$ff(w_j) = \frac{1}{2m}\sum_{i=1}^{m} error(tru_i, fm_i(w_j))^2$$

For example, for Sum function the following fitness function is used:

$$ff(w_j) = \frac{1}{2m}\sum_{i=1}^{m} error(tru_i, \max_K\{\frac{1}{J}\sum_{j=1}^{J}(\hat{P}_{ik}^{(j)})^{w_j}\})^2$$

Also, to investigate if using different fitness functions based on classifier fusion function would produce a better accuracy, the linear weight function is explored:

$$ff(w_j) = \frac{1}{2m}\sum_{i=1}^{m} error(tru_i, \max_K\{\frac{1}{J}\sum_{j=1}^{J}(\hat{P}_{ik}^{(j)}) * w_j\})^2$$

5.2.4.2 Population initialisation

The weight for each classifier is represented in each bit of a chromosome. For each combination, we have J bits. Each bit is represented by a real number between 0 and 1. In order to make sure that the weight obtained will result in higher classification accuracy, a population which covers the search space and at a possible optimum point is necessary. We propose to use the following strategy to initialise the population. Firstly, one of the populations must contain weighted average accuracy chromosomes. Secondly, the weights are randomly generated from a uniform distribution and the highest weight is assigned to the best model. Note that, the weight for the best model within the group is generated randomly between $\frac{1}{J}$ and 1. The initial population process can be summed up as follow:

1. Create a J-bit chromosome with weighted accuracy using $w_j = \frac{Accuracy^{(j)}}{\sum_{j=1}^{J} Accuracy^{(j)}}$

2. Set Lower bound $= \frac{1}{J}$ and Upper bound $= 1$
3. Randomly generate weight using range [Lower bound, Upper bound]
4. Update Lower bound value using Lower bound $= 0$
5. Update Upper bound value using Upper bound $= 1 - \sum_{j=1}^{J} W$

Repeat step 1) to 3) for $20 \times J - 1$ times. The initial population of $20 \times J$ chromosome are generated.

5.2.4.3 Crossover, mutation, and parents selection

The chromosome of the crossover kids are randomly selected where half of the genes are from each parent. The crossover kids are checked if their chromosomes are still valid. If not, the chromosome of the crossover kid can be calculated using:

$$Crossoverkid = \alpha \times parent1 + (1 - \alpha) \times parent2$$

where α is a uniform random number between 0 and 1.

The adaptive mutation is used where directions that are adaptive with respect to the last generation, are randomly generated. The feasible region is bounded by the constraint $(0 \leq w_j \leq 1)$. The mutant chromosome is calculated using:

$$Mutant = \text{parent} + \text{step size} \times \text{direction}$$

where steps size can be calculated using the following algorithm [12].

if the state before the last one is better than the last state **then**
  step size $= \min(1, \text{step size} \times 4)$
else
  step size $= \max(\sqrt{eps}, \frac{\text{step size}}{4})$
end if

where eps is the distance from 1.0 to the next largest double-precision number, which is 2^{-52}.

A mutant is checked so that linear constraints $(\sum_{j=1}^{J} w_j = 1)$ and bounds are satisfied. In case that the mutants chromosome is not valid, the parent chromosome is used.

The tournament style is used to select the parents. For each tournament, four chromosomes are selected randomly. Each chromosome is played against each other. The chromosome with the highest score i.e. fittest chromosome is the winner of the tournament. The winners for each tournament are selected as parents.

5.2.5 Experimental results

We perform classification using 3 algorithms with 16 and 24 features. In total, 6 classification models are produced. The classifications of the 6 models give mean accuracy between 94.85% and 97.20% with STD between 0.3088 and 0.4186. As expected, SVM performance is superior to other algorithms. However, when we investigate the precision and recall of each classifier, it is found that some classifiers are better than SVM in some activities.

Next, classifier fusion is performed. Data from training and validation set are used to determine the weight for SA, VACO, DMSFE and WACC techniques, whereas in GAFW, the training set is used in the fitness function and the validation set is used to select the weight. There are 57 possible combinations which can be generated from 6 classifier models. We present the classifier fusion result of the test dataset in Table 5.23. The classifiers fusion result is compared with the best individual classifier (BI) within the fusion group. Improvement column shows the percentage of mean difference between classifier fusion and BI. It can be seen that classifier fusion which utilises posterior probability can achieve a better result comparing to fusing the class output directly. Among 7 classifier fusion methods, sum is the best fusion technique. It improves classification accuracy by 0.3435% on average comparing to using only the best individual classifier. 95.79% of all possible combinations using the sum method achieve equal or higher accuracy than using the best classifier. This is followed by product, majority vote, weighted average, max, min, and ranking, respectively. In term of the fusion weight determination technique, in general, 98.25% of combination using GA achieves equal or higher accuracy than using one best classifier. VACO also achieves very good result of 93.86% accuracy equal or higher than BI followed by WACC, DMSFE-0.95, SA, unit weight, DMSFE-0.90, DMSFE-0.85 and

DMSFE-0.80, respectively.

A study on the computational cost required to obtain fusion weights using different methods is carried out. The cost is based on the time used to find the weights for combining two classifiers using training data set of 2,016,000 samples. For different classifier combination function, the cost is based on using the function to combine the result of two classifiers per sample. The results are shown in Table 5.24.

5.2.6 Discussion

In this study, experimentations on several classifier fusion and fusion weight determination techniques are performed. The results show that sum is the most effective fusion method and when used with SA, WACC, or GA, improvement on all combinations can be achieved. As sum technique uses the average probability produced by all classifiers, it is not heavily affected when some classifiers are over confident. On the contrary, the min technique uses the minimum probability that the data will belong to this class. Min selects the class that has the minimum objection by all classifiers. As Min is sensitive toward objection, it is affected when some inaccurate classifiers always produce low probability. Similarly, max technique selects the class which has the highest probability. Therefore, if the system contains bad classifiers that produce high probability, the system accuracy is affected. The experimental results show ranking is the worse fusion method. Although ranking reduces the bias caused by some classifiers being over confident, converting probabilities into rank also loses some information. Thus, fusing classifiers by ranking could produce conflict or wrong prediction if there are many inaccurate classifiers in the group.

In term of fusion weight techniques, we find that in general GA performs the best comparing to others shown in Table 5.23. The improvement over BI is significant ($p < 0.05$) at 95% confidence interval. Although this improvement is lower comparing to other techniques, the results show that by using GA with linear fitness function, 99.42% of the combination can achieve equal or better accuracy than using just one best classifier. Also, one should bear in mind that in our experiment we limit the search time to only 5 minutes. Better performance

Table 5.22: Classifier fusions results.
SA = Simple average, VACO = Variance-covariance, DMSFE = Discounted mean square forecast error, Unit weight = All weight equal 1, WACC = Weighted accuracy, GAFW = GA based fusion weight, BI = Best individual classifier, WA = Weighted average.

Weight function	Fusion function	Accuracy (%)	STD.	Improve (%)	<BI (%)	=BI (%)	>BI (%)
-	BI	96.9662	0.4158	0.0000	0.0000	100.0000	0.0000
-	MV	97.1522	0.4832	0.1918	7.0175	26.3158	66.6667
SA	Product	97.3865	0.4610	0.4203	0.0000	0.0000	100.0000
	Sum	97.3972	0.5004	0.4310	0.0000	0.0000	100.0000
	Min	97.2504	0.3978	0.2842	15.7895	0.0000	84.2105
	Max	97.1507	0.6207	0.1845	5.2632	0.0000	94.7368
	Ranking	96.7159	0.7208	-0.2502	57.8947	0.0000	42.1053
	WA	97.3463	0.5337	0.3802	1.7544	0.0000	98.2456
VACO	Product	97.3090	0.5709	0.3428	5.2632	0.0000	94.7368
	Sum	97.3596	0.5659	0.3934	5.2632	0.0000	94.7368
	Min	97.1990	0.6202	0.2329	7.0175	0.0000	92.9825
	Max	97.1990	0.6202	0.2329	7.0175	0.0000	92.9825
	Ranking	97.0793	0.4941	0.1131	5.2632	21.0526	73.6842
	WA	97.1385	0.5628	0.1724	7.0175	0.0000	92.9825
DMSFE-0.80	Product	97.1015	0.6188	0.1353	26.3158	0.0000	73.6842
	Sum	97.2328	0.5997	0.2666	12.2807	0.0000	87.7193
	Min	96.8771	0.6670	-0.0891	40.3509	0.0000	59.6491
	Max	97.1264	0.4592	0.1603	12.2807	0.0000	87.7193
	Ranking	96.7314	0.7690	-0.2348	45.6140	3.5088	50.8772
	WA	96.8250	0.7432	-0.1412	33.3333	0.0000	66.6667
DMSFE-0.85	Product	97.1729	0.6017	0.2067	12.2807	0.0000	87.7193
	Sum	97.2797	0.5878	0.3135	12.2807	0.0000	87.7193
	Min	96.9789	0.6261	0.0127	40.3509	0.0000	59.6491
	Max	97.1267	0.4445	0.1605	17.5439	0.0000	82.4561
	Ranking	96.8291	0.7139	-0.1371	40.3509	3.5088	56.1404
	WA	96.9086	0.6952	-0.0576	33.3333	0.0000	66.6667
DMSFE-0.90	Product	97.2610	0.5878	0.2948	12.2807	0.0000	87.7193
	Sum	97.3284	0.5750	0.3622	5.2632	0.0000	94.7368
	Min	97.1054	0.6145	0.1393	12.2807	0.0000	87.7193
	Max	97.1282	0.4266	0.1620	21.0526	0.0000	78.9474
	Ranking	96.9931	0.6304	0.0269	28.0702	3.5088	68.4211
	WA	97.0544	0.6336	0.0882	19.2982	0.0000	80.7018
DMSFE-0.95	Product	97.3399	0.5831	0.3737	5.2632	0.0000	94.7368
	Sum	97.3692	0.5658	0.4031	5.2632	0.0000	94.7368
	Min	97.2196	0.6343	0.2535	7.0175	0.0000	92.9825
	Max	97.0927	0.4114	0.1265	29.8246	0.0000	70.1754
	Ranking	97.1510	0.5651	0.1849	17.5439	1.7544	80.7018
	WA	97.2063	0.6064	0.2401	10.5263	0.0000	89.4737
Unit weight	Product	97.3866	0.4610	0.4204	0.0000	0.0000	100.0000
	Sum	97.3463	0.5337	0.3802	1.7544	0.0000	98.2456
	Min	97.2504	0.3978	0.2842	15.7895	0.0000	84.2105
	Max	97.1507	0.6207	0.1845	5.2632	0.0000	94.7368
	Ranking	96.7040	0.7098	-0.2621	59.6491	0.0000	40.3509
	WA	97.3463	0.5337	0.3802	1.7544	0.0000	98.2456
WACC	Product	97.3905	0.4597	0.4243	0.0000	0.0000	100.0000
	Sum	97.3994	0.4999	0.4332	0.0000	0.0000	100.0000
	Min	97.2585	0.3967	0.2923	14.0351	0.0000	85.9649
	Max	97.1481	0.6215	0.1819	5.2632	0.0000	94.7368
	Ranking	97.2049	0.4118	0.2387	21.0526	0.0000	78.9474
	WA	97.3530	0.5323	0.3868	1.7544	0.0000	98.2456

Table 5.23: Classifier fusions results (cont.)

Weight function	Fusion function	Accuracy (%)	STD.	Improve (%)	<BI (%)	=BI (%)	>BI (%)
GAFW+fusion function	Product	97.1910	0.2840	0.2248	0.0000	0.0000	100.0000
	Sum	97.1771	0.2932	0.2109	0.0000	0.0000	100.0000
	Min	97.0730	0.3773	0.1068	7.0175	5.2632	87.7193
	Max	97.1764	0.4275	0.2102	10.5263	1.7544	87.7193
	Ranking	97.0189	0.3710	0.0527	0.0000	68.4211	31.5789
	WA	97.0706	0.2548	0.1044	0.0000	0.0000	100.0000
GAFW+linear function	Product	97.1511	0.2613	0.1849	0.0000	0.0000	100.0000
	Sum	97.2067	0.3085	0.2405	0.0000	0.0000	100.0000
	Min	97.2067	0.3085	0.2405	0.0000	0.0000	100.0000
	Max	97.2067	0.3085	0.2405	0.0000	0.0000	100.0000
	Ranking	96.9783	0.4091	0.0121	3.5088	84.2105	12.2807
	WA	97.0648	0.2598	0.0986	0.0000	1.7544	98.2456

Table 5.24: Computational cost on different fusion weight functions and different classifier combination methods

Weight function	Cost (s)	Weight function	Cost (s)
GA sum	122.7658	VACO	0.0941
GA min	150.9691	DSMFE-0.80	0.1387
GA max	156.4008	DSMFE-0.85	0.1479
GA rank	71.9906	DSMFE-0.90	0.1378
GA prod	125.4562	DSMFE-0.95	0.1514
GA linear	119.0677	Unit weight	0.0001
SA	0.0105	WACC	0.1213

Fusion methods	Cost (s)	Fusion methods	Cost (s)
MV	0.012245	Max	0.000121
Product	0.001735	Ranking	0.001716
Sum	0.000847	WA	0.000067
Min	0.000115		

is expected if GA converges.

The results also reveal that using linear function as fitness function can find better weights, especially for Min and Max classifier combination function. However, GA with ranking function produces better results than GA with linear function. When we observed on the cases that used GA with linear function fail to improve the accuracy, it is found that the calculated weights are totally different. For example, GA with linear function gave 0.3 and 0.7 weight, while GA with ranking gave 0.7 and 0.3 weight. This is because the class probability has been converted into ranks which are different data representation than that used in linear function.

VACO also obtains very good results while using the low computational cost. The DMSFE technique performs the worst. DMSFE is similar to VACO, where β in VACO is 1, DMSFE is between 0.80 - 0.95. From our study, it can be seen that β value nearer to 1 achieves better combination accuracy. The results are similar in [54] which found VACO is better than DMSFE.

The results of the study also show that using GA to find the fusion weight uses a much higher computational cost than other functions especially when trying to optimise min and max function. Therefore, the proposed GAFW should be appropriate in the AR model that will be developed offline. For other system that needs to update the fusion weights in real time, other functions such as VACO and WACC should be used. For the classifier combination function, the computational cost is very low and can be applied in both online and offline applications.

5.2.7 Conclusion remarks

We have studied the use of GA to find fusion weights. Unlike previous studies, we compare GA performance with BI and test on all possible classifier combinations. The results show that for all possible classifier combinations and fusion methods, 99% of times GAFW can achieve higher or at least equal to the best classifier within the group. However, due to high computational cost, this technique is only suitable for offline training.

5.3 Classifier combination using Genetic Algorithm study

5.3.1 Study hypothesis and objectives

Although the results from previous study of using GA to find the optimum fusion weight show a promising result, it suffers from high computational cost. In this study, we propose to use GA to find the optimum combination model between classifiers, fusion weight and combiner functions. Simple weight functions such as SA, VACO, unit weight and weighted accuracy are used so that the classifier combination can be done online. It is hypothesisd that using GA, an optimum classifier combination can be found which resulted in higher classification result. The objectives of this study are to propose an algorithm using GA, called Genetic Algorithm based Combination Model (GACM), to find the optimum combination between different classification models, weight functions and classifier combination function and to evaluate the proposed algorithm against manual selection and best individual classifier.

5.3.2 Experimental design and data set

This study uses Wearable-sensor activity data set with 10-fold cross validation and 10 runs. The combination is based on six classifiers generated from MLP, RBF, and SVM using 16 features and 24 features. Four fusion weights including SA, VACO, unit weight and weighted accuracy and five classifier combiner functions including product, sum, minimum, maximum, and weighted average are studied. The study is carried out using 10-fold cross validation and 10 runs. The statistical tests are based on 95% confidence interval. The combination model obtained based on training data and the result is evaluated on test data. The results are compared with the combination model obtained from manually selection.

