Entropy Measures for Anomaly Detection

Aadel Fathe Omar Howedi

School of Science and Technology

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of

Doctor of Philosophy

December 2020

This thesis is dedicated to my parents, wife and family with great gratitude. I know you will be proud of this milestone accomplished. Undoubtedly, without their prayers and support this thesis would have been impossible.

Acknowledgements

The work described in this thesis was carried out at Nottingham Trent University between October 2017 and December 2020, while I was working as a full-time doctoral research student.

All pray and glory to Allah Almighty for making this a reality.

My sincere gratitude to my honourable director of studies, Prof. Ahmad Lotfi, for his guidance throughout the duration of my research. Your consistent motivation and kind words will never be forgotten. His office was always open for me for any inquiry despite his many responsibilities. Also, his relentless effort and extensive suggestions around my work were indispensable to the accomplishment of this research. Also, I will like to thank the team of co-supervisors I had at different stages of my research: Dr. Caroline Langensiepen I am grateful for the creative ideas you always had to offer and for the many corrections you made to the publications leading to this thesis. I wish you were on my supervisory team till completion. Thank you Dr. Amir Pourabdollah for the supervision given at the later stage, especially, for your invaluable corrections to the publications we shared. Furthermore, I really appreciate and thank my sponsor, Ministry of Higher Education and Research of Libya for awarding me this Scholarship and the cultural affairs attache office at the Libyan embassy in London.

My sincere appreciation to my family for their support, sympathy, and patience throughout the journey of this study, particularly, I would like to say thanks to my wife Nora and our lovely kids, Joury, Mohamed and Yaseen. Your joy and love have always put me up during the well-being and sadness of the last four years. My thanks are extended to my lovely father who always encouraged me to carry on in my study, also I thank my lovely mother, sisters, brothers, brothers in law and uncles. Their prayers and faith have inspired and encouraged me to continue with this study. I say God bless you all for your support.

Finally, I would like to appreciate my friends and colleagues at Nottingham Trent University for the time shared during the past few years of my research. Specifically, my fellow PhD researchers in the Computational Intelligence and Applications (CIA) research group. To everyone who in one way or the other contributed to the success of my research, I am most grateful.

Aadel Fathe Omar Howedi
December 2020

Abstract

Human activity recognition methods are used to support older adults to live independently in their own homes by monitoring their Activities of Daily Living (ADL). The gathered data and information representing different activities will be used to identify anomalous activities in comparison with the routine activities. In the related research in this area, the most recent studies have mainly focused on detecting anomalies in a single occupant environment. Although older adults often receive visits from family members or health care workers, representing a multi-occupancy environment.

This research is focused on the application of entropy measures for anomaly detection in ADLs in a single-occupancy and multi-occupancy environment. In many applications, entropy measures are used to detect the irregularities and the degree of randomness in data. However, this has rarely been applied in the context of activities of daily living.

To address the research questions identified in the thesis, three novel contributions of the thesis are; Firstly, a novel method based on different entropy measures is investigated to detect anomalies in ADLs, specifically in sleeping routine and human falls. Secondly, a novel entropy-based method is explored to detect anomalies in ADLs in the presence of a visitor, solely based on information gathered from ambient sensors. Finally, entropy measures are applied to investigate their effectiveness in identifying a visitor in a single home environment based on data gathered from ambient sensors. The results presented in this thesis show that entropy measures could be used to detect abnormality (here, irregular sleep, human fall and a visitor) in ADLs.

Publications

As a result of the research presented in this thesis, the following publications have been published:

Refereed Journal Papers:

Aadel Howedi, Ahmad Lotfi, and Amir Pourabdollah. "Exploring Entropy Measurements to Identify Multi-Occupancy in Activities of Daily Living." Entropy 21, no. 4 (2019): 416.

Aadel Howedi, Ahmad Lotfi, and Amir Pourabdollah. "An Entropy-Based Approach for Anomaly Detection in Activities of Daily Living in the Presence of a Visitor." Entropy 22.8 (2020): 845.

Aadel Howedi, Ahmad Lotfi, and Amir Pourabdollah. "Employing Entropy Measures to Identify Visitors in Multi-Occupancy Environments." Journal of Ambient Intelligence and Humanized Computing. To be published in 2021 [doi: 10.1007/s12652-020-02824-z].

Refereed Conference Papers:

Aadel Howedi, Ahmad Lotfi, and Amir Pourabdollah. "Distinguishing activities of daily living in a multi-occupancy environment." In Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments, pp. 568-574. 2019.

Aadel Howedi, Ahmad Lotfi, and Amir Pourabdollah, "Accelerometer-based Human Fall Detection Using Fuzzy Entropy," 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Glasgow, United Kingdom, 2020, pp. 1-7, doi: 10.1109/FUZZ48607.2020.9177577.

Aadel Howedi, Ahmad Lotfi, and Amir Pourabdollah. "A multi-scale fuzzy entropy measure for anomaly detection in activities of daily living." In Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments, pp. 1-8. 2020.

Contents

Nomenclature

Acronyms

ADL	Activities of Daily Living
AmI	Ambient Intelligence
ApEn	Approximate Entropy
CHMM	Coupled Hidden Markov Models
CL-HMM	Combined Label Hidden Markov Models
CNN	Convolutional Neural Network
CNDE	Consensus Novelty Detection Ensemble
CRF	Convolutional Random Field
DCNN	Deep Convolutional Neural Network.
DNN	Deep Neural Network
EDCV	Ensemble of Detectors with Correlated Votes
EDVV	Ensemble of Detectors with Variability Votes
FCRF	Factorial Conditional Random Field
FHMM	Factorial Hidden Markov Model
FuzzyEn	Fuzzy Entropy
HAR	Human Activity Recognition
HHMM	Hierarchical Hidden Markov Model
HMM	Hidden Markov Model
IE	Intelligent Environment
IM	Indoor Mobility
LSTM	Long Short-term Memory
MFE	Multi-scale Fuzzy Entropy
MMPP	Markov Modulated Poisson Process

MPE	Multi-scale Permutation Entropy
MSE	Multiscale Entropy
NBC	Naive Bayes Classifier
OC-SVM	One-Class Support Vector Machine
PerEn	Permutation Entropy
PIR	Passive Infra-Red
R-CNN	Region-based Convolutional Neural Network
RF	Random Forest
RNN	Recurrent Neural Network
SampEn	Sample Entropy
ShEn	Shannon Entropy
SVM	Support Vector Machine
TDNN	Time-Delay Neural Network
URFD	University of Rzeszow Fall Detection
X-HMM	X-factor Hidden Markov Model

Symbols

m	Embedding dimension
r	Tolerance
τ	a time delay
s	a scale factor
x_i	is the i^{th} value in the dataset
$\overline{x}$	is the average of the x-values in the dataset
N	is a finite time series length

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Globally, the population of older adults aged 65 and above is estimated to be over 1.9 billion by the year 2050 [1]. This has a major influence on the healthcare sector, as the cost of older adults care is expected to increase enormously over the years [2, 3]. Additionally, the researches have demonstrated that the number of older adults living alone at home and the number of single-occupancy homes are also growing worldwide, due partly to the high expense of residential care services [4, 5, 6]. The majority of older adults need long-term care and require continuous help in their Activities of Daily Living (ADL). Nevertheless, most older adults prefer to stay in their own homes for as long as possible rather than in residential or care home facilities, to maintain their independence [7]. In order to support older adults to live independently in their own homes, the home environments equipped with appropriate sensors, referred to as Intelligent Environments (IE) or Smart Homes (SH), are used to help support individuals with their daily activities, improve their quality of life, and allow them to stay safely and independently in their own homes [8, 9, 10, 11]. To support independent living for older adults, it is essential to have a means of monitoring and recognising their daily activities and detecting any anomalies in the recognised activities. This would need a reliable system Human Activities Recognition (HAR) [1, 5].

The HAR is the process of automatically detecting human actions from the data collected from different types of sensors. Research related to HAR has devoted particular attention to monitoring and recognising the human activities of a single occupant in a home environment, in which it is assumed that only one person is present at any given time [12, 13, 14, 15, 16, 17, 18]. Recognition of the activities is then used to identify any abnormalities within the routine activities of daily living. Different types of sensors are utilised to detect anomalies in ADLs in a home environment. Most research works so far have considered video cameras and wearable sensors to develop HAR systems in a single-occupancy environment [19, 20, 21, 22]. Video cameras allow the identification of different people moving around the house, which can be considered as the violation of users' privacy [14]. In contrast, other researchers have utilised wearable sensors, such as a wrist-worn accelerometer or gyroscopes for anomaly detection in ADLs [16, 17, 19]. Such devices can provide adequate information about the location of occupants in a home environment and capture human body movements to detect any anomalies. Using wearable sensors would be ineffective if the user forgets to wear them or may take them off when they become uncomfortable [23, 24]. Furthermore, due to improved privacy and reduced cost of equipment, recognising human activities based on ambient sensors is a preferred option.

This chapter presents an introduction to this thesis. The rest of this chapter is structured as follows: The definition of anomaly detection and multi-occupancy environments are explained in Section 1.2. In Section 1.3, an overview of the research describing the schematic of the work proposed is presented. The research questions identified are outlined in Section 1.4. Section 1.5 moves to outline the research aims and objectives, followed by the highlight of major contributions of the thesis in Section 1.6. Finally, the structure of the thesis with a summary of the contents of each chapter is outlined in Section 1.7.

1.2 Anomaly Detection and Multi-occupancy Environments

The challenge of detecting anomalous/surprising/novel patterns has increasingly attracted attention. Anomaly detection is the identification of previously unknown patterns. The problem is particularly difficult because what constitutes an anomaly can greatly differ depending on the task at hand. In a general sense, an unusual pattern significantly different from behavioural routine is referred to as an anomaly (event), and maybe an early symptom of Mild Cognitive Impairment (MCI) or of dementia in older adults [25, 26]. By monitoring the sensor data, important information regarding any irregular (or anomalies) behaviour will be identified. Anomalies are those odd patterns of data that do not match the normal behaviour. Anomalies can be recognised using different anomaly detection techniques.

In many real-life applications, these kinds of patterns are also called events, discordant observations, exceptions, surprises, or outliers. Amongst all mentioned terminology, anomalies, events, and outliers are the most frequently used terms within the context of human behaviour detection. Human behaviour is dynamic, and the behaviour of a person could vary from usual behavioural routines on some days due to some factors such as visits, and the influence of health conditions, irregular sleep and human falls. Anomaly detection aims to detect and identify any abnormal patterns in ADLs in terms of the duration of the event such as irregular sleep and time of occurrence such as human falls or identifying visiting times. For example, the individual who sleeps for a short time period compared to their usual pattern of sleep or the person goes to bed late compared to the usual days, will be detected as the detection of an event.

Most research works related to recognising ADLs have focused only on single occupancy environments, wherein, it is assumed that only one person (i.e. the prominent resident) is present in the home [16, 27, 28]. However, the assumption that home environments are occupied by one person all the time is not necessarily true [18, 29, 30, 31]. For example, it is likely that older adults will receive visits from family members or healthcare workers (referred to as a multi-occupancy environment). Visiting is considered as one of the most important activities for older adults living alone at home, which makes multi-occupancy scenarios are far more realistic [13, 14]. Moreover, Identifying visitors and the time of the visits (such as healthcare visitors) is essential for healthcare management [23]. Therefore, it is important to develop a system with the ability to identify the exact time of a visit without the need for visitors to be asked to carry a tag or wearable device to identify them.

Many current research works acknowledge the challenges of multi-occupancy in HAR [13, 32, 33]. Such challenges are, finding suitable models to represent the data association problem (i.e., the detection of a visitor) and finding an activity recognition system that captures different interactions among residents [14, 34]. Previous studies report that detecting and identifying a visitor in a home environment using only binary sensors is a primary challenge, as binary sensors are not able to provide any information about the personal identity of who triggered the sensor [18, 35]. Reliable anomaly detection in ADLs, or identifying visiting times (e.g. visits made by healthcare workers) is considered as one of the most important components of many home healthcare applications [5]. Thus, existing methods are not able to reliably detect anomalous events in activities and identify the time of visits in the presence of a visitor, therefore generating a high false alarm rate [36].

In many applications, entropy measures are used to detect the irregularities and the degree of randomness in data [37]. Hence, the hypothesis of the research reported in this thesis is to investigate the application of suitable entropy measures to identify anomalies in ADLs, and specifically in a sleeping routine, human falls and in identifying visiting times, in a single and/or multi-occupancy environment. Distinguishing and detecting anomalies in older adults' activities and identifying visitors (the time of their visits) is very important for healthcare management, and helps carers to act early to avert prospective problems. On the other hand, identifying the visit time for older adults may have a significant impact on implementing preventive social distance measures to reduce the transmission of infectious diseases, e.g., Covid-19 virus.

1.3 Overview of the Research

Ambient sensors are often used to monitor and identify daily human activities in an Intelligent Environment or Smart Home. Most of the proposed methods for anomaly detection in ADLs have focused on utilising statistical techniques, including a Hidden Markov Model (HMM) [38, 39, 40] and Random Forest (RF) model [41]. These techniques are used to detect the relationship between the temporal data generated from sensors and identify the pattern of the users' activities. However, it is very difficult to model and recognise large low-level sensory datasets due to the significant network complexity of the outputs from these methods [42, 43]. Moreover, these techniques have some challenges in terms of extracting multiple interacting activities, which could be either cooperative activities and parallel activities [44]. As an alternative to the statistical techniques, computational intelligence techniques such as Convolutional Neural Network (CNN) [8, 45], Support Vector Machine (SVM) [46, 47], and Deep Neural Network (DNN) [48], are widely used to detect anomalies in ADLs. Nevertheless, without significant training, the possible sequences consistent with a particular activity might not be recognised using these techniques [49].

Furthermore, in research related to activity recognition and anomaly detection, most recent studies have focused on detecting anomalies in a single occupant environment [8, 10, 50]. However, living environments are commonly occupied by more than one person. For instance, it is very likely that there will be visitors and/or carers who visit the older adult regularly. Anomaly detection in ADLs in a single-occupancy or multi-occupancy environment requires more investigation to provide a better understanding of the nature of activities and aid older adults to live safely and independently in their own homes. Therefore, it is essential to develop an appropriate method or algorithm that can efficiently detect such anomalies. This can be achieved by using a suitable technique, such as entropy measures, which enables analysis to distinguish between normal and anomalous cases in daily activities with a high degree of accuracy.

In many applications, entropy measures are used to quantify the concept of irregularity and the degree of randomness in a system [37]. Nevertheless, to classify ADL data representing an individual's daily activity routine as either normal or abnormal, entropy measures are considered as a useful measure to discriminate between normal and anomalous cases. This research proposes a novel framework for anomaly detection based on entropy measures through the use of data gathered from ambient sensors and wearable sensors. This is with the aim of incorporating the framework in anomaly detection of daily activities in a multi-occupancy environment.

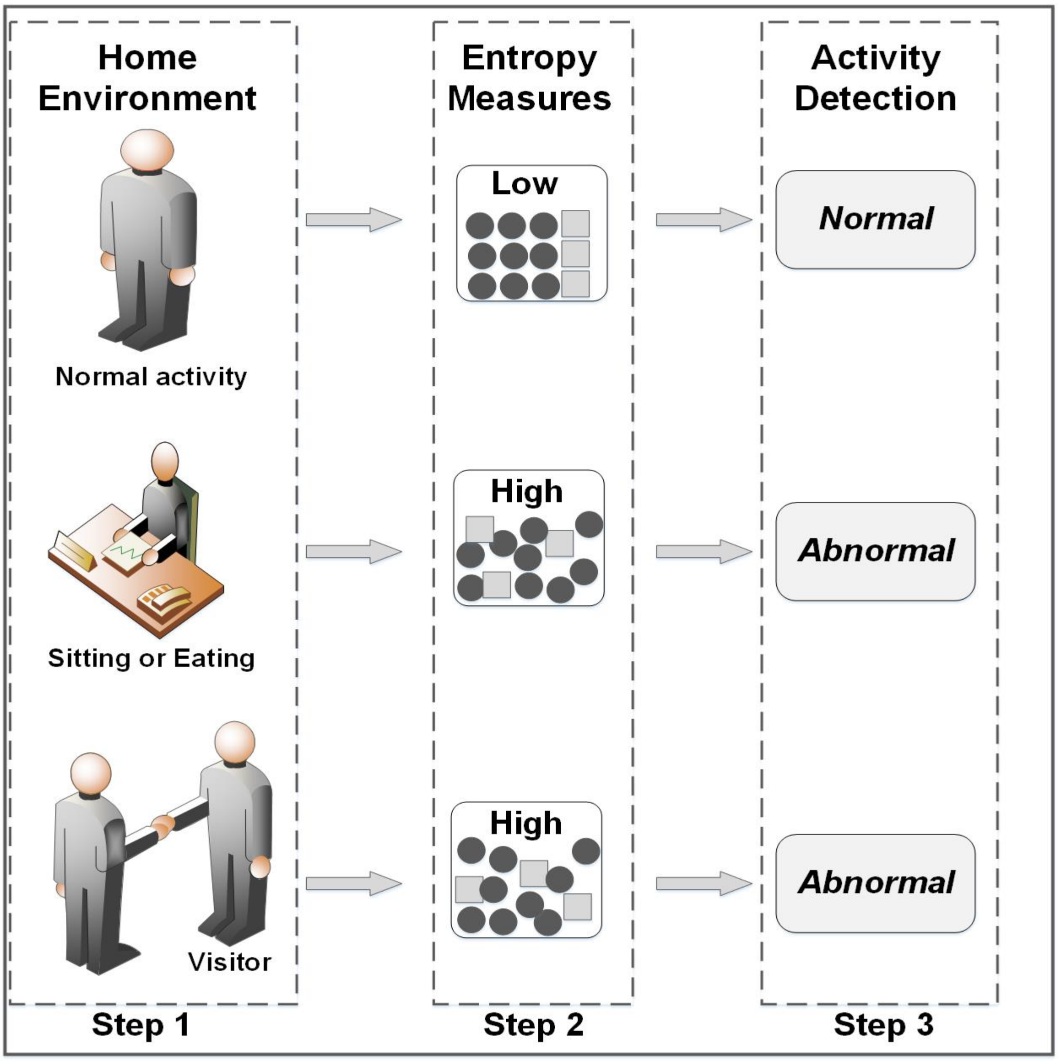


Figure 1.1: Schematic representation of the proposed anomaly detection in activities of daily living.

A schematic representation of the proposed framework in this thesis for anomaly detection based on entropy measures is shown in Figure 1.1. The framework comprises three main steps:

1. Observing daily human activities using sensor networks - Utilising sensor networks to extract and select features representing ADLs. This gives information about the location of a resident in different areas of a home environment.

2. Applying entropy measures in ADLs - Different entropy measures are applied to obtain data for detecting anomalies in the extracted activity patterns.

3. Activity detection as either normal or abnormal - Investigate the effectiveness of different entropy measures in detecting anomalies in ADLs in a single-occupancy and multi-occupancy environment. Furthermore, a pre-set threshold is used, based on the standard deviation of the occupancy data in conjunction with several entropy measures, for discriminating between normal and anomalous cases in daily activities.

1.4 Research Questions

Following the research overview, the main research question addressed in this thesis is to investigate the effectiveness of different entropy measures in detecting and identifying various types of anomalies in daily activities. In particular, this study attempts to answer the following questions:

- How to extract useful features from low-level binary data representing ADL of a single-occupancy or multi-occupancy environment?
- Is it possible to use entropy measures to detect and identify anomalies in a person's ADLs, specifically in sleeping routine?
- How a resident's daily pattern can be used with several entropy measures to decide whether there is an anomaly in their activities or not?
- Can entropy measures be utilised for detecting human falls in daily activities, solely based on the information gathered from a wearable motion-sensing device?
- Can entropy measures be used to detect anomalies in ADLs in a multi-occupancy environment, solely based on information gathered from ambient sensors? Most existing methods for anomaly detection rely on a single-occupant environment where only one individual is monitored. This is often not true.

- Can the proposed solutions be tested and validated on data obtained from real-world environments?

To address the above questions, the following section outlines the aim and objectives of this research.

1.5 Technical Objectives

To support independent living, it is essential to recognise routine ADL and distinguish any abnormality with the recognised activities. This would require accurate and reliable HAR [12, 51]. Anomaly detection in ADLs has remained a significant challenge for researchers in recent years. Considering the complexity and uncertainty associated with human activities, the existing outlier detection techniques provide some limited reliabilities in detecting the anomalous events in ADLs, particularly due to ignoring the changes in individuals' routine [52].

The aim of this research is to investigate the effectiveness of different entropy measures in detecting and identifying various types of anomalies in ADLs. This research tries to find an acceptable solution that can be used to detect and identify anomalies in ADLs in a single-occupancy and multi-occupancy environment. As a starting point for detecting anomalies in ADLs, the investigation of the effectiveness of entropy measures initially focuses on a single-occupant environment, when only one individual is monitored, and their activities are categorised. Then, the research investigates the effectiveness of entropy measures for anomaly detection in a multi-occupancy environment. Furthermore, the entropy measures are not only used to detect anomalies in ADLs but also to identify potential causes of anomalies, and to distinguish anomalies in ADL data (here, irregular sleep in the resident's activity and visitors). The proposed anomaly detection framework based on entropy measures will be applied to several datasets representing the ADLs of a single-occupancy and multi-occupancy environment.

In order to accomplish the aim of this research, the following research objectives have been identified:

1. Use a low-cost, non-intrusive ambient sensory device-based system to obtain

a dataset that represents ADL from single-occupancy and multi-occupancy smart environments and extract the required numerical features from raw data to be used as input vector sequences for the entropy measures.

2. Investigate the existing anomaly detection approaches and their suitability for detecting anomalies in ADLs.

3. Propose a novel anomaly detection method based on entropy measures to detect different anomalies in daily activities (e.g., irregular sleep in the resident's activity and falls).

4. Create a resident's daily pattern to be utilised with several entropy measures for identifying and detecting anomalies in the resident's activities.

5. To investigate the effectiveness of different entropy measures for detecting anomalies in ADLs in a multi-occupancy environment, solely based on information obtained from ambient sensors.

6. Compare the performance of different entropy measures to assess the most appropriate method for detecting anomalies in ADLs based on information gathered from different smart environments.

1.6 Major Contributions of the Thesis

The major contributions of the work presented in this thesis are summarised as follows:

- An extensive literature review of the state-of-the-art on anomaly detection, which encompasses approaches proposed and validated results from experiments.

- A novel framework based on different entropy measures for anomaly detection in ADLs where anomalies are diverse and normal samples are relatively homogeneous.

- A robust investigation into the use of entropy measures for human fall detection in daily activities, solely based on the information gathered from a wearable motion-sensing device.
- The proposed entropy measures are used not only to detect anomaly days but also to identify potential causes of anomaly days based on the calculation period of entropy measures.
- By finding the maximum entropy values in normal daily activities, it is possible to detect abnormal human behaviours in ADLs in completely unseen data.
- To identify the possible causes of anomalies (here, irregular sleep and identifying visiting times), the main door sensor along with entropy measures is used to confirm the time of the visitor's presence.
- Investigating the effectiveness of different entropy measures in distinguishing activities in a multi-occupancy home environment solely based on the information collected from motion detectors (e.g. Passive Infra-Red (PIR)) and door entry sensors. Once the presence of the main occupier is distinguished from others, the existing activity recognition and abnormality detection processes could be applied for the main occupier.
- Testing and evaluating the proposed entropy-based approach using several different datasets gathered from real home environments representing ADL for a single or a multi-occupancy. Unlike entropy measurements, most machine learning techniques require a large amount of training data and classification time.

The outlined contributions of the thesis are addressed in different chapters of this thesis. A summary of these chapters is presented in the following section.

1.7 Thesis Outline

This thesis consists of seven chapters. Figure 1.2 shows the structure of the thesis with an indication of how the chapters are linked. The idea behind this

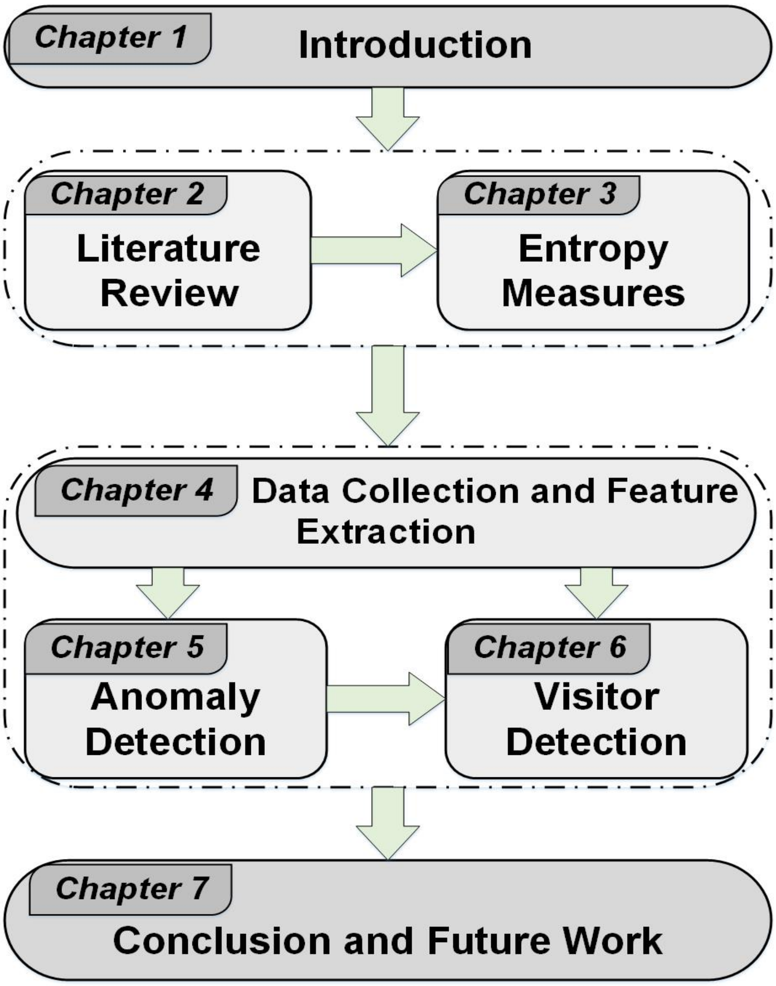


Figure 1.2: Thesis structure showing the organisation of the chapters and their respective dependencies.

figure is to give readers an overview of the organisation of the thesis and a direction on how the chapters are grouped. The summary of the contents of this thesis are presented as follows:

Chapter 2: Literature Review - This chapter gives a comprehensive review of the relevant literature in the field of anomaly detection in activities of daily living. The main areas that are covered are anomaly detection in daily activities using statistical methods and computational intelligence techniques, human fall detection using different approaches and algorithms, activity recognition, and the challenge of data association in a multi-occupancy environment. In particular, the literature focuses on utilising available technologies for detecting different anomalies in ADLs. Furthermore, a summary of the literature review is presented to identify the research gaps and highlight how this research differs from previous research works. Finally, conclusions drawn from the review are presented.