5.3.3 The proposed Genetic Algorithm based Combination Model

5.3.3.1 Problem representation

In this study, GA is used to find the optimum combination between classifiers, weight functions and combiner functions. Given a problem with m classifiers, w weight functions, and c combiner functions, the chromosome of the combination can be represented as:

$$M = \begin{bmatrix} M_1 & M_2 & \dots & M_m \\ W_1 & W_2 & \dots & W_w \\ C_1 & C_2 & \dots & C_c \end{bmatrix}$$

The values of M_i, W_i, and C_i are either 0 or 1 where 0 represents the absent of the incident i.e. classifier, weight, or combiner functions, and 1 represents the present of the incident. For example, a chromosome of

$$M = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

represents the combination model using classifiers 1 and 2 with weight function 2 and combiner function 1. There are some constraints that need to be applied on the chromosome. First, if only one classifier is selected, then the weight and combiner function must not be selected. This means that the value of the bit in the chromosome which represents the weight and combiner functions can only be 0. For example,

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

represents the combination model which consists of classifier 1. Second, if more than one classifier is selected, the weight and combiner functions must not be 0.

5.3.3.2 Fitness function

The aim is to find the combination model which yields the highest classification accuracy. Therefore, the fitness function is the mean classification error:

$$\frac{\sum_{i=1}^{N}(error(pred_i, true_i))}{N}$$

Where N is the number of training data. The error function is defined as:

$$error_i = \begin{cases} 0 & \text{if } tru_i = pred_i \\ 1 & \text{otherwise} \end{cases}$$

5.3.3.3 Fitness scaling function

The fitness score calculated from the fitness function may not be suitable for using in GA e.g. scores vary too widely or too little. If the range of the score is too wide, the chromosomes which have high scores will fill the population quickly which will prevent the GA from exploring other solution regions. On the other hand, if the score varies too little, the probability of selecting each chromosome will be similar which will make the GA progress slowly. Therefore, fitness scaling function is used for scaling the raw fitness score so that it falls into an appropriate range. In addition, we propose to add model selection criteria e.g. simpler model is preferred, etc. to the fitness scaling function such that the chromosomes which favour these criteria have higher score. The raw fitness score

$$S_i = \frac{p}{sqrt(r_i)\sum_{i=1}^{R}\frac{1}{\sqrt{(r_i)}}} \tag{5.1}$$

where p is the number of parents required for the next generation, r_i is the rank of i_{th} chromosome, and R is the total number of population. The fitness scaling algorithm is defined as follow:

In this study, the criterion is the number of classifiers used in the combination model. Therefore, according to algorithm 1, z is the index of the population sort by the number of classifiers in the ascending order.

Algorithm 1 Fitness scaling algorithm

function FITNESSSCALING($score, \#parent, population$)
  i = SORT($score$) ▷ Sort score in ascending order
  $n = \#score$
  $x = 1$
  while $x \leq n$ **do**
    y = FINDLOCATION(i, x, n) ▷ Return locations where score is equal to the minimum score in range x to n
    if y is not empty **then**
      z = Sort population index y by criteria
      $i(y) = i(y(z))$
    end if
    $x = y(end) + 1$
  end while
  Calculate scaled score using equation (5.1)
return $scaledScore$
end function

5.3.3.4 Selection function

In this study, a tournament selection strategy is used. A group of n chromosomes are competed in a tournament and a parent is the chromosome that wins the tournament. The number of parents required for the next generation can be calculated as [12]:

$$N_P = (2 \times N_{CK}) + N_{MK}$$

where N_P is the number of parents, N_{CK} is the number of cross over kids and N_{MK} is the number of mutant kids.

5.3.3.5 Crossover function

The crossover kids are reproduced based on two parents. The number of crossover kids can be calculated using:

$$N_{CK} = \eta(N_{POP} - N_E)$$

where η is the crossover fraction, N_{POP} is the population size, and N_E is the number of elite. The following strategies are used to combine chromosome between the two parents.

1. Randomly swap each genes which represent classifiers between the two parents.

2. Randomly swap the genes which represent weight function between two parents.

3. Randomly swap the genes which represent combiner function between two parents.

Note that the successful rate of the swap is $\alpha\%$ and the crossover kid is based on the parent that is the fittest. For example, figure 5.5 shows the result of the crossover between two parents, given that parent 2 is fitter than parent 1 and α is 20%. The red arrow indicates the genes that are successfully swapped, while the black arrow indicates the unsuccessful swap.

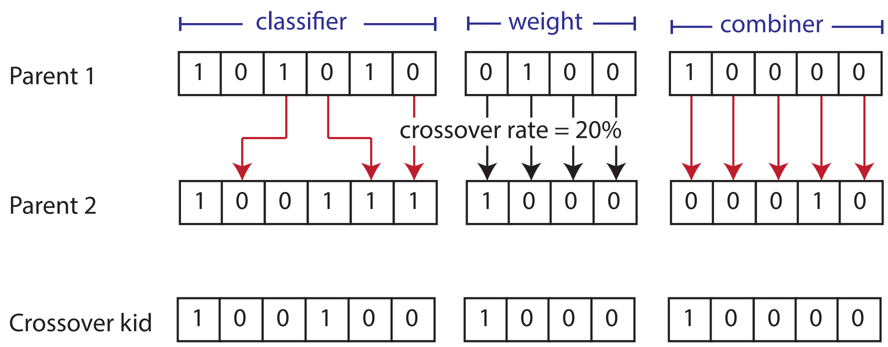

Figure 5.5: An example of crossover process between two parents

5.3.3.6 Mutation function

The mutation kids are produced based on the mutation of the genes of a parent. The size of mutation kids can be calculated from:

$$N_{MK} = N_{POP}(N_E + N_{CK})$$

The processes of mutation are as follow:

1. Randomly mutate each gene which represent classifiers by changing the value from 0 to 1, or 1 to 0.

2. Randomly mutate the genes which represent the weight function by randomly selecting a weight function.

3. Randomly mutate the genes which represent combiner function by randomly selecting a combiner function.

Note that the successful rate of the mutation is $\beta\%$. Figure 5.6 shows the mutation result of a chromosome, given β is 20%. The red arrow indicates the genes that are successfully mutated, while the black arrow indicates the unsuccessful mutation.

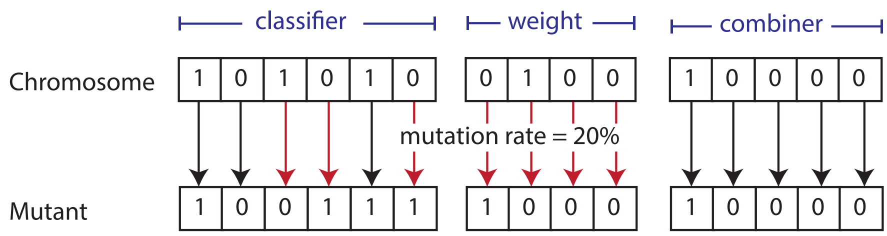

Figure 5.6: An example of mutation process from a parent

Hill climbing algorithm is used in both mutation and crossover processes to ensure that the chromosomes generated from these processes will not decrease the mean fitness values of the whole populations. Each time the chromosome is generated either from crossover or mutation process, it is checked if its fitness value is lower than the mean fitness values of the whole populations. If the fitness value of the generated chromosome is lower, then we attempt n times to generate another chromosome.

5.3.4 Experimental results

For each fold, all possible combinations (1,146 classification models) are generated and then used to calculate the mean classification error using training data. The results in Table 5.25 shows the minimum error selected manually from each fold and run. Table 5.26 shows the mean classification error of the combination model generated by GACM using training data. The bold fonts indicate where there is a difference in error between manual selection and GACM. From these tables, it can be seen that 15% of times GACM cannot find the most optimum combination. However, the average error difference between manual selection and GACM is only 0.0176.

Table 5.25: Minimum error based on all possible combination. The error are based on training data.

Run no.	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Run 1	0.2249	0.4541	0.4718	0.4630	0.4630	0.4674	0.4674	0.4718	0.4718	0.2249
Run 2	**1.0626**	**1.0758**	1.0317	1.0847	1.1155	1.0670	1.0714	1.0758	1.0670	1.0714
Run 3	**1.0317**	1.0494	1.0538	1.0714	1.0626	1.0538	1.0450	1.0406	1.0494	1.0362
Run 4	1.0494	1.0362	1.0626	1.0670	1.0670	1.0758	1.0802	**1.0626**	**1.0141**	1.0626
Run 5	**1.0670**	**1.0582**	1.1067	1.0670	**1.0802**	1.0009	1.0450	1.0670	1.0450	1.0538
Run 6	1.0362	1.0714	1.0847	**1.0847**	1.0758	**1.0758**	**1.0626**	1.1023	**1.0758**	1.0273
Run 7	**1.1332**	1.0450	1.1243	1.1023	1.0891	1.0891	1.1243	1.0714	1.0802	1.1067
Run 8	1.1023	1.1155	**1.1199**	1.1332	1.0979	1.1596	1.1023	1.1243	1.1111	1.1023
Run 9	1.0979	1.0670	1.0494	**1.0626**	1.0802	1.0802	1.1023	1.0362	1.0538	1.0935
Run 10	1.1067	1.0847	1.0714	1.1067	1.0847	1.0714	1.0935	1.1023	1.0714	1.0979

Table 5.26: Mean error of the combination model generated by GACM on training data.

Run no.	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Run 1	0.2249	0.4541	0.4718	0.4630	0.4630	0.4674	0.4674	0.4850	0.4718	0.2249
Run 2	**1.0714**	**1.0891**	1.0317	1.0847	1.1155	1.0670	1.0714	1.0758	1.0670	1.0714
Run 3	**1.0406**	1.0494	1.0538	1.0714	1.0626	1.0538	1.0450	1.0406	1.0494	1.0362
Run 4	1.0494	1.0362	1.0626	1.0670	1.0670	1.0758	1.0802	**1.0847**	**1.0229**	1.0626
Run 5	**1.1067**	**1.0979**	1.1067	1.0670	**1.1023**	1.0009	1.0450	1.0670	1.0450	1.0538
Run 6	1.0362	1.0714	1.0847	**1.0891**	1.0802	**1.0847**	**1.0802**	1.1023	**1.0935**	1.0273
Run 7	**1.1376**	1.0450	1.1243	1.1023	1.0891	1.0891	1.1243	1.0714	1.0802	1.1067
Run 8	1.1023	1.1155	**1.1464**	1.1332	1.0979	1.1596	1.1023	1.1243	1.1111	1.1023
Run 9	1.0979	1.0670	1.0494	**1.0847**	1.0802	1.0802	1.1023	1.0362	1.0538	1.0935
Run 10	1.1067	1.0847	1.0714	1.1067	1.0847	1.0714	1.0935	1.1023	1.0714	1.0979

To evaluate the proposed GACM performance, the results are compared with the performance of the models selected manually. For manual selection, the combination models are selected based on the combination which produces the lowest error in training stage. These combination models are then evaluated using testing data. The results are shown in Table 5.27. The combination models generated by GACM in training stage are also evaluated using testing data and the results are presented in Table 5.28. The mean error of using only one classifier i.e. MLP, SVM or RBF with 16 features or 24 features are presented in Tables 5.29, 5.30, 5.31, 5.32, 5.33, and 5.34, respectively. The mean error over 10 folds and 10 runs for manual selection, GACM, MLP16, MLP24, SVM16, SVM24, RBF16, RBF24 are 1.0380% ± 0.3741%, 1.0452% ± 0.3710%, 5.0444% ± 0.4576%, 2.1798% ± 0.5146%, 1.4417% ± 0.4728%, 1.2663% ± 0.4832%, 3.2238% ± 0.5827%, and 3.0813% ± 0.5803%, respectively. From the results, it can be seen that using only 1 classifier cannot achieve high classification result. MLP with 16 features perform the worst. On comparison, using GACM reduces the error from MLP16 by 79% and from SVM24 by 18%. On average, the error difference between manual selection and GACM is 3%. The results also show that GACM has difficulty in finding the optimum models in fold 1, 2, and 4.

Table 5.27: Manual selection: Mean error of the combination model on testing data. The models are obtained manually based on training data.

Run no.	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Run 1	2.2222	0.1587	0.0000	0.0794	0.0794	0.0397	0.0397	0.0000	0.0000	2.1429
Run 2	1.0714	1.0714	1.4683	0.9921	0.7143	1.1508	1.1111	1.0714	1.1508	1.1111
Run 3	1.3889	1.1905	1.0714	0.9524	0.9921	1.1111	1.1508	1.1111	1.0714	1.2698
Run 4	1.1111	1.2698	1.0317	0.9921	0.9921	0.9127	0.8730	1.0317	1.4683	1.0714
Run 5	1.1508	1.0714	0.6349	0.9921	0.8730	1.5873	1.1905	0.9921	1.1905	1.2302
Run 6	1.4683	1.1508	0.9524	0.9524	1.0317	1.0317	1.1508	0.7937	1.0317	1.4683
Run 7	0.9524	1.5079	0.7937	0.9921	1.1111	1.1111	0.7937	1.2698	1.1905	0.9524
Run 8	1.3095	1.1508	1.1111	1.0317	1.3095	0.7540	1.2698	0.9921	1.1905	1.3095
Run 9	0.9127	1.1111	1.2698	1.1508	0.9921	0.9921	0.7540	1.4683	1.2302	0.8730
Run 10	1.0714	1.1905	1.2698	0.9524	1.1508	1.2698	1.0714	0.9921	1.2698	1.0317

Table 5.28: GACM: Mean error of the combination model on testing data. The models are generated by GA based on training data.

Run no.	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Run 1	2.2222	0.1587	0.0000	0.0794	0.0794	0.0397	0.0397	0.0000	0.0000	2.1429
Run 2	1.1905	1.0317	1.4683	0.9921	0.7143	1.1508	1.1905	1.0714	1.1508	1.1111
Run 3	1.3492	1.1905	1.0714	0.9921	0.9921	1.1111	1.1508	1.1111	1.0714	1.2698
Run 4	1.1111	1.2698	1.0317	0.9921	0.9921	1.0317	0.8730	0.9524	1.5079	1.0714
Run 5	1.1508	1.1508	0.6349	0.9921	0.8333	1.5873	1.1905	0.9921	1.1905	1.2302
Run 6	1.4683	1.1508	0.9524	0.9921	1.0714	1.0317	1.3889	0.7937	0.9524	1.4683
Run 7	1.0317	1.5079	0.7937	0.9921	1.1111	1.1111	0.7937	1.2698	1.1905	0.9524
Run 8	1.3095	1.1508	1.4683	1.0317	1.3095	0.7540	1.2698	0.9921	1.1905	1.3095
Run 9	1.1067	1.0847	1.0714	1.1067	1.0847	1.0714	1.0935	1.1023	1.0714	1.0979
Run 10	1.1067	1.0847	1.0714	1.1067	1.0847	1.0714	1.0935	1.1023	1.0714	1.0979

The normality test is applied to the data and the results show that the data is not normal distribution ($p < 0.05$). The Wilcoxon Signed Ranks test is applied to the data to test if the difference in performances between using manual selection, GACM, and single classifier are significant. The statistical tests show that the differences in classification error between manual selection and GACM are not significant ($p \geq 0.05$). The differences in classification error between GACM

Table 5.29: Mean error of using only MLP with 16 features on testing data

Run no.	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Run 1	5.5952	5.1984	4.8016	4.9206	4.0079	5.0000	4.0079	4.0873	3.9683	5.5556
Run 2	4.3651	4.9603	5.2778	4.8016	4.6032	4.6825	5.5556	4.4444	5.6349	4.3651
Run 3	5.5159	5.0794	5.2381	4.4444	5.5556	5.0794	4.4444	5.1587	4.8810	5.4365
Run 4	4.6825	5.3571	4.5635	5.1587	5.0000	4.8810	5.5159	5.3175	5.6349	4.6825
Run 5	4.9603	4.6825	5.3571	4.7222	5.1984	5.1190	5.1587	5.2778	5.5952	4.9206
Run 6	5.9127	5.0000	5.1190	5.3175	5.3571	5.3968	5.2778	4.0476	5.1190	5.8730
Run 7	4.9603	5.5159	4.8810	5.4762	5.6746	4.4444	4.9206	4.7222	4.8413	4.9206
Run 8	5.3571	4.7222	4.8016	5.5159	5.8730	4.9603	4.8016	5.0000	4.8016	5.3175
Run 9	5.1984	4.0079	5.9921	5.7540	5.3968	5.3571	5.4365	5.0000	4.2063	5.1190
Run 10	4.8413	5.4762	5.1587	4.7619	4.5238	5.1587	5.2381	5.2778	5.3968	4.7619

Table 5.30: Mean error of using only MLP with 24 features on testing data

Run no.	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Run 1	3.8095	1.1905	0.9921	1.1905	1.0317	1.5079	0.7143	0.7540	0.5556	3.7698
Run 2	2.1429	2.4206	2.8175	2.2222	2.0238	1.6270	2.4206	2.1032	2.1825	2.1429
Run 3	2.8571	2.4206	2.6587	2.2619	1.8254	2.4603	2.1032	2.2619	1.6667	2.7778
Run 4	1.9841	2.5397	1.7857	1.8254	2.3413	2.0238	2.4603	2.1429	2.9762	1.9444
Run 5	2.5000	1.7857	2.0238	2.3016	1.9048	2.3413	2.6587	2.3016	2.3413	2.4603
Run 6	2.4603	2.5397	1.8254	2.3016	2.5000	2.3413	2.0238	1.7063	2.5000	2.4206
Run 7	1.9444	2.3413	1.8254	2.0238	2.6190	2.2619	2.1032	2.4603	2.4206	1.9048
Run 8	2.4603	2.1032	2.5000	2.5397	2.3413	1.5476	2.3016	2.4206	1.9841	2.4206
Run 9	2.5397	2.3413	2.5794	2.2619	2.0238	1.6667	2.0238	2.7381	2.3413	2.4603
Run 10	2.0238	2.0635	2.2619	2.1825	2.7381	2.3016	1.7063	2.2222	2.8571	1.9444

Table 5.31: Mean error of using only SVM with 16 features on testing data

Run no.	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Run 1	2.9365	0.1984	0.0794	0.0397	0.3571	0.1984	0.1587	0.1587	0.1190	2.8968
Run 2	1.5873	1.5476	1.5873	1.3889	0.8730	1.2302	1.6270	1.6667	1.3492	1.5873
Run 3	1.6667	1.3492	1.4286	1.1508	1.5476	1.7460	1.4286	1.7857	1.4683	1.5873
Run 4	1.5476	1.4286	1.5079	1.3889	1.3889	1.4286	1.3095	1.3889	1.9048	1.5079
Run 5	1.7063	1.3492	1.3492	1.5079	1.3492	1.7460	1.3492	1.7460	1.7460	1.6667
Run 6	1.8254	1.5476	1.2698	1.5476	1.2698	1.8254	2.0238	1.1111	1.6667	1.7857
Run 7	1.4683	1.8254	1.1111	1.5476	1.6667	1.7460	1.2302	1.7063	1.5873	1.4286
Run 8	1.7460	1.5873	1.4683	1.1905	1.5476	1.7857	1.6270	1.2698	1.6270	1.7063
Run 9	1.2302	1.3492	1.8651	1.6270	1.1905	1.5476	1.5873	1.9048	1.3889	1.1508
Run 10	1.5873	1.5873	1.6667	1.2698	1.5079	1.8254	1.6270	1.3492	1.5873	1.5079

Table 5.32: Mean error of using only SVM with 24 features on testing data

Run no.	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Run 1	3.0159	0.0000	0.0397	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	2.9762
Run 2	1.3492	1.2698	1.5476	1.1508	1.0714	1.4286	1.3889	1.2302	1.4286	1.3492
Run 3	1.5873	1.1905	1.3492	1.0317	1.2698	1.4286	1.3889	1.7857	1.2302	1.5079
Run 4	1.5476	1.5079	1.0714	1.1905	1.4683	1.1508	1.0714	1.4683	1.7063	1.5079
Run 5	1.5476	1.1905	0.9921	1.1905	1.0317	1.7857	1.1508	1.3889	1.5079	1.5079
Run 6	1.7460	1.1905	1.2302	1.2698	1.4286	1.1111	1.5079	1.3095	1.3095	1.7063
Run 7	1.2302	1.5079	0.9524	1.3095	1.3095	1.6667	0.9524	1.4683	1.3095	1.1905
Run 8	1.3095	1.1905	1.3492	1.1905	1.6270	1.4286	1.4286	1.1111	1.5476	1.2698
Run 9	1.2698	1.3095	1.3492	1.4286	0.8730	1.6270	1.0714	1.4286	1.3095	1.1905
Run 10	1.3889	1.5873	1.4286	0.9524	1.6667	1.7460	1.2698	0.9524	1.3095	1.3095

Table 5.33: Mean error of using only RBF with 16 features on testing data

Run no.	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Run 1	4.8413	2.3413	1.7857	2.1825	1.5476	1.9048	1.8254	1.9444	1.7857	4.8016
Run 2	3.2143	3.7302	3.2143	3.0556	2.4603	3.4524	3.2540	2.6190	3.5317	3.2143
Run 3	3.9286	3.6111	3.4524	2.8571	3.4921	3.4921	3.5317	3.0556	3.0952	3.8492
Run 4	3.0952	3.0159	3.0159	3.0952	3.3333	3.0159	3.2937	3.2143	4.2460	3.0556
Run 5	3.1746	3.4127	3.6508	3.2937	2.6190	3.6905	3.4127	2.8968	3.3730	3.1349
Run 6	4.0079	3.0952	2.8175	3.3333	2.8571	3.6111	3.8492	2.4603	3.6905	3.9683
Run 7	3.4524	3.8492	2.8571	3.4524	3.2937	3.5317	2.4206	3.9286	3.4524	3.4127
Run 8	3.8492	2.7778	3.6111	3.0952	3.3730	3.7698	3.0159	3.1746	3.0952	3.8095
Run 9	2.9365	2.9365	4.0873	3.8095	3.2540	3.0159	3.2937	3.4524	3.2143	2.8571
Run 10	2.9762	3.4127	3.4524	2.8968	2.7381	4.0476	3.2937	3.1349	3.6508	2.8968

Table 5.34: Mean error of using only RBF with 24 features on testing data

Run no.	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Run 1	4.9206	1.8651	2.3810	2.2222	1.7063	1.4683	1.8651	1.7063	1.4683	4.8810
Run 2	3.0556	3.1746	3.5317	2.6190	2.4603	2.9762	3.1746	2.5794	3.4127	3.0556
Run 3	4.1270	3.3333	2.8968	2.7778	2.9365	3.2937	3.3730	2.6984	3.1349	4.0476
Run 4	3.0952	2.8968	2.9365	3.3333	3.0159	2.8571	3.0159	2.8571	3.6905	3.0952
Run 5	3.0159	2.9762	3.4127	3.2143	2.6190	3.9683	3.1349	3.0952	3.5714	2.9762
Run 6	4.4841	2.8968	2.6984	2.9762	3.0159	3.1746	3.4524	2.4206	3.1746	4.4444
Run 7	3.0556	3.3730	2.6984	3.0952	2.9365	3.5714	2.6190	3.3333	3.2143	3.0159
Run 8	3.0556	3.0159	3.2540	2.6587	3.2143	3.3730	3.3333	3.4524	3.2540	3.0159
Run 9	3.1349	2.8175	3.8095	3.3730	3.0952	3.2540	3.4524	3.0952	2.6984	3.0556
Run 10	2.8175	3.2540	3.5317	2.7381	2.8968	3.2937	3.4921	3.0159	3.3730	2.7381

and MLP16, MLP24, SVM16, SVM24, RBF16, RBF24 are significant ($p < 0.05$). The performances of the algorithms can be expressed as $Manual > GACM >^{*} SVM24$.

5.3.5 Discussion

The results based on training data show that in general GACM can find the combination model that produces the minimum error. The results show that only 15% of times that GACM are unable to find the optimum model. This is because the number of elite chromosomes that is set in the experiment maybe too high. In the experiment, the elite chromosomes contributed to 50% of the whole population. This leaves 4 chromosomes for crossover and 1 for mutation which may limit GA in exploring a new solution region. In the experiments, only 10 chromosomes (per population) were used. This number could be increased so that the number of elite, crossover, and mutation can be adjusted such that other solution region can be explored. These GA parameters should be selected based on experiment. The results also show that the error are mainly from fold 1. This can be explained by that the data used to train GACM did not represent the data in fold 1. This implies that GACM performance is affected by the training data. Therefore, training data set should be carefully selected to ensure GACM performance. The evaluation of the proposed method on the test data show that the combination models generated by GACM produces lower error than using only one classifier. The results of the evaluation on the test data show that the error obtained from GACM is 3% higher than manual selection which is not statistically significant. This shows that in general GACM can find the optimum combination model automatically which reduces the time and effort in manual search. This is particularly useful when the search space become larger i.e. more classifiers, weight functions, or combiners.

The results between combination models and one classifier indicate that using combination model improves classification accuracy significantly. The combination model increases the classification accuracy between 0.2211% and 3.9992%. The combination model can be auto generated from GACM. The computational time in running the proposed algorithm is 20 minutes per fold where 1 fold contain 22,680 data. The proposed algorithm can be used to generate the combination model offline. The classifier combination using the model can be executed online as the weight functions and combiners are low computational cost. Other criteria for model selection can be added to suit different application. For example, model with less number of classifier, particular weight functions or classifiers or combiners are preferred.

In previous method [103] which find the combination between features, classifiers, and combiners, a high computational cost is expected as new classifier need to be built every time a new chromosome is generated. Also, using this method, optimal parameters may not be able to obtain to use in classifier construction. GACM uses existing classifiers which are already optimised for the selected features. The study also compares the GACM performances with all possible combination to demonstrate that the combination models selected by the proposed method is the optimum ones.

5.3.6 Conclusion remarks

The results of the study demonstrate that the combination model between classifier, fusion weight, and classifier combiner improve the classification accuracy. The combination model can be generated automatically using the proposed algorithm, GACM. The results also show that the combination models obtained by GACM are as good as manually selection. Also, the proposed algorithm allows other criteria, besides the minimum error, to be added to suit different purposes.

5.4 Summary

This chapter presents extensive experimental studies on multi-sensor activity classification and classifier combination. Three classification methods i.e. MLP, RBF, SVM are studied in this research. The experimental results indicate that SVM is the most powerful classification. The results also show that using a set of 24 features generated from seven sensors achieves the best classification results. Nevertheless, the results indicate that different classification models are better at classifying different activities which leads to the next investigation on classifier combination. First, classifier combination using GA weights is investigated. The experimental results indicate using GA weights can achieve equal or higher accuracy comparing to one best classifier. In this chapter another classifier combination technique called GACM is proposed where GA is used to find the optimal combination between classifiers, weight functions, and combiners. The experimental results show that in general GACM can find the optimal combination models with the minimum classification error. The combination models automatically generated by GACM are as good as manual selection. For future research, the proposed algorithms should be tested on other data sets, or activity data set with other/larger activities. Another interesting study is to modify the GACM to be adaptive to data sets. The proposed multi-sensor AR framework can be applied in health care domain such as home-based monitoring and decision support system for health care organisations. These applications are discussed in the next chapter.

Chapter 6

Conclusion and Future works

The study is set out to investigate the use of wrist-worn multi-sensor for AR of an elderly person for assisted living applications. A number of ageing population has increased rapidly worldwide over the past decades. This has major effect on health care where issues such as a rise in care cost, high demand in long-term care, burden to carers, and insufficient and ineffective care are likely to occur. AR can be used as the key part of the intelligent systems to allow elderly people to live independently at homes, reduce care cost and burden to the carers, provide ensuring for the families, and promote better care. There are a number of AR systems available. However, a majority of works mainly consider the technical aspects of the system i.e. accuracy and neglects the practical aspects such as acceptance and usability. The practicality of the system is the key factor which indicates whether the system will be used in reality or not. This research aims to develop the AR system which considers both practical and technical aspects using non-intrusive, inexpensive wearable sensors so that the acceptance and usability are increased allowing the system to be used in reality.

Firstly, an application of the proposed multi-sensor AR framework in health care domain is presented. Applications in home-based monitoring and decision support system for health care organisations are discussed. Next, the chapter revisits the objectives set at the beginning of the research. Followed by, a synthesis of the empirical findings from the study with respect to the research questions is presented. Finally, the chapter finishes off with the research limitations and suggestions for future works.

6.1 Applications in health care

6.1.1 Home-based care application

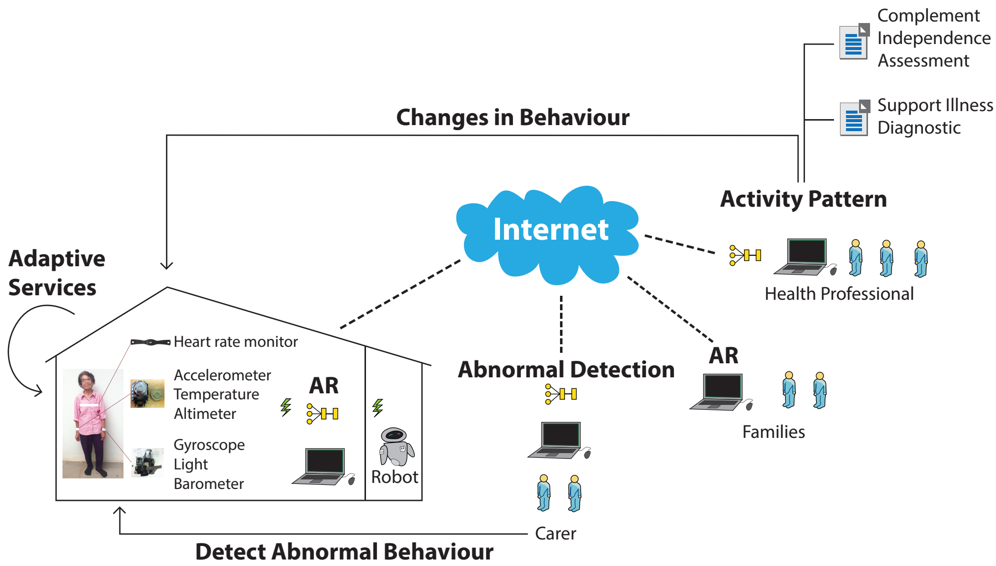

Figure 6.1: A practical multi-sensor activity recognition system for home-based care

This section presents a practical multi-sensor AR system shown in Figure 6.1 and describes how it can be used for home-based care. The elderly person wears sensors including accelerometer, temperature sensor, altimeter, gyroscope, light sensor, and barometer which are embedded on watch on their wrists and a heart rate monitor on their chests. The data from the sensor is continuously transmitted wirelessly through radio frequency to the computer in the elderly's home. The computer contains the AR application which can recognise and detect daily activities of a user. The detected activity is perceived wirelessly by a companion robot who provides assistances or services based on current activity. For example, if the robot detects that the elderly person is exercising, it can play music or video related to that exercise. If the house is equipped with smart sensors, the detected activities can be used to provide information for adaptive services. For example, if it is detected that a user is sleeping, the light and the temperature can be adjusted to the suitable condition.

The detected activities can also be used by carer, health professionals, and families. To protect the privacy of the elderly person, the system will not send the raw sensor data over the network. The detected activities are encrypted when sent over the Internet. For carers, their systems will contain an activity abnormal detection model to detect abnormality of the elderly person. When the abnormal activity is detected, a carer can visit the elderly home and provide help. This will allow independence for both elderly person and carer, while maintain safety and good care when necessary. The families of the elderly person will also benefit from the system where they can use it to monitor them online anywhere and anytime to provide a peace of mind that their love ones are doing well. Health professionals will have access to the activity records. Their systems will contain a model which interpret each activity into activity patterns. They can use this as a complement to normal independent assessment and to support illness diagnostic. Also, if they detect any changes in behaviour, they could send a request to elderly person's system to retrieve a raw sensor data for further analysis or arrange a home or hospital visit for a check up on the elderly person.