Chapter 3: Entropy Measures for Anomaly Detection - This chapter presents an overview of the entropy measures that are more relevant for measuring the complexity in time series of data gathered from an IE. Specifically, the explanation of certain entropy measures that are used for anomaly detection in this thesis are provided. These entropy measures are applied later on in Chapter 5 and 6 to propose approaches for anomaly detection in daily activities in a single-occupancy and multi-occupancy environment. The methodology proposed for anomaly detection in daily activities in this thesis is also introduced in this chapter.

Chapter 4: Data Collection and Feature Extraction - This chapter presents an overview of intelligent environments, including sensor networks that are used for gathering information representing ADLs of a single-occupant or multi-occupants. Two different environments, including real and simulated environments, are also explained in detail to validate and test the results of the proposed anomaly detection. Further details about the pre-processing and feature extraction from raw data are presented in this chapter.

Chapter 5: Anomaly Detection in Activities of Daily Living - This chapter is an extension of the explanation provided in Chapter 3, for proposing a novel framework based on different entropy measures for anomaly detection in ADLs. The chapter aims to investigate whether entropy measures can be used for anomaly detection in daily activities in a single home environment. The chapter starts with proposing a method for anomaly detection in ADLs, specifically in sleeping routine, solely based on data gathered from low-cost, non-intrusive ambient sensors. Furthermore, a novel method based on Fuzzy

Entropy measure is investigated to detect and distinguish human fall from other activities. A novel method based on different entropy measures to detect anomalies in a resident's activity in the presence of a visitor, solely based on information gathered from ambient sensors is also proposed in this chapter. In this chapter, experiments are conducted utilising the datasets obtained for the research in this thesis to test and evaluate the proposed anomaly detection method. Experimental results are also presented to demonstrate the effectiveness of entropy measures in detecting anomalies in the resident's activity. To evaluate the proposed anomaly detection carried out in this research, the results obtained by applying entropy measures are compared to the state-of-the-art approaches reviewed from existing research. The chapter concludes that the proposed anomaly detection based on entropy measures is a promising technique to distinguish between normal and anomalous events in a resident's activity in the home environment.

Chapter 6: Visitor Detection in Multi-Occupancy Environments - In this chapter, entropy measures are employed to identify visitors in multi-occupancy environments, solely based on the information collected from motion detectors (e.g. PIR) and door entry sensors. Furthermore, this chapter investigates the impact of changing the values of an embedded dimension, m, and tolerance, r, as parameters required to calculate the named entropy measures. Afterwards, to evaluate the robustness of the proposed entropy measures for visitor detection, the main door sensor is used along with entropy measures to confirm the time and duration of the visit. Also, experiments are conducted using the datasets obtained for the research in this thesis to test and evaluate the proposed anomaly detection method. The performance of proposed entropy measures for visitor detection are compared to the state-of-the-art approaches reviewed from existing research. The chapter concludes that the anomaly detection by entropy measures can be confirmed with door sensors data, particularly for identifying the exact visiting times.

Chapter 7: Conclusion and Future Work - This chapter provides a summary of the findings of the research conducted in this thesis. The major findings obtained in this thesis are discussed with a reflection on the research questions identified in this chapter. Following the summary of the achievements, the chapter also presents recommendations for applications of the work in this thesis and possible areas of future work in monitoring the activities of daily living and detecting anomalies in such activities.

Chapter 2

Literature Review

2.1 Introduction

To support older adults with their independent living, it is essential to monitor and recognise routine Activities of Daily Living (ADL) and identify any abnormality in their daily activities [8, 9]. An automated monitoring system to identify abnormalities within the ADL would require an accurate recognition of human activities. Hence, Human Activity Recognition (HAR) has gained increasing attention in recent years [6, 53, 54]. The HAR is relevant to many applications, such as healthcare and assisted living. Many data mining and machine learning algorithms are widely employed for anomaly detection in daily activities [50, 55]. In this regard, several research works have been conducted on ways to discriminate between normal and anomalous cases in ADLs, using different techniques. To justify the intent of the work in this thesis, it is essential to review the current state-of-the-art related to detecting anomalies in daily human activities. This chapter is focused on the review of the relevant literature related to detecting anomalies in ADLs, human fall detection, and the most common approaches used for anomaly detection. Moreover, the identification of a visitor in a single-occupancy home environment (represented as a multi-occupancy environment) is also reviewed.

This chapter is structured as follows: Section 2.2 provides an overview of anomaly detection in ADLs based on different approaches. Section 2.3 reviews the existing literature studies on human fall detection methods. In Section 2.4, related work in the context of activity recognition and the challenge of data association in multi-occupancy are reviewed, followed by a summary of the literature review in Section 2.5. Section 2.6 follows on from the review of previous researches to identify the research gaps and highlights how this work differs from previous research works. Section 2.7 summarises the chapter.

2.2 Anomaly Detection in Activities of Daily Living

Anomaly detection in daily activities is a challenging task, as it depends on a specific context and the unconstrained variability of practical scenarios. Anomaly, also known as an abnormality, can be defined as any significant change in usual behavioural routine and can be an early symptom of Mild Cognitive Impairment, or of dementia in older adults [26]. Additionally, anomalies in ADLs such as interrupted sleeping, or performing less active tasks during the day, can be detrimental to their well-being [56, 57].

Several anomaly detection algorithms and techniques are proposed to solve challenges in different domains, including human behaviour, computer networks, image processing, medical, etc. [46, 58, 59]. Also, several research studies have been carried out on the detection of anomalies in ADLs utilising diverse techniques. In the following sections, some related literature studies concerning anomaly detection techniques in ADLs are reviewed.

2.2.1 Statistical Techniques

Several research studies have been carried out on the detection of various types of anomalies utilising different statistical techniques, including Hidden Markov Model (HMM) [38, 40] and Random Forest (RF) approaches [41]. An HMM-based technique was developed in [38] to detect anomalies in daily activity sequences. Their experiments were based on data generated synthetically from a real-world dataset. The researchers have shown that their proposed model can detect anomalies in ADLs with an accuracy of 95.10%.

Similarly, in [40], the authors proposed an anomaly detection approach based on a dynamic Markov model. The performance of their proposed anomaly detection approach was based on both synthetic and real-world data. Their research aimed to address the challenge of reducing false alarms when compared to existing techniques. The experimental results obtained from this work indicate that the proposed approach achieved the highest true positive rate and lowest false alarm rate compared to other methods mentioned in their literature review. Nevertheless, in both [38, 40], the authors have provided few details about the usage of the synthetic data and how the work was conducted.

A study reported in [60] utilised HMM and Viterbi algorithm for real-time detection of sleep anomalies. Experiments were conducted based on a dataset generated by the authors. However, the reader is not provided with enough information about how the data synthesis was conducted. It is also reported that some further work is required to improve the proposed method by using the Pittsburgh Sleep Quality Index (PSQI).

The researchers in [61] proposed an HMM for smart home anomaly detection. The idea of the research was to tune the HMM parameters for maximising the probability of finding the anomalies by learning the typical behaviours in a smart home. They tested and evaluated their proposed method based on a synthetic dataset, and the proposed method achieved an accuracy of 97%. However, the authors also suggest that some further work is required to improve the proposed method by testing the performance of HMM in case of bigger datasets. Likewise, the researchers in [62] proposed an approach based on Two-dimensional HMM for anomaly detection in ADLs. They trained and tested their proposed model based on dataset, which is split into 70% as a training set and 30% as a testing set. The experimental results obtained from this study show that the proposed model achieved an accuracy of 92.25%. However, the authors have not provided enough details about the usage of the data and how the work was conducted.

2.2.2 Computational Intelligence Techniques

As an alternative to statistical techniques, computational intelligence techniques are widely utilised to detect and identify anomalies in daily activities. In [8], a combination of Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) is utilised to detect simulated anomalies in ADL data related to dementia in smart homes. The main dataset is used as training data for the normal class, whereas the synthesised anomalous dataset is utilised as training data for the anomalous cases. The dataset is then fed into a CNN to be trained, while LSTM is utilised to learn the activity sequences of the behavioural routine. The authors tested and evaluated their model based on two different datasets. The results obtained from their research indicate that CNN with LSTM achieved better results compared to the state-of-art methods mentioned in their literature review, with an accuracy of 89.72%. However, the authors also reported that their method could not detect every type of anomaly, such as using too much soap or leaving kitchen appliances on when they are not needed.

A relatively new research study presented in [50] have proposed an approach based on LSTM to detect anomalies in a sequence of daily activities in a home environment. The idea of their research was to compare the performance of LSTM and HMM in anomaly detection in ADLs under different sizes of the training sets. Their experiments were evaluated based on the "Aruba" dataset publicly available from the Centre of Advanced Studies in Adaptive Systems (CASAS) at Washington State University [63]. The results obtained by using these two models indicate that both LSTM and HMM achieved the same accuracy of 87.50%. However, the authors also state that the LSTM method has some limitations in terms of the requirement of input of 5 initial activities of a sequence. This means the method cannot detect anomaly in these 5 initial daily activities as the layer of LSTM utilised is unidirectional.

Novelty detection algorithms have also been applied to distinguish between normal and anomalous cases in ADLs, including a Support Vector Machine (SVM) [46]. Researchers in [64] and [65] have applied One-Class Support Vector Machine (OC-SVM) for anomaly detection in ADLs. This approach has also been utilised for unsupervised outlier detection in brain multiparametric magnetic resonance imaging [66]. Likewise, OC-SVM has been applied for anomaly detection in time-series data [57], and applied with Electroencephalogram (EEG) data for detecting seizures in humans [67].

2.2.3 Other Techniques used in Anomaly Detection

Several other techniques not mentioned above are used to distinguish between normal and abnormal cases in ADLs. For example, in [68], the authors proposed a Consensus Novelty Detection Ensemble (CNDE) approach for anomaly detection in daily activities. A novel version of the Gated Recurrent Unit (GRU), called Single-Tunnelled (SiTGRU), was proposed in [69] for anomaly detection and generalisation in videos. They trained and tested their proposed model based on three well-known video anomaly detection datasets. The researchers indicate that their proposed model achieved better performance than standard recurrent networks. However, the proposed model required some further work, that is, fusing it with other variants of recurrent and deep networks in order to improve the model introduced.

In [36], the authors proposed a system named "Holmes" for anomaly detection in ADLs, utilising Density-Based Spatial Clustering of Applications with Noise (DBSCAN). The idea of their research was to address the challenge of reducing false alarms compared to existing techniques. The evaluation of their system was based on two public datasets from the CASAS repository, which do not have ground truth for anomalies in ADLs. The experimental results obtained from this study show that their system decreases false positives and false negatives by 46% and 27%, respectively. Similarly, the researchers in [70] presented an approach based on the DBSCAN clustering algorithm in order to detect anomalies ADLs performed in a smart home.

A relatively new research work proposed a novel method, the Positive-Unlabelled deep metric learning method for anomaly detection (PUMAD), which effectively identifies various anomalies [10]. They tested and evaluated their proposed method based on two types of datasets. Their results show that the PUMAD achieves state-of-the-art performance. However, the authors also state that the PUMAD method has some limitations in terms of its potential unsuitability for normal data that has a lot of classes (clusters). It is also reported that some further research is required to improve the proposed method by extending the study to a more generalised PU anomaly setting, as a multi-class anomaly detection setting.

A summary of the related literature studies in the area of anomaly detection in daily activities is provided in Table 2.1. The table summarises the existing related research studies for anomaly detection in ADLs in the context of the publication year, the approaches used, dataset name, the type of sensors used as well as the overall system accuracy.

2.3 Human Fall Detection

Falls are considered as one of the greatest risks and a fundamental problem in healthcare for older adults living alone at home [71]. The number of older adults living alone in their own homes is increasing worldwide, and this causes an increase in the demand for healthcare services [72]. Therefore, it is important to develop an accurate system with the ability to detect human falls during daily activities.