Any sensor data sent from the elderly person must be encrypted and authorisation system must be installed and used whenever someone requests to access the data. Also, there must be a signed agreement on who can have access to what information and the elderly must give their consent prior the use of the system to ensure privacy and visibility.

6.1.2 Decision support system for health care

This section describes how the proposed multi-sensor AR can be used to enhance the decision support systems (DSS) for health care. First, the architecture of the DSS is described, then followed by examples how the proposed method can be used in DSS and improve health care. Figure 6.2 shows the design of the DDS. The proposed method is used for classifying the complex sensor data into activities to generate a database of activity records over times. The data management is used for manage databases from several sources. The operations that the data management carry out includes organise, search, query, add, update, and delete databases. It also connects to the user interface management to provide interface for the users to perform operations with the databases. Besides the activity database, other databases related to health care information such as medical records, hospital resources, carer records, independence assessments, etc. are connected with the data management so that the DDS can cooperate several sources to make reliable decisions.

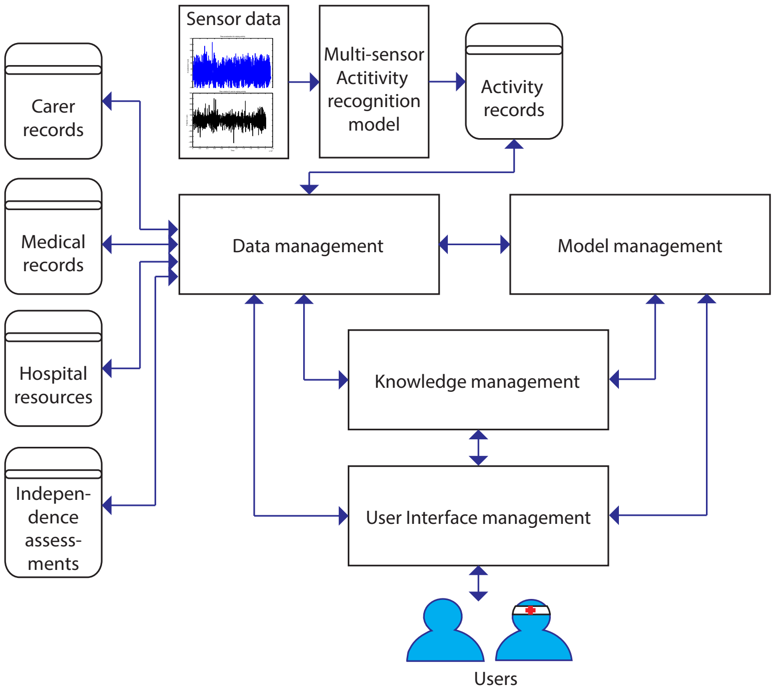

Figure 6.2: Decision support system for health care

The model management (MM) is used to manage models, select suitable models for different problems, execute the model, combine results from models. MM is connected to data management and user interface management (UI) to retrieve input data and to present outputs. The models are used to predict, simulate, schedule, classify, etc. input information. Example models are such as a model to predict decline in daily activities, schedule the carer timetable, classify independence level, simulate the use of beds in hospital, etc.

In some cases, the information from several databases may not be enough to make a decision. Especially in health care, when experiences or expertise may be needed to make critical decisions. Therefore, the DDS contains the knowledge management (KM) which is used to store the knowledge resulting from the decision made by experts. The knowledge includes the process and/or information required to make decision by experts. KM consists of several subsystems required to build the knowledge base. This includes knowledge acquisition, representation, validation, inference, and explanation or justification of the knowledge.

The DDS also contains the UI to manage different terminals for users to interact with the DDS. The UI includes several interfaces such as text, graph, web page, etc. suitable for different tasks and user groups. For example, the interface for management staffs should present overall result with graphical formats, while the interface for operational staff should present the information of a particular task in details. High usability is crucial aspect of the acceptance of DDS.

The DDS can be used to generate a monthly activity graph which shows the amount of each activity carried out in different months. This can be used to see the trend and detect changes in activities and support the decision whether to contact the person to come to the hospital and to which department or a home visit or whether further activity data should be requested from the patient. For example, if the graph shows the decline in walking over several months, this could suggest there is a problem with ambulating. This would help reduce the number of hospital visits, improve hospital resources utilisation, and increase earlier detection rate.

The DDS can be used to support the decision on the type of carer is required for different patients. For example, if an activity record shows no decline or changes in activity pattern, carer may not be needed. If the activity record suggests the person may have problem with feeding, the carer who can provide assistance with feeding or cooking should be sent. Also, based on activity database, the DDS can build a model to predict when it is likely that the person will need a carer, so that the management of carer e.g. schedule, number of carer, etc. can be done effectively.

The activity record can be used as a complement tool for the assessment of independence. The DDS can use this to make a decision whether the carer is needed and to predict when the carer will be needed in order to manage resources effectively. The activity database can be used as part of the other clinical decision support system to give more information to support the illness diagnostic or disease symptom. For example, if the activity record shows the patient has very little sleep per day could influent the decision of the specific sleeping disorder.

6.2 Objective revisits

The main aim of this research is to develop a novel method for multi-sensor based AR of ADL of an elderly for an intelligent assisted living system. This section revisits the objectives set out at the beginning of the research and discusses how these objectives are met.

1. "To carry out literature reviews in wearable sensor based AR and its application in assisted living and to identify research gap"

   Extensive literature reviews have been carried out in wearable sensor based AR and its application in assisted living. These are presented in Chapter 2. The reviews start with the history of AR where different approaches are reviewed and discussed in term of their advantages and disadvantages. The rest of the literature reviews are focused on the AR based on wearable-sensor approach where topics such as sensor types, sensor locations, sensor fusion, as well as activity classification techniques including pre-processing, segmentation, feature calculation, feature space manipulation, classification techniques, and classifier combination are covered. A review on applications of AR in various domains is presented. In addition, the application of AR in assisted living is focused where topics such as activities of daily living, existing studies in AR for assisted living, and requirements of assisted living are reviewed. Based on these literature reviews, the research gaps in AR for assisted living are identified.

2. "To design and develop hardware for sensor data collection"

   Sensors and hardware platforms have been identified, designed and developed for data collection purpose. The details of the sensors, platforms, implementation, justification for sensor location and choices of activities are presented in Chapter 3.

3. "To collect sensor data in a real home setting"

Three activity data sets are collected. The first data set is collected from seven young participants performing running, walking, sitting, standing, and lying down. This data set is used for a feasibility study. The second data set is collected from 12 elderly people performing 12 ADL in a real home setting. This data set contains data from three sensors. The third data set is collected from 12 elderly people performing 13 ADL in home setting. This data set contains data from seven wearable sensors. The descriptions and characteristics of these data sets are presented in Chapter 3.

4. "To carry out a feasibility study on using a wrist worn sensor to detect activities and to identify features and techniques for data pre-processing and segmentation for multi-sensor based AR"

   A feasibility study has been carried out to evaluate the possibility of using the wrist-worn sensor for AR, and to identify features, pre-processing and segmentation techniques suitable for human AR. This study used Young Activity data set to evaluate the feasibility. According to the study, a set of features from time and frequency domain are identified. Also, a pre-processing technique i.e. Weighted Moving Average, and segmentation technique i.e. windowing using 128-window length and 50% overlapping are selected. The results of this study are presented in Section 4.1.

5. "To investigate and evaluate techniques for feature selection and to propose novel feature selection techniques for multi-sensor based AR"

   This research proposed two feature selections: Feature Combination (FC) and Maximal Relevancy Maximal Complementary (MRMC). These techniques are based on MLP and the concept of a relationship between a group of features and the outputs. The proposed algorithms are evaluated against three popular feature selection techniques including MRMR, NMIFS, and Clamping on multi-sensor AR data sets as well as benchmark data sets. The studies of the proposed algorithms are presented in Chapter 4.2.

6. "To investigate techniques for activity classification and to evaluate classification results generated from different techniques"

This research study investigated three classification algorithms i.e. MLP, RBF, and SVM for multi-sensor AR of an elderly person. The investigation is carried out using two multi-sensor activity data sets. The results of the study are reported in Section 5.1.

7. "To investigate and evaluate techniques for classifier fusion and to propose a novel classifier fusion technique based on Genetic Algorithm"

   This research carried out an investigation on several techniques for classifier fusion including six fusion weights i.e. simple average, variance-covariance, discounted mean square forecast error, weighted accuracy, and unit weight, and seven fusion methods i.e. majority voting, product, summation, maximum, minimum, weighted average, and ranking. Also, two classifier fusion techniques based on GA are proposed. The first technique uses GA to find optimum weights for classifier fusion called Genetic Algorithm based Fusion Weight (GAFW). Unlike previous works, the evaluation of the technique is carried out on all possible classifier combinations. Also, different fitness functions of GA are investigated. The results indicate that in general, using GAFW can achieve at least equal or higher than using only one best classifier. The results of this study is presented in Chapter 5. Another classifier fusion technique is proposed where GA is used to find the optimum combination between classifiers, fusion weight, and classifier combination functions which is called Genetic Algorithm based Combination Model (GACM). Also, the proposed technique allows other model selection criteria to be added. For example, a combination which uses less number of classifiers is preferred. An investigation is carried out using 6 classifiers, 4 fusion weight functions, and 5 combiner functions. The results indicate that in general GACM can find the optimum combination automatically. The results of the investigation are presented in Sections 5.2 and 5.3.

8. "To investigate the contributions of sensors for AR"

   A study on the contributions of sensors for AR is carried out. Two techniques i.e. Mutual Information (MI) and Clamping are used to calculate the contributions of the sensors. MI is used to calculate the contribution of the sensor for the activity classification, while Clamping is used to calculate the contribution of the sensor within the classification model. The results of the investigation are presented in Section 4.4.

9. "To discuss the application of the proposed multi-sensor AR in assisted living"

   The results of each component of the proposed multi-sensor AR have been evaluated throughout the thesis using the collected data set and against other AR studies in term of technical aspect i.e. accuracy. In Section 6.1, the evaluation of the proposed multi-sensor AR in term of practical aspect against other AR works is presented. In addition, a discussion on how the proposed work can be used in the assisted living is presented.

6.3 Empirical findings

The main research problem is to recognise activities of daily living of an elderly person using multi-sensor worn on wrist. This section summarises the findings regarding the research questions.

1. How to detect the interested activities of an elderly person using multiple wearable sensors worn on wrist?

   The multi-sensor AR of an elderly person has been proposed in this research. The sensor fusion process is performed at feature and classifier levels. The proposed method uses six sensors worn on wrist including accelerometer, gyroscope, temperature sensor, altimeter, barometer, and light sensor. The proposed model receives the sensor data where they are pre-processed using weight moving average and segmented at 3.88 seconds. In training stage, the method calculates several features and performs feature selection using one of the proposed feature selection algorithms. Next, the classifiers are built using the selected features with MLP, RBF, and SVM. The combination model between classifiers, fusion weight functions, and combiners is obtained using the proposed GACM. In online stage, selected features are calculated from the sensor data and then passed to the classifiers. The prediction is the activity with the maximum probability obtained from the combination model. The explicit detail of the proposed multi-sensor AR is presented in 3.6. Some parts of the results have been published in [1, 3, 8, 9].

2. Does using multiple sensor improve classification accuracy? Does the heart rate sensor help increase the classification accuracy of the wrist-worn sensor based AR?
   The results of the study in Chapter 5.1 indicate that using multiple sensors increase classification accuracy. Accelerometer is the most important sensor in wearable AR. It is found that sensors that are not useful on its own may be useful when combine with other useful sensors. The study shows that combining heart rate with other sensor significantly improves classification accuracy. Nevertheless, the classification accuracy without using heart rate is still high which suggests that it is possible to use only wrist worn sensors to maintain its practicality while high accuracy can be achieved. The results of this study have been published in [1].

3. How to select the features using the relationship between feature and classes as well as the relationship between a group of features and classes?
   This research proposes two feature selection techniques: FC, and MRMC. Both of these techniques consider the relationship between feature and classes and the relationship between a group of features and classes. FC uses Clamping technique to calculate the relationship between a feature and outputs. It then use the modified forward selection technique to measure the relationship a group of features and classes. The investigation of this technique is reported in Chapter 4. The findings indicate that FC performance is better than popular techniques including MRMR, NMIFS, and Clamping at 95% confidence interval. It is found that the evaluation between a group of features and classes along the selection help to make sure redundant features are not selected. MRMR, NMIFS, and Clamping only measure the redundancy between 2 variables which is shown not enough to reduce the overlapped features. However, FC has two limitations. Firstly, it is possible that redundant features may be selected at earlier stage of selection. Secondly, good features may be eliminated in early stage due to

the use of forward selection.

To overcome these limitations, MRMC is proposed. MRMC is based on the criteria of maximum relevance and maximum complementary of the feature. MLP is used to calculate the relevancy and complementary score of the feature. The feature with maximum score is then selected. The study of MRMC is presented in Chapter 4.3. The findings indicate that in general MRMC provide a good result comparing to the MRMC, NMIFS, and Clamping. The algorithm is capable of detecting completely overlapped and partial overlapped features. In addition, it is found that the proposed complementary criteria improve the performance of Clamping.

MRMC has limitation such that it cannot guarantee optimum result when apply to a small feature set. This is due to the selection of the first feature obtained from Clamping. However, when the number of features is increased, MRMC is demonstrated to be superior to the other three algorithms. This is because although the first feature selected by Clamping algorithm may not always be the most important but it is somewhat important and the optimum feature set can still be obtained by use of complementary score. The results of this study have been published in [3].

4. How to combine classifiers by utilising class probabilities and are generalise enough to be apply in other data set?

   Two strategies based on GA are proposed for classifier combination. The first strategy used GA to find the optimum fusion weight for classifier combination called GAFW. An investigation of this technique is reported in Chapter 5.2. The results indicate that in most cases the combination using GAFW can achieve equal or better accuracy than using just one best classifier. GAFW based on linear fitness function yield better performance than combiner function-specific fitness function, especially when minimum or maximum combiner is used. However when ranking combiner is used, GAFW based on ranking fitness function gives better results than GAFW based on linear fitness function. This is because the class probability has been converted into ranks which are different data representation than that used in linear function. It is also found that GAFW should be appropriate for the AR model that will be developed offline due to high computational cost. For online system, it is recommended that other weight functions such as Variance-Covariance and Weighted Accuracy should be used.

Due to the limitation from high computational cost, another classifier combination method is proposed. This technique uses simple fusion weight and combiner functions. GA is employed to find the optimum combination between classifier, fusion weight and combiner functions. The study of this technique is presented in Chapter 5.3. The results indicate that in general GACM can find the optimum combinations automatically. The study against manual selection among 1,146 combinations reveals that there is no statistical significant in the performances of GACM and manually selection. In addition, GACM allows other criteria for model selection to be added e.g. simpler model is preferred. Some parts of the results have been published in [4].

6.4 Research limitations

This section discusses the research limitations. First, the limitations are identified, and their impacts on research results are discussed. Next, a reflection on the limitations and the justifications of the choices made during the research process. Finally, the suggestions on how these limitations can be overcome in the future are discussed.

1. Data not collected under natural setting
   This research collected the data from a group of elderly people in a real home. However, the protocol used in the data collection process is controlled. For example, the participants are asked to perform different activities for a period of time. This may prevent the participants to carry out the activities as continuous and natural as possible. Therefore, the reduction in classification performance of the developed AR model when used in reality is expected.

   The reason that controlled protocol is used in this research is to reduce the complexity of the data collection. For example, an experimental set up is required for each participants home. Data annotation in natural setting is more complex and video camera may be required to assist the process. Also, longer hours of data collection are required in order to collect enough data on specific activities which have lower probability of occurrences. For example, the elderly person may use stairs only a few times a day. The research used a controlled protocol which has been used widely in AR studies. However, the following strategies are used to minimise the control effect in order to encourage the participants to perform the activities as natural as possible:

Flexibility: The participants are allowed to perform the activities in any order and they could have breaks in between the activities.

Minimal supervision: The participants are left to perform the activities on their own without direct supervision. The researcher is only at the start and the end of each activity to set up the equipment and marked down activity information. This helps reduce the participants anxiety and encourage them to carry out activities as natural as possible.

Mimic home setting: Unlike other studies which collect the data in a laboratory, the data collection in this research is carried out in a real home. The purpose is to mimic the home setting and environment, therefore help the participants to perform activities more natural.

To overcome this limitation, it is suggested that the data collection should be carried out in a natural setting i.e. in a real home of the participant or an instrumented house that allow natural behaviour without using controlled protocol. In addition, the AR should be tested in real application or on data set collected under natural setting.

2. A limited skills in electronics

   This research requires multi-disciplinary skills including computing, and electronics as the hardware need to be designed and developed for the data collection purpose. However, the researcher has limited skills in electronics and also due to time restriction, it is not possible to develop a new built-for-purpose hardware to collect the sensor data. This affects the process of data collection such that all sensors are not presented in one platform. Also, this makes it difficult to envisage the real use of the proposed work in real environment.

Due to this limitation, ready sensor platforms or ones which are easy to implement are selected. The EZ-430 watch is selected as it contains multi-sensor on-board. Also, it allows programming and debugging of the software on the watch. This allows the researcher to modify the program on the watch to suit different purposes of the data collection. It can also be paired with a heart rate monitor which is one of the sensors used in the research. In addition, a Microsoft Gadgeteer platform is used. It is an open-source toolkit for building small electronic devices based on the .NET Micro Framework and Visual Studio/Visual C# Express. There are several sensors for Gadgeteer available in the market.

Since the sensors are implemented on different platforms, it is decided to separate the sensors between two wrists. The following strategies are used to reduce the effect cause by separating the sensors over two wrists. The sensors are separated in a way that it should not interfere with the AR. The sensors which are related to the movement i.e. accelerometer and gyroscope are worn on the dominant wrist in order to capture the activitys movement. Also, barometer and light sensors are also worn on the dominant wrist as they are parts of the Gadgeteer platform. The temperature sensor which captures the body temperature and altimeter are worn on the non-dominant wrist. In real application, we are expected to implement all the sensors into a single watch and will be worn on the dominant wrist of the elderly person. This location will not disrupt a user from performing an activity and/or cause discomfort in wearing sensors.

To overcome this limitation for future research, it is suggested that the sensor platforms which is easy to implement, has a variety ranges of sensors are considered. The sensor hardware should be developed as soon as possible at the beginning stage of the research. Also, being able to identify the lacks of skills in hardware in early stage of the study will lead to better problem management e.g. whether to employ other people to develop the hardware, take relevant courses, use ready hardware, etc.

3. Low number of participants

   The number of ageing population (people aged 65 years and over) in the world is 520 million in 2010. Using this population with 95% confidence level and 5% confidence interval, the sample size of 385 is required in order to build a generalised classification model. This limitation affects the studys results such that the model may not be generalised enough for the world population.

   One of the main challenges in activity research is the data collection. It is difficult to recruit a large number of people to participate in the data collection. There are several challenges in recruiting participants for the data collection. Here, only the problems related to this research are discussed. Firstly, it requires the participant to participate for a rather long period of data collection in order to collect enough data for all interested activities. This includes time to setup equipment, performing activities, breaks between activities for the participants, and transferred data. Also, enough data collection tools are required. If the number of equipment is not sufficient, the process of data collection could be slow and extended time will be required. In this research, one set of equipment is used. Also, the data annotation is done manually which is a time-consuming process. Secondly, it is difficult or in some cases not possible to use the available public data sets. This is due to the variability between sensors, equipment, participants, and activity descriptions. Thirdly, it is more difficult to recruit elderly people to participate in the data collection. The number of elderly people is generally lower than the young people. In addition, since this research is related to physical activities, only healthy elderly people are eligible.

   Table 2.6 shows the characteristics of the participants in other studies. Based on this table, the number of participant varies from 1 to 60 people. The mean number of participants is 10.97 and median is 8 people. Figure 6.3 shows the histogram of the number of participants which illustrates that the

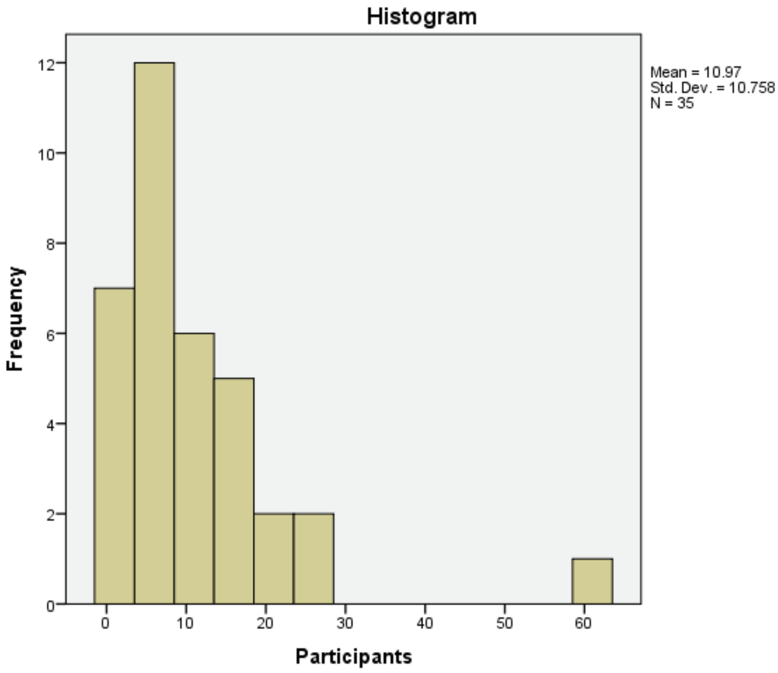

Figure 6.3: A histogram of number of participants in AR studies.

number of participants is skewed toward the lower number of participants. In addition, only 7% of the studies collected the data from elderly people. Among these studies, a small number of elderly people is recruited. For example, [145] used eight elderly people in their study. [83] work is based on two elderly persons. A larger number of elderly people is recruited in [101] study. However, mainly posture and transition activities are studied in that work. In this research, each activity data set is based on 12 elderly people, containing both males and females. This number is slightly larger than the average number of participants in previous AR studies.

Collaborations between the research project and nursing homes, hospitals, or any organisations related with elderly people could attract more number of participants. To overcome this limitation in future research, it is suggested that, if possible, such collaborations are established prior research commences.

6.5 Future works

In this research, a multi-sensor AR of an elderly person has been studied. Future research may be conducted to overcome the limitations discussed in Section 6.4. Also, this research could be extended into various new research directions.

1. Extension of the MRMC
   Maximal Relevancy Maximal Complementary feature selection has been proposed in the research. However, MRMCs performance is depended on the first feature when a number of feature set is low. MRMC uses Clamping to select the first feature and the results of the experiment show that the selected feature is not always the most relevant. Future research may be conducted to improve MRMC by focusing on the identification of the first feature in order to improve the performance of the algorithm. In addition, future work may consider the cases that there are more than one important features with equal scores. It is recommended that in order to correctly identify the first feature, the next important feature also need to be taken into account.

2. Abnormal activity detection algorithm

   The results of the proposed multi-sensor AR are activity records of an elderly person. Future research could be conducted to use these results to develop an abnormal activity detection algorithm. This algorithm can be used by the carers or the healthcare professionals to detect any changes in behaviour in both long-term and short-term care. Future study may focus on the discovery of activity pattern over times and outlier detection.

3. Hardware and battery

   Future research can be focused on the hardware aspects to improve the systems usability and acceptance. Note that, the system mentioned here refers to the equipment used for collecting sensor data. The system could be designed and developed so that all sensors are implemented on a single micro-controller board. Also, the future research could be focused on the extension of the batterys life of the system.

4. Data privacy and security

   One of the main issues in monitoring system is privacy and security. This problem is often concerned by the users when they consider adopting the system. The monitoring system is designed to be used by several users such as relatives, carers, and healthcare professionals who should have different access rights to the data. Future research can be conducted to focus on the data privacy and security. Research problems such as how/what/where the data should be stored, data encryption, authorisation, the design of the system architecture i.e. distributed or centralised could be investigated.

5. More activities

   The multi-sensor AR proposed in this research is based on 13 activities of daily living. Future research can extend the number of activities to cover more daily activities as well as fall. Suggested activities are shower, using toilet, using stairs, and dressing. These activities may be difficult to detect if only use wearable sensors. It is suggested that sensors which can be used to identify the location of the user such as motion sensor and RFID use in conjunction with wearable sensor to help with the classification. Note that the video sensor should be avoided. Location has shown to have a relationship with activity. It can be used to eliminate activities that are not possible at certain locations or increase probability of possible activities.

Appendix A

The Barthel index used to evaluate the participants and activity selection.

Barthel Index of Activities of Daily Living

Instructions: Choose the scoring point for the statement that most closely corresponds to the patient's current level of ability for each of the following 10 items. Record actual, not potential, functioning. Information can be obtained from the patient's self-report, from a separate party who is familiar with the patient's abilities (such as a relative), or from observation. Refer to the Guidelines section on the following page for detailed information on scoring and interpretation.

The Barthel Index

Bowels
0 = incontinent (or needs to be given enemata)
1 = occasional accident (once/week)
2 = continent
Patient's Score: ____________

Bladder
0 = incontinent, or catheterized and unable to manage
1 = occasional accident (max. once per 24 hours)
2 = continent (for over 7 days)
Patient's Score: ____________

Grooming
0 = needs help with personal care
1 = independent face/hair/teeth/shaving (implements provided)
Patient's Score: ____________

Toilet use
0 = dependent
1 = needs some help, but can do something alone
2 = independent (on and off, dressing, wiping)
Patient's Score: ____________

Feeding
0 = unable
1 = needs help cutting, spreading butter, etc.
2 = independent (food provided within reach)
Patient's Score: ____________

Transfer
0 = unable – no sitting balance
1 = major help (one or two people, physical), can sit
2 = minor help (verbal or physical)
3 = independent
Patient's Score: ____________

Mobility
0 = immobile
1 = wheelchair independent, including corners, etc.
2 = walks with help of one person (verbal or physical)
3 = independent (but may use any aid, e.g., stick)
Patient's Score: ____________

Dressing
0 = dependent
1 = needs help, but can do about half unaided
2 = independent (including buttons, zips, laces, etc.)
Patient's Score: ____________

Stairs
0 = unable
1 = needs help (verbal, physical, carrying aid)
2 = independent up and down
Patient's Score: ____________

Bathing
0 = dependent
1 = independent (or in shower)
Patient's Score: ____________

Total Score: ______________

(Collin et al., 1988)

Scoring:
Sum the patient's scores for each item. Total possible scores range from 0 – 20, with lower scores indicating increased disability. If used to measure improvement after rehabilitation, changes of more than two points in the total score reflect a probable genuine change, and change on one item from fully dependent to independent is also likely to be reliable.

Sources:

- Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: a reliability study. *Int Disabil Stud.* 1988;10(2):61-63.
- Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Md State Med J.* 1965;14:61-65.
- Wade DT, Collin C. The Barthel ADL Index: a standard measure of physical disability? *Int Disabil Stud.* 1988;10(2):64-67.

Guidelines for the Barthel Index of Activities of Daily Living

General
- The Index should be used as a record of what a patient ***does***, NOT as a record of what a patient ***could do***.
- The main aim is to establish degree of independence from any help, physical or verbal, however minor and for whatever reason.
- The need for supervision renders the patient not independent.
- A patient's performance should be established using the best available evidence. Asking the patient, friends/relatives, and nurses will be the usual source, but direct observation and common sense are also important. However, direct testing is not needed.
- Usually the performance over the preceding 24 – 48 hours is important, but occasionally longer periods will be relevant.
- Unconscious patients should score '0' throughout, even if not yet incontinent.
- Middle categories imply that the patient supplies over 50% of the effort.
- Use of aids to be independent is allowed.

Bowels (preceding week)
- If needs enema from nurse, then 'incontinent.'
- 'Occasional' = once a week.

Bladder (preceding week)
- 'Occasional' = less than once a day.
- A catheterized patient who can completely manage the catheter alone is registered as 'continent.'

Grooming (preceding 24 – 48 hours)
- Refers to personal hygiene: doing teeth, fitting false teeth, doing hair, shaving, washing face. Implements can be provided by helper.

Toilet use
- Should be able to reach toilet/commode, undress sufficiently, clean self, dress, and leave.
- 'With help' = can wipe self and do some other of above.

Feeding
- Able to eat any normal food (not only soft food). Food cooked and served by others, but not cut up.
- 'Help' = food cut up, patient feeds self.

Transfer
- From bed to chair and back.
- 'Dependent' = NO sitting balance (unable to sit); two people to lift.
- 'Major help' = one strong/skilled, or two normal people. Can sit up.
- 'Minor help' = one person easily, OR needs any supervision for safety.

Mobility
- Refers to mobility about house or ward, indoors. May use aid. If in wheelchair, must negotiate corners/doors unaided.
- 'Help' = by one untrained person, including supervision/moral support.

Dressing
- Should be able to select and put on all clothes, which may be adapted.
- 'Half' = help with buttons, zips, etc. (*check!*), but can put on some garments alone.

Stairs
- Must carry any walking aid used to be independent.

Bathing
- Usually the most difficult activity.
- Must get in and out unsupervised, and wash self.
- Independent in shower = 'independent' if unsupervised/unaided.

(Collin et al., 1988)

Appendix B

An example of an information sheet and consent form used in sensor data collection.

Multi-sensor activity recognition for an intelligent assisted living system

STAFFORDSHIRE UNIVERSITY

Introduction

Over the years, the number of ageing population has increased significantly. In 2050, 1.5 billion people will age 65 years and over. A new model of care that facilitates self-care and extends the independence of ageing population is required. Automatic recognition of daily activities allows activity monitoring, judging independence level, detect changes in behaviour over time which leads to intelligent assisted living system. The research proposes a solution in human activity recognition which overcomes privacy violation, cost and wearability issues by using non-obtrusive, non-intrusive, low-cost sensors.

The aim of this data collection is to collect sensory data (acceleration, temperature and altitude) from different activities in order to analyse and develop novel method for activity recognition of Activities of Daily Living of an elderly. The participants will be required to answer their personal information i.e. age, weight, height and personal illness and will be assessed on their independence level using Barthel Index. The participants will be required to wear watches (which have integrated sensors) on their wrists and perform 11 activities:

1) Walking
2) Sweeping floor
3) Watching television
4) Walking upstairs
5) Walking downstairs
6) Sleeping/Lie down
7) Dressing
8) Brushing teeth
9) Feeding
10) Washing dishes
11) Ironing shirts

Participants will be asked to perform these activities in their own pace and there will be short break between activities. For 'Feeding' activity, meal will be provided. The activities will be done in private area where no direct observation will be made. Also, the participants will be asked their opinion toward wearing the watches for monitoring activities.

If you have any additional questions about the research, your rights, or a research-related injury, you may contact: Miss Saisakul Chernbumroong, Staffordshire University, Faculty of Computing, Engineering and Technology, Stafford, ST180AD. Or by email at: s.chernbumroong@staffs.ac.uk

Privacy and Rights of Participant
1. Data collected from this study is solely for the purpose of investigation and analysis of activity recognition of Activities of Daily Living of an elderly. The statistical and qualitative data compiled will be used for research purposes, which will contribute to the knowledge of human activity recognition and intelligent assisted living system. 2. Data collection is collected in an anonymous and confidential manner. No personal details are required and hence individuals will be non-identifiable. An email address is required only if you wish to be informed about the findings of this study. 3. Your participation in this study is completely voluntary. You have the right to withdrawn from participation at any time. Such requests can be made in form of a verbal statement, written statement or an electronic mail clearly stating your wish to withdraw. There is no need to state a reason for withdrawal.