To support older adults with their independent living, assistive technologies, such as automated fall detectors are utilised to assist and support them to live safely in their own homes [73]. Several research studies have been carried out on detecting human falls during daily activities, using different types of sensor. These studies can be classified into three main categories, namely; ambient sensor-based [74, 75, 76], vision-based methods [77, 78, 79] and wearable sensor-based [45, 80, 81]. Ambient sensors such as pressure sensors are installed on the floor. They are used to capture vibrations and sound that detect the presence of a person [82]. Alternatively, several studies have been carried out based on computer vision for human fall detection utilising single [83], multiple [84, 85], and omni-directional [86] cameras. Recently, depth sensors such as Microsoft Kinect [47, 87] have been utilised for human fall detection. The Kinect sensor is a motion-sensing device which integrates a Red Green Blue (RGB) camera and a depth sensor to capture moving objects in 3D [87]. On the other hand, several works have utilised wearable sensors, such as a wrist-worn accelerometer or gyroscopes to detect human falls during ADLs [45, 88, 89]. These types of sensors are widely utilised to capture human body movements. Thus, analysis of the movement of the human body through an accelerometer or gyroscope allows for detection when there is a fall [90].

Based on the reviewed literature, different approaches and algorithms are used for human fall detection, which are divided into two main categories: statistical techniques and computational intelligence techniques. In the following sections, these two techniques and other techniques used for detecting human falls during daily activities are reviewed.

2.3.1 Statistical Techniques

Detecting human falls in a home environment is still a significant challenge for researchers. In recent years, research has been carried out on detecting human falls using statistical techniques, including HMM [39] and Hierarchical Hidden Markov Model (HHMM) [91]. In [39], the authors proposed a model, namely three X-Factor Hidden Markov Models (XHMMs), for human fall detection using a wearable device. The idea of their study was to detect unseen falls by modelling transitions between normal daily activities to train an HMM and adding a new state to model unseen falls. Their experiments were based on two human activity recognition datasets collected using an accelerometer and gyroscope. The experimental results obtained from this study show that two of the XHMM models can detect human falls with an accuracy of 96.6%.

The researchers in [91] proposed an HHMM based video analysis method for fall detection during daily activities. They used HHMM with two layers; in the first layer, two states are utilised, one related to an upright standing pose and the other to a laying pose. The object of their research was to study the relationship between angle sequences in the 3D world and their projection onto the image plane. The results obtained from their research indicate that the overall system could correctly detect 98% of human falls in a home environment. Similarly, in [92], the authors proposed a system based on HMM for temporal detection of social interactions. The idea of the research was to detect intervals where an individual or social activity is occurring. The performance of their proposed detection approach was based on the publicly available RGB-D dataset. It is also reported that the proposed system achieved an accuracy of 85.56%. However, the authors also suggest that some further work is required to improve the proposed method by using online learning techniques to improve the classification over time.

A study reported in [93] proposed an HMM-based fall detection system that can automatically detect falls using a single motion sensor for real-life home monitoring scenarios. The HMM is trained and used to detect falls based on acceleration signal data gathered from motion sensors. They tested and evaluated their proposed method based on both synthetic and real-world data. The results of their study show that when HMM is applied on both datasets, a sensitivity of 99.2% and a positive predictive value of 98.1% were achieved for their first dataset, whereas 100% sensitivity and 78.6% positive predictive value have resulted from their second dataset. Whereas the proposed model demonstrated a promising result, there are some constraints to the study; however, the data in this study is a snapshot of one event, not many events from one subject over time.

In [94], the researchers proposed a new method using acceleration data and HMM to detect fall events. The idea of their research was to extract Feature sequences from the acceleration data to be used as a sequence of observations to train an HMM of fall detection. They tested and evaluated their proposed method based on a synthetic dataset, and the proposed method achieved detection rates of 91.7% sensitivity and 97.2% accuracy. However, the authors also highlight some limitations of their study in terms of the dataset, which the training samples are from simulated motion process, but not falls in real practice.

2.3.2 Computational Intelligence Techniques

As an alternative to statistical methods, computational intelligence techniques, such as the SVM [47, 87, 95], Recurrent Neural Network (RNN) [89], DNN [48] and CNN [45, 71, 83, 96, 97] are widely used to detect human falls. An SVM was utilised in [47] to distinguish a falling pose from normal daily activities using machine vision from RGB-D images. Their experiments were evaluated based on the publicly available University of Rzeszow Fall Detection (URFD) dataset [87]. The dataset contains 30 videos capturing different cases of falling and 40 videos demonstrating ADLs. The experimental results obtained from this study show that the proposed approach outperformed similar studies where images or accelerometers were utilised, achieving a sensitivity and specificity of 100% and

97.5%, respectively. However, there are some limitations, which includes the proposed model failing to detect falling on a bed or sofa, as well as the inherent limitations of the Kinect camera.

A relatively new research study, [98], has proposed a novel camera-based real-time automated human fall detection in a home environment using SVM. The idea of the research was to detect the moving person in the home and utilise features of the bounding ellipse, then apply SVM to classify the activities into fall and non-fall events. The authors evaluated their model based on the publicly available URFD dataset, and the proposed method achieved detection rates of 98.15% sensitivity and 97.1% specificity.

Recently, several research studies have been conducted to detect human falls in daily activities employing deep learning techniques. A study reported in [99] used a CNN based on dynamic motion and shape variations to detect older adults' falls during daily activities. They utilised a new vision system based on novel two-stream CNNs for older adult fall detection. Firstly, the human image is extracted based on person recognition and background subtraction. Then, History of Binary Motion Image (HBMI) is integrated into the first stream, distinguishing human shape variations. Experiments were conducted based on two publicly available datasets, which are the Multiple Cameras Fall (MCF) dataset [100] and the URFD dataset. It is also reported that the proposed system achieved a sensitivity and specificity of 100% and 92.5%, respectively. However, the authors also suggest that some further work is required to improve the proposed method by utilising depth cameras and using Region-based CNNs (R-CNN) to improve the shape-based stream by extracting features from different body shapes.

In [89], a fall detection method is proposed based on an RNN method, which can process and encode the inherent information contained in sequential data. The authors used a dataset gathered from an accelerometer placed near the pelvis area of the user, and depth cameras. The results obtained from their research indicate that the proposed method achieved better results compared to the previous methods mentioned in their literature review, with an accuracy of 98.57%.

Accelerometer-based human fall detection utilising CNNs is proposed in [45]. The authors evaluated their approach using three open datasets and compared the results to other methods. The experimental results for this approach showed that around 99.86% of human falls could be detected. The authors also suggest that some further work is required to evaluate other deep learning techniques for human fall detection, improve the proposed method to detect multi-class events and distinguish various activities.

2.3.3 Other Techniques used in Human Fall Detection

There are some other techniques not mentioned above, utilised to distinguish and detect human falls during daily activities [101, 102]. For instance, the study in [103] proposed human fall detection from a depth image based on the velocity and position of the subject. The research aimed to detect the potential fall activity and the position of the subject to confirm human fall. The results obtained from the research showed that the proposed system can correctly distinguish human falls from a non-fall with an average accuracy of 93.94%. Despite the high accuracy, the authors suggested that the proposed system could be further improved by focusing on the acceleration of joints together with the velocity.

In [104], a novel slow feature analysis-based framework for distinguishing human fall from normal daily activities is proposed. Their experiments were evaluated based on two different publicly available multiple-camera fall dataset [105] and SDUFall datasets [106]. The results obtained from their study showed that the proposed method achieved better results compared to the previous methods mentioned in their literature review, with an accuracy of 81.33%. In addition, in [107], a Single Shot Human Fall Detector (SSHFD) is proposed to detect human fall from a single image. They tested and evaluated their proposed method on the public multiple-camera fall dataset [105] and the Le2i fall dataset [108].

Table 2.2 provides a summary of the related literature research studies in the area of human fall detection. The table presented summarises the existing publications in the context of the publication year, the approaches used, dataset name, the type of sensors used as well as the results obtained.

2.4 Visitor Detection

Distinguishing and detecting a visitor in a single-occupancy home environment (represented as a multi-occupancy environment) based on ambient sensors is still a significant challenge area, as most of the sensors do not provide any information regarding the personal identity of who triggered the sensor. Therefore, the use of wearable sensors, visual sensors and video cameras have become the norm to monitor ADLs in a multi-occupancy environment [18, 23]. However, few studies have focused on the detection and identification of multi-occupancy activities using ambient sensors, especially those with binary sensors [51, 109, 110]. Several research works have been carried out on detecting and identifying multi-occupancy and monitoring activities by using different techniques and algorithms [34, 32, 111, 112].

In this thesis, the reviewed techniques are classified into two main categories, which are statistical techniques and computational intelligence techniques presented in Section 2.4.1 and Section 2.4.2, respectively. Other techniques not included in these two main groups are reviewed in Section 2.4.3. A brief review of the related work in the context of the challenge of data association in the multi-occupancy environments are also presented in Section 2.4.4.

2.4.1 Statistical Techniques

Most of the research which has been conducted in the context of activity recognition in multi-occupancy has utilised statistical techniques. These techniques are utilised to detect the relationship between the temporal data generated from sensors and identify the pattern of the user. Graphical probabilistic models are the most popular techniques utilised to identify human activity recognition. In a recent survey [35], the authors provided an overview of the latest investigations on activity recognition in multi-occupancy environments. Their survey includes the existing approaches and current practices used for activity recognition, such as an HMM, Naive Bayes Classifier (NBC), Conditional Random Field (CRF), and Dynamic Bayesian Network (DBN). Moreover, it outlines data association and interactions between occupants as the main challenges in a multi-occupancy environment. Some commonly utilised statistical techniques for detecting and identifying multi-occupancy in a home environment are reviewed in the following sections.

2.4.1.1 Hidden Markov Model Based Techniques

HMM-based techniques are widely applied in many studies to identify the activities of a resident from sensor data and distinguish the activities within a multi-occupancy and identify whether the environment is utilised by one or more than one person. These techniques are utilised to detect relationships between temporal data generated by the sensors and identify the pattern of the user.

There are many published papers related to pattern recognition that conducted their research to detect HAR in a home environment using a range of different machine learning techniques, including HMM [113, 114, 115]. For example, in [13], the Factorial Hidden Markov Model (FHMM) and Nonlinear Bayesian Tracking method are applied and compared for tracking and recognising human activity. The FHMM is used to model two separate Markov chains corresponding to two users, whereas Nonlinear Bayesian Tracking is used to break down the observation area into the number of users. The authors indicated that the Nonlinear Bayesian Tracking method performs better than FHMM (the performance of Bayesian Tracking was 67.9%, while the performance of FHMM was 59.5%).

In [116], the researchers proposed an unsupervised method for detecting visits as abnormal activity in the homes of older adults. They utilised a method based on a Markov Modulated Multidimensional non-homogeneous Poisson Process (M3P2) to model daily and weekly characteristics, as well as to distinguish between regular and irregular visits in a home environment. The results obtained from their research demonstrate that the M3P2 method performs better than the Markov Modulated Poisson Process (MMPP). They also state that the performance of M3P2 in terms of precision was 64%, while the performance of MMPP was 56%. However, the proposed model generates a high false alarm rate, which reduces precision.

A relatively new research work [12] has proposed a new model based on

MMPP, which is an unsupervised method that models regular activity patterns and detects visits in homes of older adults living alone. The ambient sensors are installed in specific locations to cover most of the movement without affecting the routine activities of the occupier. Their experiments were based on the data obtained from two apartments using different sensor networks. The results of their study show that when MMPP is applied on both datasets, a recall of 78.4% and a precision of 74.9% were achieved for their first dataset, whereas 80.1% recall and 84.2% precision have resulted from their second dataset. The only issue, however, with this method is the difficulty in processing a large amount of low-level data such as the data gathered from ambient sensory devices. In [109], the authors investigated the challenge of detecting multi-occupancy in a home environment with different sensor networks using HMM. The authors evaluated their model based on data obtained using a binary sensor in a living lab. Likewise, in [14], the authors investigated the challenge of modelling multi-occupancy activities. Specifically, they explored different models based on HMM, known as CL-HMM to attempt to deal with cooperative activities and parallel activities in a multi-occupancy environment. The authors have evaluated their model based on a CASAS multi-occupancy dataset [117]. Whereas these methods demonstrate a promising result, there are some constraints to the study; however, since the collected data was limited to only one room and the number of sensors used was small.

Some other research works have addressed the challenge of identifying multi-occupancy activities utilising wearable sensors [19] or video sensors [20]. For example, in [29], researchers investigated the challenge of recognising multi-occupancy activities utilising wearable sensors in a home. Their idea was to study two probabilistic temporal models; the Coupled Hidden Markov Model (CHMM) and Factorial Conditional Random Field (FCRF) to model and classify multi-occupancy activities. Their proposed model was tested and evaluated using a dataset obtained from two subjects over two weeks. The results obtained by utilising these two models showed that CHMM performs better than FCRF. Nevertheless, the authors also highlight some limitations of their study in terms of the dataset, which was collected in a mock scenario, rather than being conducted in a real home environment.

2.4.1.2 Naive Bayes Classifier

The researcher in [51] investigated two different models in which multi-users can be detected in a home environment with various sensor networks using a NBC and HMM. The idea of this study was to detect a visitor in an office environment equipped with binary sensors and a video camera to record the visits to the office. The results obtained using these two models indicated that the HMM performs better than the NBC, with an accuracy of 92% and 83%, respectively. While the proposed method demonstrated a promising result, there are some constraints to the study; the data obtained was limited to only one room, and the number of sensors utilised was small.

2.4.2 Computational Intelligence Techniques

As an alternative to statistical methods, computational intelligence techniques are widely utilised to recognise the ADLs in a multi-occupancy environment. The following sections summarise some of these techniques.

2.4.2.1 Support Vector Machine

The SVM is widely utilised for detecting and distinguishing multi-occupancy based on data gathered from a home environment and detecting the users' abnormal activities. In [110], SVM has been utilised to identify the periods when visitors are present in a home. They have used dwell time, the number of transitions between main living places (dining room, kitchen, living room, and bathroom), and the number of sensor firings as features in the SVM. Their model was evaluated on data obtained from only motion sensors in a living lab. Some limitations are however evident, such as the visits not being recorded overnight. Likewise, the researchers in [23] proposed a system based on SVM to detect visitors in the home environment using wearable devices and an ambient sensor network. Their experiments were based on the data gathered from a Swiss-Korean project on healthcare monitoring of older adults living alone in a home environment. The results obtained from their study show that the method can correctly detect 58% - 83% of visits in a home environment.

However, the main challenge facing the authors is that they did not have fully annotated data to label every visit in the life of the older adults at home.

2.4.2.2 Deep Learning Techniques

Machine learning algorithms have been utilised to recognise the ADLs in a multi-occupancy environment in recent years. The most common techniques are Deep Convolutional Neural Network (DCNN) and CNN. For instance, a study reported in [7] has used a novel RGB activity image based on a DCNN classifier for the unobtrusive recognition of multi-occupancy activities of older adults. They have used a labelled open dataset gathered by environmental sensors (i.e., PIR sensors and temperature sensors) in a Cairo testbed, which is one of the testbeds from the CASAS Project [118]. It is also reported that the dataset is pre-processed with a sliding window, RGB activity image conversion steps, and activity segmentation. The experimental results obtained from this study demonstrate that the proposed model outperformed the previous methods mentioned in their literature review, with an accuracy of 95.2%. However, the authors also suggested that further work is required to classify more intertwined and more complex activities using real-life long-term multi-occupancy activity recognition.

2.4.3 Other Hybrids Techniques used in Multi-occupancy Activity Recognition

There are other techniques not mentioned above, used to identify and detect activities in a multi-occupancy environment. For example, researchers in [15] applied a platform based on active learning techniques, known as Smart ADL Recogniser and Resident Identifier in Multi-resident Accommodations (SARRIMA), to recognise ADLs in multi-resident environments by utilising only passive sensors. They used semi-supervised algorithms to detect ADLs in order to reduce the trade-off between the training time and data labelling. The SARRIMA is used to solve both the problem of data association by using an identification module of the resident, and the problem of activity recognition by utilising the module of ADL recognition. The result of this approach showed that around 96% of the activity instances can be detected. Also, it can be used to identify the activity of residents without utilising RFID tags or wearable sensors. However, the disadvantage of this research is that it lacked sufficient data to robustly test the model, as it was tested only on data obtained from two people in a smart home.

In [119], the authors investigated the essential problem of recognising activities of both a single-occupant and a multi-occupant environment from ambient sensors, by using Emerging Patterns (EPs) to distinguish between the activities of a single person and multi-users in a smart environment. The datasets in their research were collected from two people through a period of two weeks in a smart home environment. The authors also emphasised that EPs can be used to recognise the cooperative activities in multi-occupancy environments. Bluetooth enabled smartphones were used in [120] to identify and track a resident in a smart home. The research also shows that the solution based on Bluetooth technology was the best option to achieve the goals of this study rather than Wi-Fi because of the lower power drain on mobile devices. Likewise, the researchers in [121] presented a wireless distributed pyroelectric infrared sensor network and a novel method utilising an Empirical Mode Distributed (EMD) algorithm to identify and track multi-occupancy in a home environment.

Many published papers addressed the challenge of recognising and identifying multi-occupancy activities using wearable sensors. The researchers in [18] propose an automatic multi-occupancy activity labelling approach in a smart home for resident localisation, using wearable sensors and a Bluetooth Low Energy (BLE) technology. The BLE devices are used as the tag to localise and label their activities in a multi-occupancy intelligent environment. The experiments were based on data obtained from a real smart home. The smart home is equipped with three types of sensors, including five PIR sensors, two switch sensors, and one power sensor, which they used to monitor the user' activities. The results obtained from their research indicate that BLE approaches can achieve high accuracy. However, it is also mentioned that forgetting to wear the tag is considered as one of the main problems with this research which affects the performance of activity labelling accuracy.

Similarly, in [122], it was shown that human identification in a multi-occupancy environment could be detected using three sensing/communication modules, including a PIR sensor, an ultrasound array, and a BLE device. The PIR sensor is used to detect the user's movement in different locations in a home environment, while the ultrasonic array module is used to detect the moving user's height. Then, the BLE mode is utilised to communicate the data from the PIR and ultrasound array to the data server. The study demonstrated that there is a limitation of the proposed model in distinguishing residents if they have similar heights.

A recent survey by [21] presents an overview of wearable sensors and bespoke sensors' usage in activity recognition in multi-occupant environments. The paper highlights the cooperative interaction activities and complex activity recognition in smart homes. The authors of [30] proposed a hybrid approach to recognising complex ADLs using a smartphone-based sensor. First, different activities such as walking and sitting are extracted by the smartphone accelerometer data and used as inputs to an HMM algorithm. The hidden states are used as the locations of the occupant. Finally, CHMM is constructed to infer the persons' activities in a multi-occupancy environment. The hidden states of the CHMM and HMM refer to the activities, whereas the observations of the CHMM and HMM indicate both the location and posture of the individual. The results obtained with five people demonstrated that their proposed method improves the accuracy up to 70%, compared to 30% when only accelerometer data is used. Nevertheless, the cooperative activities, where many residents work together in a cooperative manner such that each person partakes in the same activity separately or together (e.g., two persons moving a table by holding it by the ends), were ignored in this research.

In [19] the authors present an overview of different classification techniques used to recognise human activity based on wearable sensors. They used four supervised classification techniques, namely K-Nearest Neighbor (K-NN), SVM, Gaussian Mixture Models (GMM), and RF as well as three unsupervised classification techniques namely K-Means, GMM, and HMM. These were compared in terms of correct classification rate, recall, precision, and specificity. The results obtained from their study indicate that the K-NN classifier gives the best performance compared to other supervised classification algorithms, whereas the HMM classifier is the model that provides the best result among the unsupervised classification algorithms.

Several other techniques are used for identifying activities in a multi-occupancy environment. A new research work reported in [123], introduced a hybrid mechanism between ontology-based and unsupervised machine learning for detecting and separating the activities of a single person in a multi-occupancy environment. The authors tested and evaluated their method based on a CASAS Spring dataset. The results obtained from this work show that the proposed method achieved an average activity recognition rate of 95.83% in the context of a multi-occupancy home environment. Another new research study [4], presented a daily activity recognition method based on time clustering for multi-occupancy in a smart home environment. The method required features that are extracted from raw data using a de-noising method. Then, cluster techniques are used to separate activities which occur at the same location but at various times. Finally, a similarity matching method is used to complete daily activity recognition. The authors tested the performance of the proposed method based on two multi-occupancy datasets provided by the CASAS repository. The results obtained from their research indicate that the proposed method for recognition of daily activities of multi-occupancy in a smart home environment achieved an accuracy of 92%.

Some other research studies have addressed the challenge of detecting daily activities in a multi-occupancy home environment using wearable sensors [124] or video sensors [22]. The major drawback of using these types of sensors is that they are not widely accepted by individuals, due to privacy and ethical concerns [14, 15, 16, 17]. Thus, it is often a preferred solution to utilise ambient sensors to identify and recognise multi-occupancy in a home environment [34].

2.4.4 Data Association in Multi Occupancy Environments

Many of the previous studies on multi-occupancy HAR have used ambient sensors. In this context, some previous studies have focused on the data association in multi-occupancy environments to recognise the residents [34, 125]. For example, in [126], CRF is applied to deal with the problem of data association in multi-resident activity recognition using the CASAS dataset [117]. The results of the study indicate that data association is a fundamental problem in dealing with a multi-occupant environment. It also mentions that modelling human interaction is a critical issue when modelling activities in a multi-resident environment. Likewise, in [127], the authors proposed two HMM models to recognise activities in the multi-resident environment. The first model of HMM is used to identify the resident. The second model is used to identify each of the separate activities. The results of these studies show that the performance of the proposed HMM models is low due to the sensors incapable of distinguishing who activated them in the absence of any tagging system to distinguish individuals in the environment. A study reported in [128] used Incremental Decision Trees (IDT) in an attempt to deal with ADL in a multi-occupancy environment. The proposed method was evaluated using the ARAS dataset, a real dataset collected from a smart home environment. However, the results showed that only about 40% rate of classification was achieved.

Most of the previous studies disregarded the possible interactions between occupants due to the data association problem when recognising multi-occupancy activities [129, 130, 131, 132]. The authors in [129, 130] used two different activity recognition models, HMM and CRF; whereas the study in [132] used five different classifiers namely, HMM, Decision Trees (DT), KNN, Time-Delay Neural Network (TDNN), and Multi-Layer Perceptron (MLP) to evaluate the activity recognition performance of a single resident in the multi-occupancy environment. They used these methods to recognise multi-occupancy activities by combining labels. The results of their research showed that the TDNN method gives the best performance in terms of accuracy and precision compared to the other methods.

Table 2.3 provides a summary of the related research studies for a multi-occupancy environment in the context of the publication year, the approaches used, name of the dataset, the type of sensors used as well as the results obtained.

Table 2.1: A summary of the related research studies for anomaly detection in daily activities in the context of the publication year, the approaches used, dataset name, the type of sensors used as well as the the system overall accuracy.

Reference	Year	Approach	Dataset name	Type of Sensors	Overall Accuracy
[66]	2020	OC-SVM	Self-gathered	Camera vision	61%
[46]	2020	SVM	Self-gathered	Ambient sensors	86.07%
[69]	2020	SiTGRU	UCSD Ped1 and UCSD Ped2 [133] & CUHK dataset [134]	Camera vision	-
[10]	2020	PUMAD	tabular dataset [135] & MNIST dataset [136]	Camera vision	93.29%
[8]	2019	CNN,LSTM	CASAS dataset [137] & Aruba dataset [138]	Ambient sensors	89.72%
[50]	2019	LSTM	Aruba dataset [63]	Ambient sensors	87.5%
[68]	2019	CNDE	Self-gathered & CASAS dataset [138]	Ambient sensors	95.7%
[40]	2017	HHM	Video surveillance dataset [139]	Visual-based sensors	89.15%
[41]	2016	RF	Simulated dataset [140]	Ambient sensor	-
[60]	2016	HMM	Self-gathered	Ambient sensor	-
[38]	2015	HMM	Self-gathered	Ambient sensors	95.10%
[36]	2015	DBSCAN	CASAS dataset [137]	Ambient sensors	-

Table 2.2: A summary of the related research studies for human fall detection in the context of the publication year, the approaches used, dataset name, the type of sensors used as well as the results obtained.

Reference	Year	Approach	Dataset name	Type of Sensors	Overall Accuracy
[141]	2020	CNN	DLR dataset [142] & tFall dataset [143]	Wearable sensors	98.78%
[141]	2020	CNN, LSTM	ASLH dataset [144]	Wearable sensors	96.64%
[107]	2020	SSHFD	Multiple-camera fall dataset [105] & Le2i fall dataset [108]	Camera vision	90%
[45]	2019	CNNs	SmartWatch and Notch datasets [145]	Wearable sensors	99.86%
[98]	2019	SVM	URFD dataset [87]	Visual-based sensors	97.5%
[99]	2019	CNN	Multiple Cameras Fall (MCF) dataset [100] & URFD dataset	Visual-based sensors	-
[47]	2018	SVM	URFD dataset	Visual-based sensors	98.15%
[89]	2018	RNN	URFD dataset	Wearable sensors & depth cameras	98.57%
[39]	2017	XHMMs	MobiFall dataset [146] & German Aerospace Center dataset [147]	Wearable sensors	96.6%
[91]	2006	HHMM	Self-gathered	Visual-based sensors	98%

Table 2.3: A summary of the related research studies for a multi-occupancy environment in the context of the publication year, the approaches used, dataset name, the type of sensors used as well as the results obtained.