Participant Number *You will need to state this number if you later request to withdraw from participation*	**Email address: (OPTIONAL)** *Only needed if you wish to be informed about the findings of this study*

If you have any additional questions about the research, your rights, or a research-related injury, you may contact: Miss Saisakul Chernbumroong, Staffordshire University, Faculty of Computing, Engineering and Technology, Stafford, ST180AD. Or by email at: s.chernbumroong@staffs.ac.uk

Consent form

Place: ....

Date: ..../............../........................

Participant number: ....

I, who have signed at the end of this form, give consent to participate in the following study:

Research Title: ....Multi-sensor activity recognition for an intelligent assisted living system....

Researcher: ....Miss Saisakul Chernbumroong...

Contact address: ....80/4 Bumroongrat Road, Soi 2, Chiang Mai, Thailand, 50000..............

Telephone number: ....+6653246226..

I volunteer to participate in this study as a participant to answer personal information and sensory data collection which I am required to wear watches on my wrists and performing 11 activities as described in Information Sheet.

I understand that my participation is voluntary and that I have the right to withdrawn at any time without providing reasons and without my rights being affected.

I understand that my personal information will be looked at by researcher solely for the purpose of investigation and analysis of activity recognition of Activities of Daily Living of an elderly. My data will be kept anonymous and in confidential manner.

By singing on the consent below, I confirm that I have read the above information about this study, and that you understand the purpose of the study as well as the potential risks that are involved agree to participate in this study.

Signature of Participant: .... **Date:** ..../.........../.............

Signature of Researcher: .... **Date:** ..../.........../.............

Signature of Witness: .... **Date:** ..../.........../.............

If you have any additional questions about the research, your rights, or a research-related injury, you may contact: Miss Saisakul Chernbumroong, Staffordshire University, Faculty of Computing, Engineering and Technology, Stafford, ST180AD. Or by email at: s.chernbumroong@staffs.ac.uk

References

[1] Saisakul Chernbumroong, Shuang Cang, and Hongnian Yu. A practical multi-sensor activity recognition system for home-based care. *Decision Support Systems*, 66(0):61 – 70, 2014. ISSN 0167-9236. 104, 171, 223

[2] Saisakul Chernbumroong, Anthony S Atkins, and Hongnian Yu. Wrist-worn accelerometer based activity classification using decision tree and neural network for smart health application. *The Mediterranean Journal of Computers and Networks*, 9(1):10–19, Jan 2013. ISSN 1744-2397. 104

[3] Saisakul Chernbumroong, Shuang Cang, and Hongnian Yu. Maximum relevancy maximum complementary feature selection for multi-sensor activity recognition. *Expert Systems with Applications*, 42(1):573 – 583, 2015. ISSN 0957-4174. 104, 223, 224

[4] S. Chernbumroong, S. Cang, and H. Yu. Ga-based classifiers fusion for multi-sensor activity recognition of elderly people. *Biomedical and Health Informatics, IEEE Journal of*, PP(99):1–1, 2014. ISSN 2168-2194. 171, 225

[5] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003. 179

[6] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University, 2003. 175

[7] C Okan Sakar and Olcay Kursun. A method for combining mutual information and canonical correlation analysis: Predictive mutual information

and its use in feature selection. *Expert Systems with Applications*, 39(3): 3333–3344, 2012. 123

[8] Saisakul Chernbumroong, Shuang Cang, Anthony Atkins, and Hongnian Yu. Elderly activities recognition and classification for applications in assisted living. *Expert Systems with Applications*, 40(5):1662–1674, 2013. 104, 171, 223

[9] Saisakul Chernbumroong, Anthony S Atkins, and Hongnian Yu. Activity classification using a single wrist-worn accelerometer. In *Software, Knowledge Information, Industrial Management and Applications (SKIMA), 2011 5th International Conference on*, pages 1–6. IEEE, 2011. 104, 223

[10] Diogo Ayres-de campos, Joo Bernardes, Antonio Garrido, Joaquim Marques-de s, and Luis Pereira-leite. Sisporto 2.0: A program for automated analysis of cardiotocograms. *Journal of Maternal-Fetal and Neonatal Medicine*, 9(5):311–318, 2000. 87, 147

[11] Texas Instruments. Accelerometers and how they work, 2007. 25

[12] MATLAB. *version 8.2.0 (R2013b)*. The MathWorks Inc., Natick, Massachusetts, 2013. 196, 205

[13] Nishkam Ravi, Nikhil D, Preetham Mysore, and Michael L. Littman. Activity recognition from accelerometer data. In *In Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence(IAAI*, pages 1541–1546. AAAI Press, 2005. 25, 70

[14] Merryn J Mathie, Adelle C F Coster, Nigel H Lovell, and Branko G Celler. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiological Measurement*, 25(2):R1, 2004. 25

[15] C. E. Shannon. A mathematical theory of communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, 5(1):3–55, January 2001. ISSN 1559-1662. 122, 155

[16] H. G. Nguyen, A. Bouzerdoum, and S. L. Phung. *Learning Pattern Classification Tasks with Imbalanced Data Sets*. Pattern Recognition. intechopen, 2009. ISBN 978-953-307-014-8. 127

[17] Yishi Zhang and Zigang Zhang. Feature subset selection with cumulate conditional mutual information minimization. *Expert Systems with Applications*, 39(5):6078 – 6088, 2012. ISSN 0957-4174. 122

[18] W. Wang, P. Jones, and D. Partridge. Assessing the impact of input features in a feedforward neural network. *Neural Computing & Applications*, 9(2): 101–112, 2000. ISSN 0941-0643. 39, 124, 134, 155

[19] R. Setiono and Huan Liu. Neural-network feature selector. *Neural Networks, IEEE Transactions on*, 8(3):654–662, 1997. ISSN 1045-9227. 39, 124

[20] Roberto Battiti. Using mutual information for selecting features in supervised neural net learning. *Neural Networks, IEEE Transactions on*, 5(4): 537–550, 1994. 39

[21] M. Losch, S. Schmidt-Rohr, S. Knoop, S. Vacek, and R. Dillmann. Feature set selection and optimal classifier for human activity recognition. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 1022–1027, Aug 2007. 51

[22] Douglas L Vail and Manuela M Veloso. Feature selection for activity recognition in multi-robot domains. In *AAAI*, volume 8, pages 1415–1420, 2008.

[23] Tomoo Inoue. Future dining table: Dish recommendation based on dining activity recognition. In *Human-Computer Interaction*, pages 39–49. Springer, 2010. xiv, 51

[24] Keng-hao Chang, Shih-yen Liu, Hao-hua Chu, JaneYung-jen Hsu, Cheryl Chen, Tung-yun Lin, Chieh-yu Chen, and Polly Huang. The diet-aware dining table: Observing dietary behaviors over a tabletop surface. In KennethP. Fishkin, Bernt Schiele, Paddy Nixon, and Aaron Quigley, editors, *Pervasive Computing*, volume 3968 of *Lecture Notes in Computer Science*,

pages 366–382. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-33894-9. 51

[25] Dhruv Kumar Mahajan. *A Frameouork for Activity Recognition and Detection of Unusual Activities*. PhD thesis, Indian Institute of Technology Delhi, 2004.

[26] Justin Muncaster and Yunqian Ma. Hierarchical model-based activity recognition with automatic low-level state discovery. *Journal of Multimedia*, 2(5), 2007.

[27] Charalampos Doukas and Ilias Maglogiannis. Intelligent pervasive healthcare systems. In Margarita Sordo, Sachin Vaidya, and LakhmiC. Jain, editors, *Advanced Computational Intelligence Paradigms in Healthcare - 3*, volume 107 of *Studies in Computational Intelligence*, pages 95–115. Springer Berlin Heidelberg, 2008. ISBN 978-3-540-77661-1. 48, 49

[28] Thurmon E. Lockhart, Adam T. Barth, Xiaoyue Zhang, Rahul Songra, Emaad Abdel-Rahman, and John Lach. Portable, non-invasive fall risk assessment in end stage renal disease patients on hemodialysis. In *Wireless Health 2010*, WH '10, pages 84–93, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-989-3. 49

[29] Inaki Maurtua, Pierre T. Kirisci, Thomas Stiefmeier, Marco Luca Sbodio, and Hendrik Witt. A wearable computing prototype for supporting training activities in automotive production. In *Applied Wearable Computing (IFAWC), 2007 4th International Forum on*, pages 1–12, March 2007. 50

[30] Thomas Stiefmeier, Clemens Lombriser, Daniel Roggen, Holger Junker, Georg Ogris, and Gerhard Troester. Event-based activity tracking in work environments. In *Applied Wearable Computing (IFAWC), 2006 3rd International Forum on*, pages 1–10, March 2006. xiv, 50

[31] C. Zhu and Weihua Sheng. Human daily activity recognition in robot-assisted living using multi-sensor fusion. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, pages 2154–2159, May 2009. 49

[32] Yi He and Ye Li. Physical activity recognition utilizing the built-in kinematic sensors of a smartphone. *International Journal of Distributed Sensor Networks*, 2013, 2013.

[33] Martha E. Pollack. Intelligent technology for an aging population: The use of ai to assist elders with cognitive impairment. *AI Magazine*, 26(2):9–24, 2005.

[34] MM Massin, Marie-Christine Lebrethon, Danièle Rocour, Paul Gèrard, and Jean-Pierre Bourguignon. Patterns of physical activity determined by heart rate monitoring among diabetic children. *Archives of disease in childhood*, 90(12):1223–1226, 2005. 26, 70

[35] Yasuaki Ohtaki, Hikaru Inooka, Koichi Sagawa, Akihiro Suzuki, Zhan Xiumin, Mitsuharu Okutsu, and Ryoichi Nagatomi. Recognition of daily ambulatory movements utilizing accelerometer and barometer. *Power*, 100: 102, 2004. 26, 70

[36] Y. Nam and J.W. Park. Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor. *Biomedical and Health Informatics, IEEE Journal of*, 17(2):420–426, 2013. 26, 70

[37] Clemens Lombriser, Nagendra B. Bharatula, Daniel Roggen, and Gerhard Tröster. On-body activity recognition in a dynamic sensor network. In *Proceedings of the ICST 2nd international conference on Body area networks*, BodyNets '07, pages 17:1–17:6, 2007. 26, 70

[38] Hristijan Gjoreski and Matjaž Gams. Activity/posture recognition using wearable sensors placed on different body locations. *Proceeding of signal and image processing and applications*, 2011. 26, 70

[39] The Regulation and Quality Improvement Authority. Staffing guidance for nursing homes, version 1.5. http://www.rqia.org.uk/cms_resources/Staffing%20Guidelines%20for%20Nursing%20Home%20Version.pdf, Jun 2009. 3

[40] SAGA. Final cost of long-term care predicted to double in 20 years. http://www.saga.co.uk/money-and-finance/retirement-solutions/care-funding/press-releases/final-cost.aspx, Feb 2008. 2

[41] Christopher Chantrill. Time series chart of uk public spending. *UK Central Government and Local Authority Public Spending 2013*, 2012. 2

[42] Richard W Johnson. Will health care costs bankrupt aging boomers?, 2010. 2

[43] Bernd Rechel, Yvonne Doyle, Emily Grundy, and Martin McKee. How can health systems respond to population ageing. *World Health Organization, Regional Office for Europe and European Observatory on Health Systems and Policies*, 2009. 2

[44] Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat. World population ageing: 1950-2050, 2002. URL http://www.un.org/esa/population/publications/worldageing19502050. 2

[45] United Nations. World population prospects: The 2008 revision, 2009. URL http://www.un.org/esa/population/publications/wpp2008/wpp2008_highlights.pdf. xiv, 1, 2

[46] David E Goldberg and John H Holland. Genetic algorithms and machine learning. *Machine learning*, 3(2):95–99, 1988. 194

[47] A.B. Santos, A.de A. Araujo, and D. Menotti. Combining multiple classification methods for hyperspectral data interpretation. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 6(3): 1450–1459, 2013. ISSN 1939-1404. 192

[48] D. Trabelsi, S. Mohammed, F. Chamroukhi, L. Oukhellou, and Y. Amirat. An unsupervised approach for automatic activity recognition based on hidden markov model regression. *Automation Science and Engineering, IEEE Transactions on*, 10(3):829–835, 2013. 190

[49] Zhelong Wang, Ming Jiang, Yaohua Hu, and Hongyi Li. An incremental learning method based on probabilistic neural networks and adjustable fuzzy clustering for human activity recognition by using wearable sensors. *Information Technology in Biomedicine, IEEE Transactions on*, 16(4):691–699, 2012. 190

[50] Ming-Zher Poh. *Continuous assessment of epileptic seizures with wrist-worn biosensors*. PhD thesis, Massachusetts Institute of Technology, 2011.

[51] Akane Sano and Rosalind W Picard. Toward a taxonomy of autonomic sleep patterns with electrodermal activity. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 777–780. IEEE, 2011.

[52] I.H. Witten, E. Frank, and M.A Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, USA, 2011.

[53] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York, NY, USA, 1995. 134, 172, 173

[54] Shuang Cang. A non-linear tourism demand forecast combination model. *Tourism Economics*, 17(1):5–20, 2011. 201

[55] I. Cohen, F.G. Cozman, N. Sebe, M.C. Cirelo, and T.S. Huang. Semisupervised learning of classifiers: theory, algorithms, and their application to human-computer interaction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(12):1553–1566, 2004. ISSN 0162-8828. 88, 149

[56] Boonserm Kijsirikul, Sukree Sinthupinyo, and Kongsak Chongkasemwongse. Approximate match of rules using backpropagation neural networks. *Machine Learning*, 44(3):273–299, 2001. ISSN 0885-6125. 88, 149

[57] Kristin P. Bennett and O. L. Mangasarian. Robust linear programming discrimination of two linearly inseparable sets, 1992. 87, 144

[58] Md. Monirul Kabir, Md. Monirul Islam, and Kazuyuki Murase. A new wrapper feature selection approach using neural network. *Neurocomputing*,

73(1618):3273 – 3283, 2010. ISSN 0925-2312. 10th Brazilian Symposium on Neural Networks (SBRN2008). 39, 87, 124, 144

[59] András Antos, Balázs Kégl, Tamás Linder, and Gábor Lugosi. Data-dependent margin-based generalization bounds for classification. *J. Mach. Learn. Res.*, 3:73–98, March 2003. ISSN 1532-4435. 87, 144

[60] Ping Zhong and Masao Fukushima. Regularized nonsmooth newton method for multi-class support vector machines. *Optimization Methods and Software*, 22(1):225–236, 2007. 87, 143

[61] Jennifer G. Dy, Carla E. Brodley, and Stefan Wrobel. Feature selection for unsupervised learning. *Journal of Machine Learning Research*, 5:845–889, 2004. 87, 143

[62] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1226–1238, 2005. 39, 122, 123

[63] P.A. Estevez, M. Tesmer, C.A. Perez, and J.M. Zurada. Normalized mutual information feature selection. *Neural Networks, IEEE Transactions on*, 20 (2):189–201, 2009. 122, 124, 140

[64] K. Bache and M. Lichman. UCI machine learning repository, 2013. 86, 138

[65] Shuang Cang and Hongnian Yu. Mutual information based input feature selection for classification problems. *Decision Support Systems*, 54(1):691 – 698, 2012. ISSN 0167-9236. 39, 122, 138, 140

[66] Jamie A Ward, Nagendra Bharatula, Gerhard Troster, and Paul Lukowicz. Continuous activity recognition in the kitchen using miniaturised sensor button. *Technical Notes*, 2002. 46

[67] D.H. Wilson and C. Atkeson. Simultaneous tracking and activity recognition (star) using many anonymous, binary sensors. In Hans-W. Gellersen, Roy Want, and Albrecht Schmidt, editors, *Pervasive Computing*, volume

3468 of *Lecture Notes in Computer Science*, pages 62–79. Springer Berlin Heidelberg, 2005. ISBN 978-3-540-26008-0. 46

[68] Dairazalia Sánchez, Monica Tentori, and Jesús Favela. Activity recognition for the smart hospital. *Intelligent Systems, IEEE*, 23(2):50–57, 2008. 46