Reference	Year	Approach	Dataset name	Type of Sensors	Overall Accuracy
[4]	2020	Time clustering	CASAS dataset	Ambient sensors	92%
[123]	2020	ontology-based and unsupervised machine	CASAS dataset	Ambient sensors	95.83%
[148]	2019	Multi Label Classification (MLC)	ARAS dataset	Ambient sensors	74.8%
[7]	2018	DCNN	CASAS dataset	Ambient sensors	95.2%
[13]	2017	FHMM	ARAS dataset	Ambient sensors	64%
[23]	2017	SVM	Self-gathered	Ambient and wearable sensors	58%-83%
[12]	2017	MMPP	Self-gathered	Ambient sensors	82.3%
[30]	2016	CHMM and HMM	Self-gathered	Ambient and wearable sensors	70%
[19]	2015	K-NN, SVM, GMM and RF	Self-gathered	Wearable sensors	-
[15]	2015	SARRIMA	Self-gathered	Ambient sensor	96%
[128]	2014	IDT	ARAS dataset	Ambient sensor	40%
[132]	2014	TDNN	Self-gathered	Ambient sensors	84.6%
[110]	2012	SVM	Self-gathered	Ambient sensors	83.5%
[149]	2010	Bayesian framework	Self-gathered	Ambient and wearable sensors	80.2%
[22]	2005	Linear Signal Model for Hybrid and Video Decoding	Self-gathered	Camera vision	90%
[20]	2004	HMM	Self-gathered	Camera vision	98.3%

2.5 Literature Review Summary

Based on the knowledge gained from the literature review in this chapter, it is found that several research studies paid attention to the detection of various types of anomalies in daily activities in an environment equipped with different sensors. Although statistical and computational intelligence techniques are commonly utilised to detect and identify anomalies in ADLs, there are some challenges associated with their employment. For example, HMMs have some challenges in terms of extracting multiple interacting activities (either cooperative activities or parallel activities). Moreover, without significant training, the possible observation sequences consistent with a particular activity might not be recognised utilising an HMM [44]. Therefore, more investigation is required to develop an appropriate method or algorithm that can efficiently detect such anomalies. The conclusions from the reviewed literature studies are outlined below:

- Most of the existing methods for anomaly detection in ADLs are constrained to low dimensional data and small data size because of the legacy of their original algorithms [150, 151]. These approaches often under perform resulting in too many false alarms (having normal instances identified as anomalies) or too few anomalies being detected and therefore generate a high false alarm rate [36]. For example, without significant training, the possible observation sequences consistent with a particular activity might not be recognised utilising an HMM or SVM [44]. A method with a high false alarm rate might not be suitable for reliably detecting anomalous events in the daily activities of users, especially for older adults. To reduce false alarms, the user's behaviour needs to be monitored and recognised accurately. This can be achieved by utilising a suitable technique, which enables the ADL data to be identified as either normal or anomalous.

- Based on the reviewed literature studies, most of the current research in detecting anomalies in ADLs focuses on a single-occupant environment when only one person is monitored, and their activities are categorised

[8, 10, 50]. The assumption that home environments are occupied by one person all the time is often not true. It is common for the resident to receive visits from family members or health care workers (represented as a multi-occupancy environment). Visiting is considered as one of the most significant activities for older adults living alone at home [23]. The resident's activity pattern is expected to be different when there is a visitor in the same environment, which could also be considered as an abnormal pattern in the resident's activities. The behaviour of a person could vary due to some personal factors such as visits and the influence of health conditions. Reliable anomaly detection in ADLs, or identifying visiting times (e.g., visits made by healthcare workers) is considered as one of the most important components of many home health care applications [5]. Thus, existing methods are not able to reliably detect anomalous events in the resident's activities in the presence of a visitor and to identify the time of visits. Therefore, detecting a time of visit in a single-occupancy home environment (represented as a multi-occupancy environment) requires more investigation to provide a better understanding of the nature of activities. It is important to develop a system with the ability to identify the exact time of a visit without the need for visitors to be asked to carry a tag or wearable device to identify them.

2.6 Research Gap

Recognising human activities based on the data coming from a range of simple to complex sensors is an interesting area of research. Different kinds of sensors, such as wearable sensors and cameras, are used for anomaly detection in daily activities in a single-occupancy environment. Some studies have used video cameras for anomaly detection during daily activities due to the number of features that could be extracted from such data [10]. However, the use of a camera is not accepted by many users, mainly because of privacy concerns. Alternatively, several studies have been carried out based on wearable sensors, such as an accelerometer, for anomaly detection in ADLs [152]. The disadvantage of using these kinds of sensors is that they tend to make residents uncomfortable, and older adults can sometimes forget to wear or use them. Furthermore, there is also the challenge of increases power consumption with the sensors when used for a long-term [153]. This begets limitations in effectively developing appropriate methods that can efficiently detect anomalies in daily activities. On the other hand, due to the privacy and cost issues, identifying and detecting anomalies in daily human activities based on ambient sensors, such as PIR sensors and door sensors, is a preferred option. These kinds of sensors can be easily installed in the home environment and allow people to live normally without feeling they are restrained by the technology.

Several research studies have investigated methods to detect normal and abnormal human behavioural activities using different computational methods [50, 55]. There are some limitations to these approaches, however, including the fact that they do not take into account changes in individual routine [52]. Human behaviour is dynamic, and behaviour changes through an individual's life, due to factors such as visits and health influences. Reliable anomaly detection in ADLs is considered as one of the most important components of many home health care applications [154]. However, the majority of the anomalies detection methods proposed in daily activities are too simplistic and therefore generate a high false alarm rate, and they are focused on a single-occupant environment where only one individual is monitored. An approach with a high rate of false alarms may not be appropriate to reliably detect anomalies in ADLs, resulting in dissatisfaction on the part of users and caregivers [68, 36]. In order to restrict the false alarm rate, human behaviours need to be recognised and monitored accurately. This can be achieved by using an appropriate technique, such as entropy measures, which enables analysis to distinguish between normal and anomalous cases in daily activities with a high degree of accuracy. Unlike entropy measurements, most classification techniques require a large amount of training and classification time. Entropy measures analysis has not been given much attention for anomaly detection in daily activities. Therefore, entropy can be suitable for real-time anomaly detection systems.

To address the gaps identified, this research investigates the effectiveness of different entropy measures to detect and identify various types of anomalies in daily activities in single-occupant and multi-occupant environments based on information obtained using low-cost, non-intrusive ambient sensors. Besides, in one case study, entropy measures are used to investigate their effectiveness in detecting anomalies in daily activities based on wearable sensors.

2.7 Summary

This chapter presented the state-of-the-art research related to anomaly detection in daily activities, human fall detection as well as visitor detection in a single-occupancy home environment (represented as a multi-occupancy environment). The review also presented various anomaly detection techniques that have been investigated. Some limitations on utilising these techniques were also discussed in this chapter. In assisted living, technologies such as smart homes are used to help and support older adults to live safely and independently in their own homes. Although there are still gaps in practical implementations of such systems, its significance cannot be overemphasised.

Based on the knowledge gained from the literature review in this chapter, entropy measures analysis has not been given much attention for anomaly detection in daily activities. Entropy analysis is an established method for irregularity detection in many applications; however, it has rarely been applied in the context of ADL and HAR. To classify ADL data representing the individual's daily activity routine as either normal or abnormal, entropy measures are considered as a useful measure to detect different anomalies in ADLs. This is investigated in this research. To reiterate the focus of this research, different entropy measures are employed to investigate their effectiveness in detecting various types of anomalies in daily activities in a single-occupant and multi-occupant environment.

In the following chapter, a description of different applied entropy measures for anomaly detection in daily activities is presented.

Chapter 3

Entropy Measures for Anomaly Detection

3.1 Introduction

Entropy measures are used to detect the irregularities and the degree of randomness in data. This thesis draws on this concept to detect various anomalies in the resident's daily activities in a single-occupancy and multi-occupancy environment. The idea is to develop a framework based on entropy measures for anomaly detection in ADL, such as irregular sleep, human falls and identifying visitors. In the previous chapter, a broad review of previous works on HAR with a focus on anomaly detection in ADLs, human fall detection and detection of visitors in a multi-occupancy environment were discussed. However, the analysis of entropy measures has not been given much attention in the literature. To the best of our knowledge, none of the studies in the literature has applied any entropy measures for anomaly detection in daily activities.

This chapter presents the proposed entropy measures framework developed in this thesis. Moreover, the Indoor Mobility (IM) method is briefly described in this chapter, which was compared with the proposed entropy measures. The IM method is defined as the frequency of the transition from room to room in a home environment. This chapter is organised as follows: Section 3.2 gives an insight into the concept of entropy measures and definition of entropy. Section 3.3 presents the explanation of certain entropy measures that are used for anomaly detection in this thesis. In this section, an in-depth explanation of how these entropy measures are carried out is provided, followed by a description of indoor mobility method in Section 3.4. Section 3.5 follows by presenting an explanation of entropy-based thresholding and the methodology proposed for anomaly detection in daily activities in this thesis is presented in Section 3.6. Finally, a summary of this chapter is presented in Section 3.7.

3.2 Background and Definition

Entropy was proposed in the nineteenth century by Rudolph Clausius [155] as a suitable complexity measure to determine the amount of disorder or uncertainty in a system or time-series data [156]. The concept of entropy is utilised in many fields of science, including statistical mechanics, information theory, neural networks, taxonomy, and mathematical linguistics [37]. Considering different methods, entropy can be utilised as a measure of randomness or uncertainty in a system. Entropy increases as the system's randomness increases. For example, if the degree of randomness is low, the system will become organised. A system is considered as completely organised when the entropy value is zero. Whereas, a high disorder in the data will give higher entropy values, as shown in Figure 3.1.

In [157], Shannon proposed entropy for information theory to describe the distribution of signal components.

Given a discrete random sequence A, Shannon Entropy (SE) can be defined as:

$$SE = -k \sum_{i} p(i) \ln p(i) \tag{3.1}$$

where $p(i)$ is the probability that it occurs during the system's fluctuations and k is Boltzmann constant.

Thus far, numerous entropy algorithms have been proposed and are extensively used to quantify the irregularity of signals, and image-processing applications [158]. The computations, however, are frequently confronted with

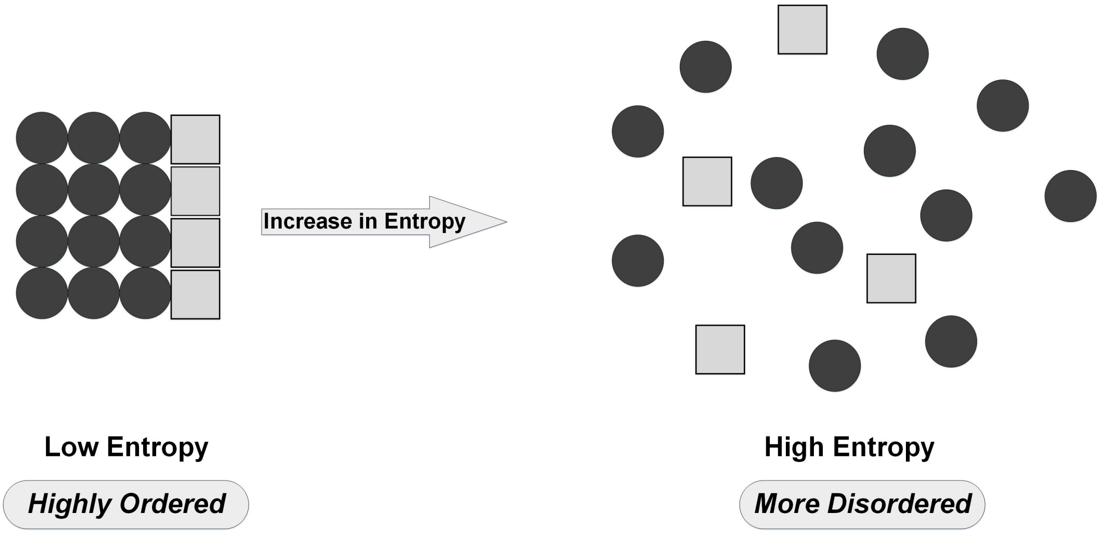


Figure 3.1: An illustration of entropy measurement definition.

the challenge of an insufficient number of data points. Moreover, certain recorded data are, to a certain degree, contaminated by noise. To deal with this problem of rather short and noisy recordings in physiological signals, Approximate Entropy (ApEn) was proposed in [159] to avert challenges in the finite length of a time series and in need to discriminate the nature of the generating systems. High regularity and low randomness in the data produce smaller entropy values, whereas, less regularity gives higher entropy values. However, the disadvantage of ApEn is that it lacks relative consistency, and it is strongly dependent on the length of a time series [37]. Authors in [160] introduced Sample Entropy (SampEn) to overcome the drawbacks of ApEn by excluding self-matches; thus, decreasing the calculation time by one-half in comparison with ApEn. The SampEn is less dependent on the data length and shows relative consistency; however, matching vectors in both ApEn and SampEn are either 1 or 0 values. Therefore, this is not realistic when dealing with real-world examples where boundaries are not fixed [161]. To overcome such cases, Fuzzy Entropy (FuzzyEn) was proposed in [162], as a method to compute the regularity in a time series. In FuzzyEn, the concept of an exponential function, $\exp(-(d_{ij}^m)^n/r)$, is applied as a fuzzy function that evaluates the similarity degree of two points (vectors).

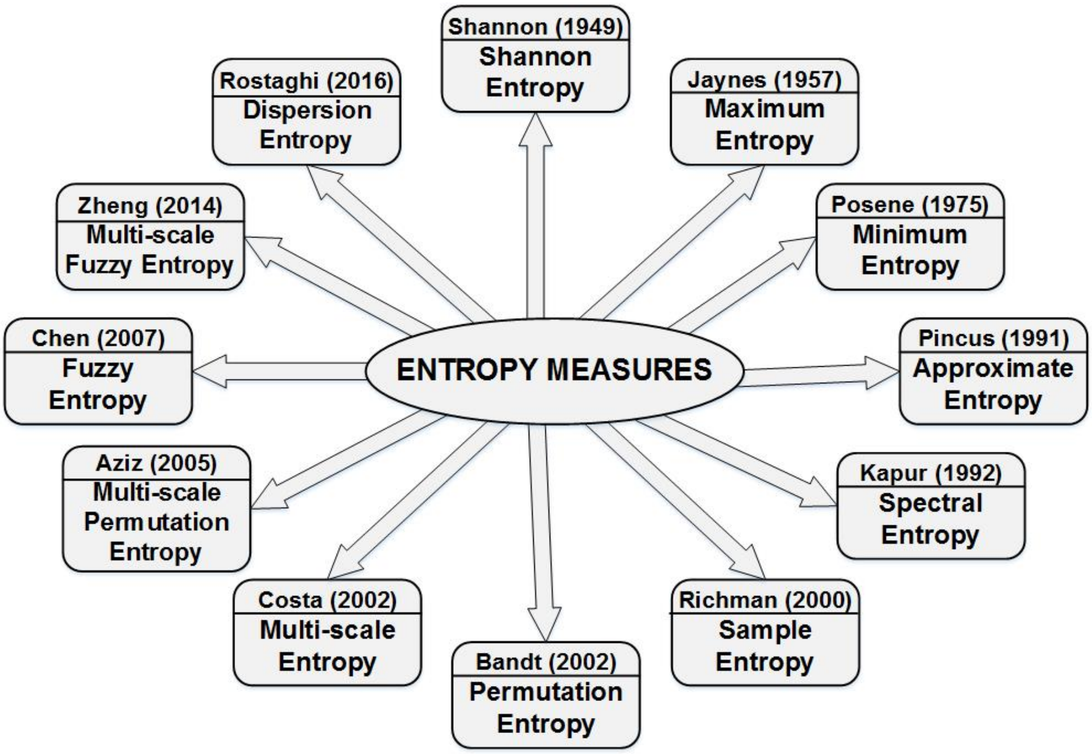


Figure 3.2: Different types of entropy measures, presented in chronological order.

A further commonly utilised regularity indicator is Permutation Entropy (PerEn), proposed in [163]. It is based on the arrangement relations between signal values and on the measure of the relative frequencies of ordinal patterns. The PerEn is considered a simple measure that generates fast calculations. However, the measure does not consider the variation among amplitude values and the average value of amplitudes [158]. Existing entropy measures, such as ApEn, SampEn, PerEn, and FuzzyEn are widely utilised to measure the irregularity of signals at single-scale. Nevertheless, these measures fail to compute the multiple time scales engrained in biomedical recordings [164]. To overcome this limitation, Multi-scale Entropy (MSE) was proposed in [165] and it is employed to quantify the irregularity of univariate time series, notably physiological time series.

The possibility of using entropy to determine the degree of disorder or uncertainty in a system resulted in the definition of different types of entropy. Figure 3.2 shows various entropy measures presented in chronological order.

More detailed information about the applied entropy measures are provided in the next section.

3.3 Applied Entropy Measures

To evaluate the relevance of entropy measures in ADLs, different entropy measures are investigated. Some of the entropy measures are proven to be more relevant than others. As part of the investigation, many different measures are investigated. A description of the entropy measures utilised in the rest of this thesis is presented below.

3.3.1 Shannon Entropy

Shannon Entropy (ShEn) was initially proposed by Shannon [157]. ShEn is a method to measure the degree of uncertainty in data associated with the occurrence of the result. In particular, ShEn quantifies the predictive value of the information contained in a message. Since then, it has been widely utilised in the information sciences [166].

For a given time series $A = \{a(i) : i = 1, 2, ..., N\}$, the ShEn is defined as:

$$ShEn = \sum_{i=1}^{N} p(a_i) \log_2 \frac{1}{p(a_i)} = -\sum_{i=1}^{N} p(a_i) \log_2 p(a_i) \tag{3.2}$$

where $p(a_i)$ is the probability of acceptance by the random variable A that takes the values a_i. The entropy of variable A is a measure of the expected randomness obtained through the measurement of the values in variable A. A higher entropy value is obtained by more uncertainty in the data and is more difficult to predict [37].

3.3.2 Approximate Entropy

Approximate Entropy (ApEn) was initially introduced by Pincus [159] to classify the concept of complex systems. It is a technique used to quantify the concept of regularity and uncertainty within a sequence of data in a system

[161]. High regularity and low randomness in the data produce smaller entropy values, whereas, less regularity gives higher entropy values. To compute the ApEn, the parameters of the embedding dimension m and tolerance r are required as input parameters. The following is the explanation of the procedure for the ApEn-based algorithm as described in [159].

For time series A with N samples, the sequences of vector A_i^m can be written as:

$$A_i^m = \left[a(i), a(i+1), ..., a(i+m-1)\right],\ for\, i = 1, ..., (N-m+1) \tag{3.3}$$

where m is the embedding dimension. The distance between two vectors A_i^m and A_j^m is represented as the maximum absolute variation between their scalar components:

$$d\left[A_i^m, A_j^m\right] = \max_{k=0,1,...,m-1} \left(|a(i+k) - a(j+k)|\right) \tag{3.4}$$

For each A_i^m, the number of $j \leq N-m+1$ such that $d[A_i^m, A_j^m] \leq r$, where r is the tolerance, is given as $N_i^m(r)$. The parameters $C_i^m(r)$ are then defined as:

$$C_i^m(r) = \frac{1}{N-m+1} N_i^m(r) \tag{3.5}$$

where $C_i^m(r)$ represent the number of $j \leq N-m+1$ such that $d[A_i^m, A_j^m] \leq r$. The $\phi^m(r)$ represent the mean value of parameters $C_i^m(r)$, which can be defined as:

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \tag{3.6}$$

Using $\phi^m(r)$ and $\phi^{m+1}(r)$, the ApEn (m, r) is defined as:

$$ApEn(m, r) = \lim_{N\to\infty} \left[\phi^m(r) - \phi^{m+1}(r)\right] \tag{3.7}$$

Finally, the ApEn is calculated for finite time series length N as:

$$ApEn(m, r, N) = \phi^m(r) - \phi^{m+1}(r) \tag{3.8}$$

The following are the properties of ApEn [167]:

- The ApEn algorithm requires datasets that are equally spaced over time, which is dependent on the computational time of ApEn.
- Non-linearity leads to a higher ApEn value.
- To compute the ApEn, the parameters of the m and r are required to be defined.
- Recommended values: m have to be low, $m = 2$ or 3 are typical options, and r must be in range 0.1 to 1.
- The number of data (N) required to distinguish between systems is in the range of 10^m to 30^m.
- The ApEn algorithm uses the data vector A_i^m instead of utilising the probabilities connected with the occurrence of each result.

3.3.3 Sample Entropy

ApEn bias has two essential challenges. The first one is that the relative consistency is not secured, and the results could be different depending on the value of tolerance r. The second one is that the ApEn value is strongly dependent on the length of the data series [167]. To avert these two challenges, Sample entropy (SampEn) was introduced by Richman and Moorman [160]. It is a method used to measure regularity and complexity in time series data, which is mostly used for nonlinear analysis and does not have self-counting. To compute the SampEn, the parameters of m and r are required to be defined [168]. The SampEn is the negative natural logarithm of the conditional probability that two similar vectors of m will be matched for $[m + 1]$ samples without allowing self-matches. The following is a description of the procedure for SampEn-based algorithm, as provided in [160].

For vector sequences, the distance between the two vectors, $d\left[A_i^m, A_j^m\right]$ are calculated as in $ApEn$. For a given A_i^m, we calculate $b_i^m(r)$ as $(N-m-1)^{-1}$ multiplied by the number of A_j^m within r of A_i^m, where j ranges from 1 to $N-m$ and $(j \neq i)$. $b^m(r)$ is then calculated as:

$$b^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} b_i^m(r) \tag{3.9}$$

Similarly, by increasing the embedding dimension m to $m+1$, the $a_i^m(r)$ is defined as$(N-m-1)^{-1}$ multiplied by the number of vectors A_j^{m+1} within r of A_i^{m+1}, whereas j ranges from 1 to $N-m$ and $(j \neq i)$. Furthermore, $a^m(r)$ is defined as:

$$a^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} a_i^m(r) \tag{3.10}$$

Therefore, the probability that two vectors will be matched for m samples is given by $b^m(r)$, while $a^m(r)$ represents the probability that two vectors will be matched for $[m+1]$ samples. Then, sample entropy can be calculated as:

$$SampEn(m,r) = \lim_{N\to\infty} \left(-\ln\left[\frac{a^m(r)}{b^m(r)]}\right]\right) \tag{3.11}$$

SampEn is defined for a finite time series length N as:

$$SampEn(m,r,N) = \ln\left[\frac{a^m(r)}{b^m(r)]}\right] \tag{3.12}$$

3.3.4 Permutation Entropy

Permutation Entropy (PerEn) was introduced by Bandt and Pompe [163]. It is based on the measure of the relative frequencies of ordinal patterns and combines the concept of Shannon Entropy with ordinal pattern analysis, through the estimation of the related frequencies of the ordinal patterns obtained from time-series [169]. There are two parameters, embedding dimension m and time delay τ, which must be defined to calculate the PerEn. Thus, the algorithm for PerEn measure is impacted by the selection of these values. When m and τ are too small, the algorithm may not work, as there are too few distinct states. On the other hand, too large of an m and τ is also unsuitable for detecting the dynamical changes in data. Therefore, it is motioned that the PerEn with $m = 3$ and $\tau = 1$ may be the most appropriate choice [163, 170]. The following is a description of the procedure for the PerEn-based algorithm as provided in [163].

For vector sequences, the m-dimensional delay embedding vector at time i is defined as:

$$A_i^m = \big[a(i), a(i+\tau), ..., a(i+(m-2))\tau, a(i+(m-1))\tau\big] \tag{3.13}$$

where m is the embedding dimension and τ is time delay. The vector A_i^m has a permutation $\pi = (r_0 r_1 ... r_{m-1})$ if it satisfies:

$$a(t+r_0\tau) \leq a(t+r_1\tau) \leq ... \leq a(t+r_{m-1}\tau) \tag{3.14}$$

where $0 \leq r_i \leq m-1$ and $r_i \neq r_j$.

There are $m!$ permutations π of order m, which are considered as possible order kinds of m different numbers. For each permutation π, the relative frequency is determined by:

$$p(\pi) = \frac{Number\{t|t \leq N-(m-1)\tau, A_i^m \text{ has type } \pi\}}{N-m+1} \tag{3.15}$$

The permutation entropy (PerEn) of the m dimension is then defined as:

$$PerEn(m,\tau) = -\sum_{i=1}^{N} p(\pi)\log p(\pi) \tag{3.16}$$

The maximum value of $PerEn(m)$ is $\log(m!)$ where all possible permutations appear with the same probability. Therefore, the Normalised Permutation Entropy (NPE) is defined as:

$$NPE = \frac{PerEn(m)}{\ln(m!)} \tag{3.17}$$

3.3.5 Multi-scale Permutation Entropy

The drawback of PerEn is the requirement of a large dataset for it to be reliable. To overcome this problem, the Multi-scale Permutation Entropy (MPE) was proposed by Aziz and Arif [171], which has been utilised as an efficient method to measure complexity over a range of scales. The MPE is an extension of the PerEn by utilising the multiscale entropy proposed in [165]. Multi-scaling is especially helpful in quantifying the information content in long-range trends. In MPE analysis, the entropy of the multiple coarse-grained time series at each scale is computed by the PerEn [172, 173]. The following procedure explains the MPE calculation, as described in [171].

For vector sequences, multiple coarse-grained time series are converted by taking the average of the data points inside non-overlapping windows of length s. The coarse-grained time series $y_j^{(s)}$ is defined by utilising the following equation:

$$y_j^{(s)} = \frac{1}{s} \sum_{i=(j-1)s+1}^{js} a(i), \quad 1 \leq j \leq \frac{N}{s} \tag{3.18}$$

where s represents the scale factor. The length of each coarse-grained time series is equal to the length of the original time series divided by the scale factor s.