[69] S Szewcyzk, K Dwan, B Minor, B Swedlove, and D Cook. Annotating smart environment sensor data for activity learning. *Technology and Health Care*, 17(3):161–169, 2009. 47

[70] Jos Antonio Iglesias, Plamen P. Angelov, Agapito Ledezma, and Araceli Sanchs. Human activity recognition based on evolving fuzzy systems. *Int. J. Neural Syst.*, 20(5):355–364, 2010. 47

[71] A. Jehad Sarkar, Sungyoung. Lee, and Young-Koo. Lee. A Smoothed Naive Bayes-Based Classifier for Activity Recognition. *IETE Technical Review*, 27(2):107–119, 2010. 47

[72] Lei Yu and Huan Liu. Efficient feature selection via analysis of relevance and redundancy. *J. Mach. Learn. Res.*, 5:1205–1224, December 2004. ISSN 1532-4435. 7

[73] Ajith H. Gunatilaka and Brian A. Baertlein. Feature-level and decision-level fusion of noncoincidently sampled sensors for land mine detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):577–589, June 2001. ISSN 0162-8828. 6, 7, 28

[74] U. Maurer, A. Smailagic, D.P. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In *Wearable and Implantable Body Sensor Networks, 2006. BSN 2006. International Workshop on*, pages 4 pp.–116, April 2006. 46

[75] Uwe Maurer, Anthony Rowe, Asim Smailagic, and Daniel Siewiorek. Location and activity recognition using ewatch: A wearable sensor platform. In Yang Cai and Julio Abascal, editors, *Ambient Intelligence in Everyday Life*, volume 3864 of *Lecture Notes in Computer Science*, pages 86–102. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-37785-6. 24, 30, 46, 61, 180, 190

[76] Po-Cheng Huang, Sz-Shian Lee, Yaw-Huang Kuo, and Kuan-Rong Lee. A flexible sequence alignment approach on pattern mining and matching for human activity recognition. *Expert Systems with Applications*, 37(1):298 – 306, 2010. ISSN 0957-4174. 60, 180

[77] P. Siirtola, P. Laurinen, Eija Haapalainen, J Roning, and H. Kinnunen. Clustering-based activity classification with a wrist-worn accelerometer using basic features. In *Computational Intelligence and Data Mining, 2009. CIDM '09. IEEE Symposium on*, pages 95–100, March 2009. 49

[78] Martha E. Pollack, Laura Brown, Dirk Colbry, Colleen E. McCarthy, Cheryl Orosz, Bart Peintner, Sailesh Ramakrishnan, and Ioannis Tsamardinos. Autominder: an intelligent cognitive orthotic system for people with memory impairment. *Robotics and Autonomous Systems*, 44(34):273 – 282, 2003. ISSN 0921-8890. Best papers presented at IAS-7. xiv, 3, 47, 48

[79] A. Fleury, M. Vacher, and N. Noury. Svm-based multimodal classification of activities of daily living in health smart homes: Sensors, algorithms, and first experimental results. *Information Technology in Biomedicine, IEEE Transactions on*, 14(2):274–283, March 2010. ISSN 1089-7771. 4, 6, 7, 26, 30, 47, 53, 58, 70

[80] Bo-Chao Cheng, Yi-An Tsai, Guo-Tan Liao, and Eui-Seok Byeon. Hmm machine learning and inference for activities of daily living recognition. *The Journal of Supercomputing*, 54(1):29–42, 2010. ISSN 0920-8542. 4, 47, 53, 57

[81] Ahmad Lotfi, Caroline Langensiepen, SawsanM. Mahmoud, and M.J. Akhlaghinia. Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour. *Journal of Ambient Intelligence and Humanized Computing*, 3(3):205–218, 2012. ISSN 1868-5137. 4, 53

[82] Xin Hong, Chris Nugent, Maurice Mulvenna, Sally McClean, Bryan Scotney, and Steven Devlin. Evidential fusion of sensor data for activity recognition in smart homes. *Pervasive and Mobile Computing*, 5(3):236 – 252, 2009. ISSN 1574-1192. 4, 47, 53, 56, 60

[83] Josef Diermaier, Katharina Neyder, Franz Werner, Paul Panek, and WolfgangL. Zagler. Distributed accelerometers as a main component in detecting activities of daily living. In Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, editors, *Computers Helping People with Special Needs*, volume 5105 of *Lecture Notes in Computer Science*, pages 1042–1049. Springer Berlin Heidelberg, 2008. ISBN 978-3-540-70539-0. 4, 25, 46, 53, 56, 70, 230

[84] Seon-Woo Lee and K. Mase. Activity and location recognition using wearable sensors. *Pervasive Computing, IEEE*, 1(3):24–32, 2002. ISSN 1536-1268. 6, 23, 28, 30, 46

[85] Jie Yin, Qiang Yang, and J.J. Pan. Sensor-based abnormal human-activity detection. *Knowledge and Data Engineering, IEEE Transactions on*, 20(8): 1082–1090, 2008. ISSN 1041-4347. 4, 17, 21, 23, 46, 54, 55

[86] Somboon Hongeng, Ram Nevatia, and Francois Bremond. Video-based event recognition: activity representation and probabilistic recognition methods. *Computer Vision and Image Understanding*, 96(2):129 – 162, 2004. ISSN 1077-3142. 4, 17

[87] Zhongna Zhou, Xi Chen, Yu-Chia Chung, Zhihai He, T.X. Han, and J.M. Keller. Activity analysis, summarization, and visualization for indoor human activity monitoring. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(11):1489–1498, 2008. ISSN 1051-8215. 4, 17

[88] Weiyao Lin, Ming-Ting Sun, R. Poovendran, and Zhengyou Zhang. Activity recognition using a combination of category components and local models for video surveillance. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(8):1128–1139, 2008. ISSN 1051-8215. 4, 17

[89] M. Singh, A. Basu, and M.K. Mandal. Human activity recognition based on silhouette directionality. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(9):1280–1292, 2008. ISSN 1051-8215. 4, 17

[90] Barthel DW Mahoney FI. Functional evaluation: the barthel index. *Maryland State Med Journal*, 14:56–61, 1965. 76, 79

[91] Mary Shelkey and Meredith Wallace. Katz index of independence in activities of daily living (adl). *The Gerontologist*, 10(1):20–30, 1998. 76

[92] Simon Günter and Horst Bunke. Optimization of weights in a multiple classifier handwritten word recognition system using a genetic algorithm. *Electronic Letters on Computer Vision and Image Analysis*, 3(1):25–41, 2004. 12, 44, 63, 190, 191, 192

[93] M. J. Valadan Zoej Y. Maghsoudi, A. Alimohammadi and B. Mojaradi. Weighted combination of multiple classifiers for the classification of hyperspectral images using a genetic algorithm. In *SPRS Commission VII, Mid-term Symposium,Remote Sensing: From Pixels to Processes*, Enschede, Netherlands, 2006. 12, 44, 63, 190, 191, 192

[94] Pekka Kahri. Ict based solutions for advancement of social for advancement of social interaction of elderly people interaction of elderly people, 2008. 59

[95] Ricardo Andr Fernandes Costa. *Collaborative Networks in Ambient Assisted Living*. PhD thesis, University of Minho, 2009. 59

[96] Hong Sun, V. De Florio, Ning Gui, and C. Blondia. Promises and challenges of ambient assisted living systems. In *Information Technology: New Generations, 2009. ITNG '09. Sixth International Conference on*, pages 1201–1207, 2009. 59

[97] Hong Sun, V. De Florio, Ning Gui, and C. Blondia. Participant: A new concept for optimally assisting the elder people. In *Computer-Based Medical Systems, 2007. CBMS '07. Twentieth IEEE International Symposium on*, pages 295–300, 2007. 59

[98] Hong Sun, V. De Florio, and C. Blondia. A design tool to reason about ambient assisted living systems. In *Intelligent Systems Design and Applications, 2006. ISDA '06. Sixth International Conference on*, volume 2, pages 1136–1144, 2006. 59

[99] Haoyong Yu, Matthew Spenko, and Steven Dubowsky. An adaptive shared control system for an intelligent mobility aid for the elderly. *Autonomous Robots*, 15(1):53–66, 2003. ISSN 0929-5593. 59

[100] Leon Harper Ada-Helen Bayer. Fixing to stay: A national survey of housing and home modification issues, 2000. 58

[101] B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C.J. Bula, and P. Robert. Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly. *Biomedical Engineering, IEEE Transactions on*, 50(6):711–723, 2003. ISSN 0018-9294. 4, 6, 24, 29, 30, 37, 46, 52, 54, 55, 230

[102] George Demiris, Brian K. Hensel, Marjorie Skubic, and Marilyn Rantz. Senior residents perceived need of and preferences for smart home sensor technologies. *International Journal of Technology Assessment in Health Care*, 24:120–124, 1 2008. ISSN 1471-6348. doi: 10.1017/S0266462307080154. 59

[103] Bogdan Gabrys and Dymitr Ruta. Genetic algorithms in classifier fusion. *Applied Soft Computing*, 6(4):337 – 347, 2006. ISSN 1568-4946. 12, 13, 43, 44, 63, 192, 212

[104] Kevin K.F. Wong, Haiyan Song, Stephen F. Witt, and Doris C. Wu. Tourism forecasting: To combine or not to combine? *Tourism Management*, 28(4):1068 – 1078, 2007. ISSN 0261-5177. 44

[105] Chang Woo HAN, Shin Jae KANG, and Nam Soo KIM. Implementation of hmm-based human activity recognition using single triaxial accelerometer. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E93.A(7):1379–1383, 2010. 24, 47

[106] A. G. Bonomi, G. Plasqui, A. H. C. Goris, and K. R. Westerterp. Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. *Journal of Applied Physiology*, 107(3): 655–661, 2009. 24, 47, 52, 106

[107] T. Choudhury, S. Consolvo, B. Harrison, J. Hightower, A. LaMarca, L. Legrand, A. Rahimi, A. Rea, G. Bordello, B. Hemingway, P. Klasnja, K. Koscher, J.A. Landay, J. Lester, D. Wyatt, and D. Haehnel. The mobile sensing platform: An embedded activity recognition system. *Pervasive Computing, IEEE*, 7(2):32–41, 2008. ISSN 1536-1268. 6, 7, 24, 30, 46, 47, 61, 62

[108] D.M. Karantonis, M.R. Narayanan, M. Mathie, N.H. Lovell, and B.G. Celler. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *Information Technology in Biomedicine, IEEE Transactions on*, 10(1):156–167, 2006. ISSN 1089-7771. 4, 24, 46, 52, 54, 55

[109] Maged N. Kamel Boulos, Ricardo Castellot Lou, Athanasios Anastasiou, Chris D. Nugent, Jan Alexandersson, Gottfried Zimmermann, Ulises Cortes, and Roberto Casas. Connectivity for healthcare and well-being management: Examples from six european projects. *International Journal of Environmental Research and Public Health*, 6(7):1947–1971, 2009. ISSN 1660-4601.

[110] Reinhard Oppermann, Rossen Rashev, and Kinshuk. Adaptability and Adaptivity in Learning Systems. *Knowledge Transfer*, II:173–179, 1997.

[111] E. Aarts. Ambient intelligence: a multimedia perspective. *MultiMedia, IEEE*, 11(1):12–19, 2004. ISSN 1070-986X.

[112] Thomas Kleinberger, Martin Becker, Eric Ras, Andreas Holzinger, and Paul Mller. Ambient intelligence in assisted living: Enable elderly people to handle future interfaces. In Constantine Stephanidis, editor, *Universal Access in Human-Computer Interaction. Ambient Interaction*, volume 4555 of *Lecture Notes in Computer Science*, pages 103–112. Springer Berlin Heidelberg, 2007. ISBN 978-3-540-73280-8. 58, 59, 60

[113] Ethem Alpaydin. *Introduction to Machine Learning*. The MIT Press, Cambridge, Massachusetts, London, England, 2004. ISBN 978-0470294666. 28, 98, 108

[114] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm. 173, 175

[115] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on*, 13(2): 415–425, 2002. ISSN 1045-9227. 40, 122

[116] Jhun-Ying Yang, Jeen-Shing Wang, and Yen-Ping Chen. Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers. *Pattern Recognition Letters*, 29(16):2213 – 2220, 2008. ISSN 0167-8655. 24, 41, 46, 61, 106

[117] Mitja Luštrek and Boštjan Kaluža. Fall detection and activity recognition with machine learning. *Informatica*, 33(2):197–204, 2009. 4, 6, 23, 30, 40, 46, 54, 55

[118] Andrea Mannini, Stephen S Intille, Mary Rosenberger, Angelo M Sabatini, and William Haskell. Activity recognition using a single accelerometer placed at the wrist or ankle. *Medicine and science in sports and exercise*, 2013. 37

[119] Weihua He, Yongcai Guo, Chao Gao, and Xinke Li. Recognition of human activities with wearable sensors. *EURASIP Journal on Advances in Signal Processing*, 2012(1):1–13, 2012. 38

[120] JohnPaul Varkey, Dario Pompili, and TheodoreA. Walls. Human motion recognition using a wireless sensor-based wearable system. *Personal and Ubiquitous Computing*, 16(7):897–910, 2012. ISSN 1617-4909. 39

[121] Jian-Xun Peng, Stuart Ferguson, Karen Rafferty, and Paul D. Kelly. An efficient feature selection method for mobile devices with application to activity recognition. *Neurocomputing*, 74(17):3543 – 3552, 2011. ISSN 0925-2312. 39

[122] A. Dalton and G. Olaighin. Comparing supervised learning techniques on the task of physical activity recognition. *Biomedical and Health Informatics, IEEE Journal of*, 17(1):46–52, 2013. ISSN 2168-2194. 39

[123] M. Zhang and A.A. Sawchuk. Human daily activity recognition with sparse representation using wearable sensors. *Biomedical and Health Informatics, IEEE Journal of*, 17(3):553–560, 2013. ISSN 2168-2194. 39

[124] M. Ermes, J. Parkka, J. Mantyjarvi, and I. Korhonen. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *Information Technology in Biomedicine, IEEE Transactions on*, 12(1):20–26, 2008. ISSN 1089-7771. 6, 8, 23, 30, 39, 41, 46, 49, 119

[125] Oresti Banos, Miguel Damas, Hector Pomares, Fernando Rojas, Blanca Delgado-Marquez, and Olga Valenzuela. Human activity recognition based on a sensor weighting hierarchical classifier. *Soft Computing*, 17(2):333–343, 2013. ISSN 1432-7643. 39

[126] Yvan Saeys, Iaki Inza, and Pedro Larraaga. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19):2507–2517, 2007. doi: 10.1093/bioinformatics/btm344. 39

[127] B. Adar Emken, Ming Li, Gautam Thatte, Sangwon Lee, Murali Annavaram, Urbashi Mitra, Shrikanth Narayanan, and Donna Spruijt-Metz. Recognition of physical activities in overweight hispanic youth using knowme networks. *Journal of Physical Activity & Health*, 9(3):432 – 441, 2012. ISSN 15433080. 38

[128] Myong-Woo Lee, AdilMehmood Khan, and Tae-Seong Kim. A single triaxial accelerometer-based real-time personal life log system capable of human activity recognition and exercise information generation. *Personal and Ubiquitous Computing*, 15(8):887–898, 2011. ISSN 1617-4909. 38

[129] A.M. Khan, Young-Koo Lee, S.Y. Lee, and Tae-Seong Kim. A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *Information Technology in Biomedicine,*

IEEE Transactions on, 14(5):1166–1172, 2010. ISSN 1089-7771. 6, 24, 30, 38, 39, 41, 47, 52, 106

[130] Yen-Ping Chen, Jhun-Ying Yang, Shun-Nan Liou, Gwo-Yun Lee, and Jeen-Shing Wang. Online classifier construction algorithm for human activity detection using a tri-axial accelerometer. *Applied Mathematics and Computation*, 205(2):849 – 860, 2008. ISSN 0096-3003. Special Issue on Advanced Intelligent Computing Theory and Methodology in Applied Mathematics and Computation. 24, 38, 42, 46, 61, 106

[131] Ganapathiraju Balakrishnama. Linear Discriminant Analysis - A brief Tutorial, 1998. 38

[132] Yunji Liang, Xingshe Zhou, Zhiwen Yu, and Bin Guo. Energy-efficient motion related activity recognition on mobile devices for pervasive healthcare. *Mobile Networks and Applications*, pages 1–15, 2013. ISSN 1383-469X. 38

[133] Yiqiang Chen, Zhongtang Zhao, Shuangquan Wang, and Zhenyu Chen. Extreme learning machine-based device displacement free activity recognition model. *Soft Computing*, 16(9):1617–1625, 2012. ISSN 1432-7643. 38

[134] Georg Ogris, Paul Lukowicz, Thomas Stiefmeier, and Gerhard Trster. Continuous activity recognition in a maintenance scenario: combining motion sensors and ultrasonic hands tracking. *Pattern Analysis and Applications*, 15(1):87–111, 2012. ISSN 1433-7541. 38

[135] T. Pawar, S. Chaudhuri, and Siddhartha P. Duttagupta. Body movement activity recognition for ambulatory cardiac monitoring. *Biomedical Engineering, IEEE Transactions on*, 54(5):874–882, 2007. ISSN 0018-9294. 4, 24, 38, 46, 54, 55

[136] Javier Andreu and Plamen Angelov. An evolving machine learning method for human activity recognition systems. *Journal of Ambient Intelligence and Humanized Computing*, 4(2):195–206, 2013. ISSN 1868-5137. 36, 38

[137] S. Allen Broughton and Kurt M. Bryan. *Discrete Fourier Analysis and Wavelets: Applications to Signal and Image Processing*. Wiley-Blackwell, USA, 2008. ISBN 978-0470294666. 36

[138] Ricardo Chavarriaga, Hamidreza Bayati, and Jos delR. Milln. Unsupervised adaptation for acceleration-based activity recognition: robustness to sensor displacement and rotation. *Personal and Ubiquitous Computing*, 17(3):479–490, 2013. ISSN 1617-4909. 35

[139] Kerem Altun, Billur Barshan, and Orkun Tunel. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognition*, 43(10):3605 – 3620, 2010. ISSN 0031-3203. 6, 7, 17, 21, 23, 29, 30, 35, 38, 40, 47

[140] Davide Figo, Pedro C. Diniz, Diogo R. Ferreira, and João M. Cardoso. Preprocessing techniques for context recognition from accelerometer data. *Personal Ubiquitous Comput.*, 14(7):645–662, oct 2010. ISSN 1617-4909. 35, 36, 106, 107

[141] David G. Stork Richard O. Duda, Peter E. Hart. *Pattern Classification*. Wiley & Sons Inc., Canada, 2nd edition, 2000. ISBN 978-0-471-05669-0. 34

[142] S Preece, JY Goulermas, LPJ Kenney, D Howard, K Meijer, and R Crompton. Activity identification using body-mounted sensors – a review of classification techniques. *Physiological Measurement*, 30:R1–R33, 2009. 37

[143] Emmanuel Munguia Tapia. *Using Machine Learning for Real-time Activity Recognition and Estimation of Energy Expenditure*. PhD thesis, School of Architecture and Planning, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2008. 26, 33, 34, 97, 166

[144] Illapha Cuba Gyllensten. Physical activity recognition in daily life using a triaxial accelerometer. Master's thesis, School of Computer Science and Engineering, Royal Institute of Technology, Stockholm, Sweden, 2010. 33, 36, 105, 106

[145] AdilMehmood Khan, Young-Koo Lee, Sungyoung Lee, and Tae-Seong Kim. Accelerometers position independent physical activity recognition system for long-term activity monitoring in the elderly. *Medical & Biological Engineering & Computing*, 48(12):1271–1279, 2010. ISSN 0140-0118. 4, 33, 35, 38, 47, 54, 57, 230

[146] Oliver Amft and Gerhard Trster. Recognition of dietary activity events using on-body sensors. *Arificial Intelligence in Medicine*, 42:121–136, 2008. 6, 23, 25, 30, 35, 46

[147] P. Lopez-Meyer, S. Tiffany, Y. Patil, and E. Sazonov. Monitoring of cigarette smoking using wearable sensors and support vector machines. *Biomedical Engineering, IEEE Transactions on*, 60(7):1867–1872, 2013. ISSN 0018-9294. 25, 27

[148] Ioannis Kouris and Dimitris Koutsouris. A comparative study of pattern recognition classifiers to predict physical activities using smartphones and wearable body sensors. *Technology and Health Care*, 20:263–275, 2012. 26

[149] Henar Martn, AnaM. Bernardos, Josu Iglesias, and JosR. Casar. Activity logging using lightweight classification techniques in mobile devices. *Personal and Ubiquitous Computing*, 17(4):675–695, 2013. ISSN 1617-4909. 25, 33

[150] Che-Chang Yang and Yeh-Liang Hsu. A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors*, 10(8): 7772–7788, 2010. ISSN 1424-8220. 25, 27, 165

[151] Kazushige Ouchi and Miwako Doi. Living activity recognition using off-the-shelf sensors on mobile phones. *annals of telecommunications - annales des tlcommunications*, 67(7-8):387–395, 2012. ISSN 0003-4347. 25

[152] Norbert Györbíró, Ákos Fábián, and Gergely Hományi. An activity recognition system for mobile phones. *Mob. Netw. Appl.*, 14(1):82–91, February 2009. ISSN 1383-469X. 6, 17, 21, 23, 25, 30, 35, 41, 47, 172, 179

[153] U Ekelund, A Yngve, and M Sjöström. Total daily energy expenditure and patterns of physical activity in adolescents assessed by two different methods. *Scandinavian journal of medicine & science in sports*, 9(5):257–264, 1999. 26, 70

[154] J Booyens and GR Hervey. The pulse rate as a means of measuring metabolic rate in man. *Canadian Journal of Biochemistry and Physiology*, 38(11):1301–1309, 1960. 26, 70

[155] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen. Activity classification using realistic data from wearable sensors. *Information Technology in Biomedicine, IEEE Transactions on*, 10(1):119–128, 2006. ISSN 1089-7771. 6, 8, 17, 21, 23, 25, 26, 29, 30, 33, 35, 39, 41, 70, 106, 119, 172, 190

[156] Liming Chen and Chris Nugent. Ontology-based activity recognition in intelligent pervasive environments. *International Journal of Web Information Systems*, 5(4):410–430, 2009. ISSN 1744-0084. 19

[157] Jhun-Ying Yang, Yen-Ping Chen, Gwo-Yun Lee, Shun-Nan Liou, and Jeen-Shing Wang. Activity recognition using one triaxial accelerometer: A neuro-fuzzy classifier with feature reduction. In Lizhuang Ma, Matthias Rauterberg, and Ryohei Nakatsu, editors, *Entertainment Computing ICEC 2007*, volume 4740 of *Lecture Notes in Computer Science*, pages 395–400. Springer Berlin Heidelberg, 2007. ISBN 978-3-540-74872-4. 24, 25, 38, 46, 61, 70, 106

[158] Sen Zhang, Marcelo H. Ang, Wendong Xiao, and Chen Khong Tham. Detection of activities by wireless sensors for daily life surveillance: Eating and drinking. *Sensors*, 9(3):1499–1517, 2009. ISSN 1424-8220. 24, 25, 35, 42, 47, 61, 70

[159] Takuya Maekawa and Shinji Watanabe. Training data selection with users physical characteristics data for acceleration-based activity modeling. *Personal and Ubiquitous Computing*, 17(3):451–463, 2013. ISSN 1617-4909. 25, 70

[160] Zhongna Zhou, Xi Chen, Yu-Chia Chung, Zhihai He, T.X. Han, and J.M. Keller. Activity analysis, summarization, and visualization for indoor human activity monitoring. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(11):1489–1498, 2008. ISSN 1051-8215. 4, 17

[161] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea. Machine recognition of human activities: A survey. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(11):1473–1488, 2008. ISSN 1051-8215. doi: 10.1109/TCSVT.2008.2005594. 17

[162] Sarvesh Vishwakarma and Anupam Agrawal. A survey on activity recognition and behavior understanding in video surveillance. *The Visual Computer*, pages 1–27, 2012. ISSN 0178-2789. doi: 10.1007/s00371-012-0752-6. 17

[163] EmmanuelMunguia Tapia, StephenS. Intille, and Kent Larson. Activity recognition in the home using simple and ubiquitous sensors. In Alois Ferscha and Friedemann Mattern, editors, *Pervasive Computing*, volume 3001 of *Lecture Notes in Computer Science*, pages 158–175. Springer Berlin Heidelberg, 2004. ISBN 978-3-540-21835-7. doi: 10.1007/978-3-540-24646-6_10. 4, 17, 19, 21, 46, 47

[164] Naoharu Yamada, Kenji Sakamoto, Goro Kunito, Yoshinori Isoda, Kenichi Yamazaki, and Satoshi Tanaka. Applying ontology and probabilistic model to human activity recognition from surrounding things(mobile computing). *IPSJ Journal*, 48(8):2823–2834, aug 2007. ISSN 03875806. 19, 46

[165] Niels Landwehr, Bernd Gutmann, Ingo Thon, Luc De Raedt, and Matthai Philipose. Relational transformation-based tagging for activity recognition. *Fundam. Inf.*, 89(1):111–129, January 2009. ISSN 0169-2968. 17, 19, 46

[166] Yu-Jin Hong, Ig-Jae Kim, Sang Chul Ahn, and Hyoung-Gon Kim. Mobile health monitoring system based on activity recognition using accelerometer. *Simulation Modelling Practice and Theory*, 18(4):446 – 455, 2010. ISSN 1569-190X. doi: http://dx.doi.org/10.1016/j.simpat.2009.09.002. Model-

ing and Simulation Techniques for Future Generation Communication Networks. 4, 17, 19, 20, 31, 47, 53, 57, 60, 180, 190

[167] Patrice C. Roy, Abdenour Bouzouane, Sylvain Giroux, and Bruno Bouchard. Possibilistic activity recognition in smart homes for cognitively impaired people. *Applied Artificial Intelligence*, 25(10):883–926, 2011. doi: 10.1080/08839514.2011.617248. 19

[168] Francisco E. Martnez-Prez, Jose ngel Gonzlez-Fraga, Juan C. Cuevas-Tello, and Marcela D. Rodrguez. Activity inference for ambient intelligence through handling artifacts in a healthcare environment. *Sensors*, 12(1): 1072–1099, 2012. ISSN 1424-8220. doi: 10.3390/s120101072. 19, 47

[169] Belkacem Chikhaoui, frui Wang, and Hlne Pigot. Adr-splda: Activity discovery and recognition by combining sequential patterns and latent dirichlet allocation. *Pervasive and Mobile Computing*, 8(6):845 – 862, 2012. ISSN 1574-1192. doi: http://dx.doi.org/10.1016/j.pmcj.2012.08.004. ¡ce:title¿Special Issue on Pervasive Healthcare¡/ce:title¿. 19

[170] A. Fleury, M. Vacher, and N. Noury. Svm-based multimodal classification of activities of daily living in health smart homes: Sensors, algorithms, and first experimental results. *Information Technology in Biomedicine, IEEE Transactions on*, 14(2):274–283, 2010. ID: 1. 17, 19, 179, 180, 190

[171] Liming Chen, C.D. Nugent, and Hui Wang. A knowledge-driven approach to activity recognition in smart homes. *Knowledge and Data Engineering, IEEE Transactions on*, 24(6):961–974, 2012. ISSN 1041-4347. doi: 10.1109/TKDE.2011.51. 19

[172] Jie Yin, Qiang Yang, Dou Shen, and Ze-Nian Li. Activity recognition via user-trace segmentation. *ACM Trans. Sen. Netw.*, 4(4):19:1–19:34, September 2008. ISSN 1550-4859. doi: 10.1145/1387663.1387665. 17, 20, 46

[173] Ching-Hu Lu and Li-Chen Fu. Robust location-aware activity recognition using wireless sensor network in an attentive home. *Automation Science and Engineering, IEEE Transactions on*, 6(4):598–609, 2009. ISSN 1545-5955. doi: 10.1109/TASE.2009.2021981. 4, 21, 47, 53, 56

[174] DJ Cook, M Schmitter-Edgecombe, et al. Assessing the quality of activities in a smart environment. *Methods of information in medicine*, 48(5):480, 2009. 4, 21, 47, 53, 56

[175] Chun Zhu and Weihua Sheng. Motion and location-based online human daily activity recognition. *Pervasive and Mobile Computing*, 7:256–269, 2011. doi: 10.1016/j.pmcj.2010.11.004. 6, 21, 30, 47, 172

[176] Laurent Millet. *Structural methods in pattern recognition*. North Oxford Academic, 1986. 21

[177] Ling Bao and StephenS. Intille. Activity recognition from user-annotated acceleration data. In Alois Ferscha and Friedemann Mattern, editors, *Pervasive Computing*, volume 3001 of *Lecture Notes in Computer Science*, pages 1–17. Springer Berlin Heidelberg, 2004. ISBN 978-3-540-21835-7. doi: 10.1007/978-3-540-24646-6_1. 6, 17, 21, 23, 25, 27, 29, 30, 33, 35, 36, 40, 46, 70

[178] Takuya Maekawa, Yasue Kishino, Yasushi Sakurai, and Takayuki Suyama. Activity recognition with hand-worn magnetic sensors. *Personal and Ubiquitous Computing*, 17(6):1085–1094, 2013. ISSN 1617-4909. doi: 10.1007/s00779-012-0556-8. 21

[179] P. Barsocchi. Position recognition to support bedsores prevention. *Biomedical and Health Informatics, IEEE Journal of*, 17(1):53–59, 2013. ISSN 2168-2194. doi: 10.1109/TITB.2012.2220374. 20, 48

[180] D.J. Cook, N.C. Krishnan, and P. Rashidi. Activity discovery and activity recognition: A new partnership. *Cybernetics, IEEE Transactions on*, 43(3): 820–828, 2013. ISSN 2168-2267. doi: 10.1109/TSMCB.2012.2216873. 19

[181] Corinne Belley, Sebastien Gaboury, Bruno Bouchard, and Abdenour Bouzouane. An efficient and inexpensive method for activity recognition within a smart home based on load signatures of appliances. *Pervasive and Mobile Computing*, 2013. 19

[182] J.A. Ward, P. Lukowicz, G. Troster, and T.E. Starner. Activity recognition of assembly tasks using body-worn microphones and accelerometers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(10): 1553–1567, 2006. ISSN 0162-8828. doi: 10.1109/TPAMI.2006.197. 6, 23, 26, 28, 30, 31, 34, 39, 41, 46, 97

[183] Faouzi Sebbak, Abdelghani Chibani, Yacine Amirat, Aicha Mokhtari, and Farid Benhammadi. An evidential fusion approach for activity recognition in ambient intelligence environments. *Robotics and Autonomous Systems*, 61(11):1235–1245, 2013. ISSN 0921-8890. doi: http://dx.doi.org/10.1016/j.robot.2013.05.010. 19

[184] PO Rocher, B Bouchard, and A Bouzouane. A new platform to easily experiment activity recognition systems based on passive rfid tags: Experimentation with data mining algorithms. *International Journal of Smart Home*, 6(2):7 – 24, 2012. 20

[185] Oresti Banos, Miguel Damas, Hector Pomares, Alberto Prieto, and Ignacio Rojas. Daily living activity recognition based on statistical feature quality group selection. *Expert Systems with Applications*, 39(9):8013 – 8021, 2012. ISSN 0957-4174. doi: 10.1016/j.eswa.2012.01.164. 25, 37, 70

[186] Holger Junker, Oliver Amft, Paul Lukowicz, and Gerhard Tröster. Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recogn.*, 41(6):2010–2024, jun 2008. ISSN 0031-3203. doi: 10.1016/j.patcog.2007.11.016. 23, 46