The Permutation Entropy as described in the previous sub-section is calculated for each coarse grained time series. The PerEn values for each scale factor are then plotted as a function of the scale factor. Therefore, MPE can be defined by:

$$MPE = PerEn(m, \tau, y_j^{(s)}) \tag{3.19}$$

3.3.6 Fuzzy Entropy

ApEn and SampEn produce matching vectors with either 1 or 0 values. This is unrealistic when dealing with real-world examples where the partition between classes may be cryptic or uncertain. Therefore, in the case of SampEn and ApEn, the input patterns cannot be determined [161]. To overcome such łcases, fuzzy sets and membership degrees are introduced. The membership degree is

introduced by a fuzzy membership function $\mu c(x)$ which allows each point x to be associated with a real value within a range $[0, 1]$. The theory introduces a mechanism to measure the degree to which a pattern belongs to a given category, so the membership degree of x in dataset C will become higher when the value of $\mu c(x)$ is nearer to unity. Fuzzy entropy (FuzzyEn) was proposed by Chen et al. [162], which is defined as a method to compute regularity in time series. In FuzzyEn, the concept of exponential function $\exp(-(d_{ij}^m)^n/r)$ is used as a fuzzy function that evaluates the similarity degree of two points (vectors). The following is a description of the procedure for the FuzzyEn-based algorithm, as presented in [162]. FuzzyEn accepts self-matches and beholds only the first $(N - m)$ vectors of length m to confirm that A_i^m and A_i^{m+1} are determined for all $(1 \leq i \leq N - m)$.
Where $a_0(i)$ is the average value of A_i^m over the set of m values defined as:

$$a_0(i) = \frac{\sum_{j=0}^{m-1} a(i+j)}{m} \tag{3.20}$$

The distance between vectors A_i^m and A_j^m is given by d_{ij}^m and calculated as:

$$d_{ij}^m = Max_{k=0,\dots,m-1}\Big|\Big(a(i+k) - a_0(i)\Big) - \Big(a(j+k) - a_0(j)\Big)\Big| \tag{3.21}$$

Based on the fuzzy membership function $\mu\Big(d_{ij}^m, r\Big)$, the similarity degree D_{ij}^m between the vector A_i^m and the next vector A_j^m is defined as:

$$D_{ij}^m = \mu\Big(d_{ij}^m, r\Big) \tag{3.22}$$

where the fuzzy membership function $\mu\Big(d_{ij}^m, r\Big)$ is an exponential function defined as:

$$\mu\Big(d_{ij}^m, r\Big) = exp\big(-(d_{ij}^m)^n/r\big) \tag{3.23}$$

where n and r are the gradient and width of the exponential function, respectively.

For each vector $A_i^m; i = 1, ..., N - m + 1$, averaging all the similarity degree of its next vectors $A_j^m; j = 1, ..., N - m + 1$, and $j \neq i$ is defined as:

$$\phi_i^m(r) = \frac{\sum_{j=1, j\neq i}^{N-m} D_{ij}^m}{N - m - 1} \tag{3.24}$$

Then, the $\phi^m(r)$ is defined as:

$$\phi^m(r) = \frac{\sum_{i=1}^{N-m} \phi_i^m(r)}{N - m} \tag{3.25}$$

and for A_i^{m+1}, averaging all the similarity degree of its next vectors is defined as:

$$\phi^{m+1}(r) = \frac{\sum_{i=1}^{N-m} \phi_i^{m+1}(r)}{N - m} \tag{3.26}$$

The FuzzyEn(m,r) is then calculated as:

$$FuzzyEn(m, r) = \lim_{N\to\infty} \Big[\ln \phi^m(r) - \ln \phi^{m+1}(r) \Big] \tag{3.27}$$

Finally, the Fuzzy Entropy can be defined for the finite time series of length N as:

$$FuzzyEn(m, r, N) = \ln \phi^m(r) - \ln \phi^{m+1}(r) \tag{3.28}$$

3.3.7 Multi-scale Fuzzy Entropy

The Multi-scale Fuzzy Entropy (MFE) was proposed by Zheng et al. [174]. Based on the definition of FuzzyEn, the following procedure explains the MFE calculation, as described in [174].

For vector sequences, the coarse grained time series $y_j^{(s)}$ is calculated. The FuzzyEn measure, as described in the previous section, is then calculated for each coarse grained time series. The FuzzyEn values for each scale factor are then plotted as a function of the scale factor. Therefore, MFE can be defined by:

$$MFE(A, s, m, r) = FuzzyEn(m, r, y_j^{(s)}) \tag{3.29}$$

3.4 Indoor Mobility Method

The Indoor Mobility (IM) is defined as the frequency of the transition from room to room as described in [175]. This is a measure representing the degree of mobility. Given a smart home installed with several PIR sensors, a resident's transitions in the home can be detected. Binary sensors are considered only, where S at location L with its value at time t, can be defined as:

$$S_t^L = \begin{cases} 0 & OFF \\ 1 & ON \end{cases} \tag{3.30}$$

The sequence of any sensor data for all times can be defined as: $S^L = \{S_0^L, S_1^L, ..., S_t^L\}$ and the transition from room to room can be written as:

$$Tr = \{(S^{L_1}, S^{L_2})_{st_1}, (S^{L_2}, S^{L_3})_{st_2}, ..., (S^{L_i}, S^{L_j})_{st_i}\} \tag{3.31}$$

where $i \neq j$ and st_i is the time when the resident enters the location i. IM is defined as the total number of transitions from a room to another between time $T1$ and $T2$, and can be written as:

$$IM = \left|(S^{l_1}, S^{L_2})st_1, ..., (S^{L_i}, S^{L_j})st_i\right|_{T1}^{T2} \tag{3.32}$$

In a smart home consisting of many areas, the activity pattern of the resident can be defined as the number of movements from a place to a different place (transition) as well as the time spent by the resident in each place (duration). For example, Figure 3.3 shows sensor data collected from 5 PIR sensors in different locations (e.g., Kitchen, Bedroom, Bathroom, Corridor, and Living room) over a one-day period and the duration in hours, spent by the resident in each room. Where the y-axis represents the sensor status (on/off) as a binary value in different locations, and x-axis represents time in hours. The Figure also

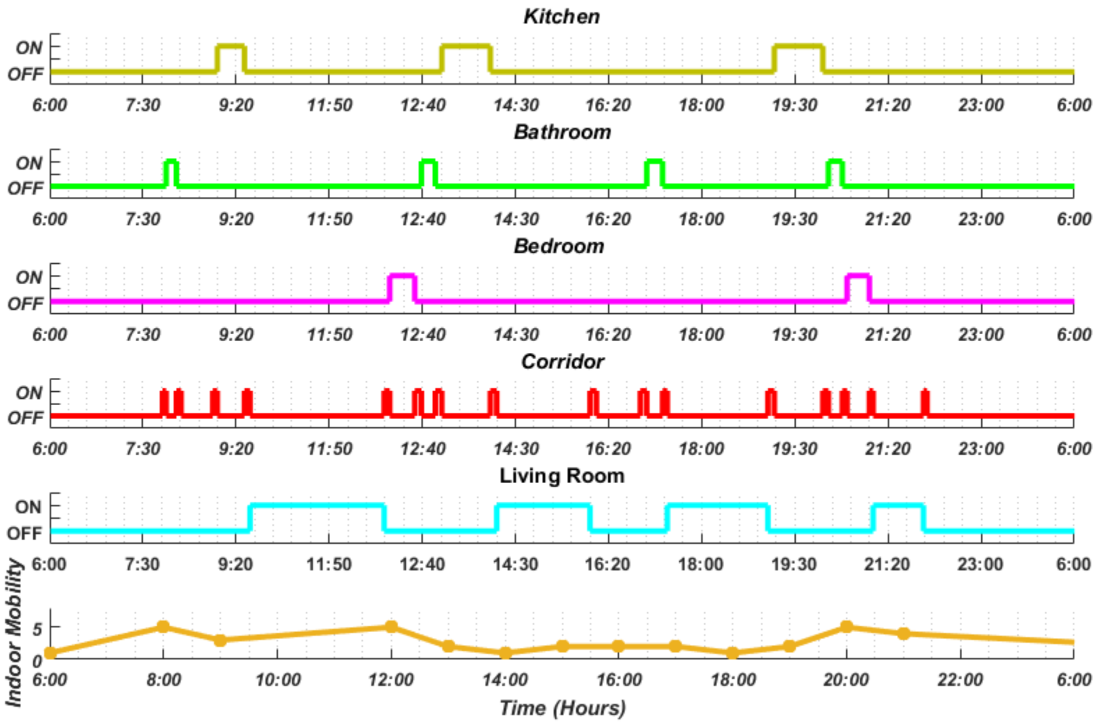


Figure 3.3: An example of raw sensor data gathered from PIR sensors in different locations, where the y-axis represents the sensor status (on/off) as a binary value in different locations; and x-axis represents time in hours, and the computed indoor mobility over a 24-hour period.

illustrates the computed indoor mobility over a 24-hour period by computing the transition from room to room.

3.5 Entropy-based Thresholding

The thresholding technique is relevant in different applications, including image processing and anomaly detection. There are several thresholding techniques that can be used for distinguishing between normal and abnormal events, such as a threshold based on the standard deviation and maximum threshold [176, 177]. The threshold based on the standard deviation is calculated by finding the standard deviation for a given data. The standard deviation is a method used to measure the amount of variation or dispersion of a set of values. It is commonly

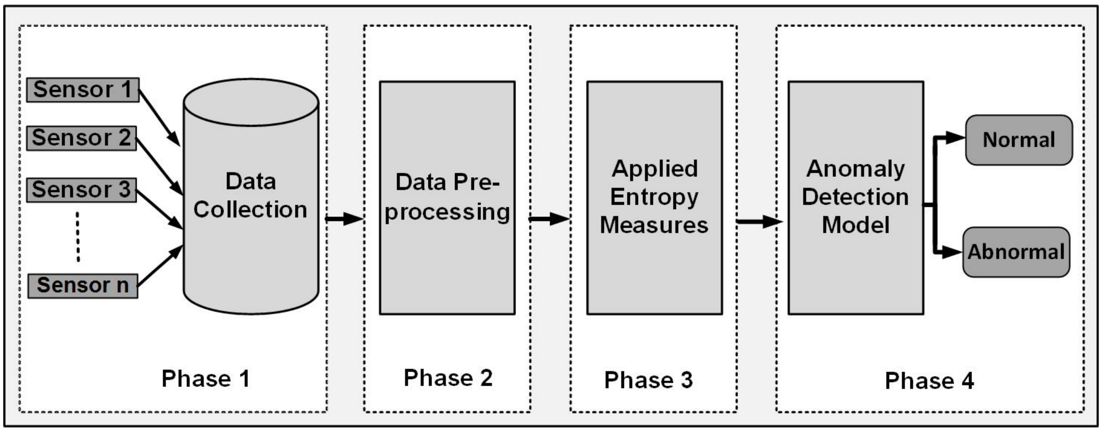


Figure 3.4: An schematic diagram of the proposed methodology framework for anomaly detection in ADLs using entropy measures.

utilised in statistical conclusions. Hence, the standard deviation is defined as:

$$SD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{3.33}$$

where x_i is the i^{th} value in the dataset, $\overline{x}$ is the average of the x-values in the dataset, and N is the number of frames.

The maximum threshold technique is aimed at finding the maximum values of normal data to be used as a proper threshold in order to detect any abnormality in such data. This means that by finding the maximum threshold value on normal data, it is possible to detect and identify abnormal events in completely unseen data.

3.6 Proposed Methodology

This research presents a novel anomaly detection based on entropy measures from data obtained using low-cost non-intrusive ambient sensors, which include PIR sensors and a door entry sensor. Furthermore, this research also investigates whether entropy measures can be used for anomaly detection based on the data obtained using wearable sensors. Figure 3.4 shows an overall schematic diagram representing the proposed stages for anomaly detection in ADLs. There are four phases; 1) Data Gathering, 2) Data Pre-processing, 3) Applied Entropy Measures, and 4) Anomaly Detection Model.

The first phase is to gather the dataset representing ADLs in a home environment based on ambient sensors. We are primarily concentrating on the movement data representing the occupancy of different areas in a home environment. Without loss of generality, data gathered from other sensors, including door entry sensors and wearable sensors, could also be used. The process of the data-gathering phase is explained in the next Chapter 4. The second phase is to extract relevant features from the raw data to be used for calculating the input vector sequences of the entropy measures that can distinguish between normal and abnormal cases in daily activities. More details about the process of data pre-processing are provided in Chapter 4. In the third phase, different entropy measures are applied to the extracted vector sequence from the raw data to detect different anomalies in the extracted activity patterns. Then, in phase four, the threshold based on the standard deviation of the occupancy data in conjunction with several entropy measures is used to distinguish whether there are anomalies in the resident's activity or not. Novel anomaly detection based on entropy measures is proposed in Chapter 5 to detect anomalies in ADLs, specifically in sleeping routine and detecting human fall from other activities. Furthermore, in Chapter 6, different entropy measures are employed to investigate their effectiveness in identifying visitors (the time of their visits) based on non-intrusive sensors.

To evaluate the proposed concept for identifying anomalies in activities of daily living, five different datasets gathered from real environments and one dataset collected from HOME I/O 3D simulation environment are investigated. These datasets comprise information regarding ADLs, including preparing a meal (kitchen activity), staying in the living room, eating (dining room activity), sleeping, toileting, and going out of the home. Besides, each activity includes information within the data, such as the date, start time, end time, and the location of activities.

3.7 Summary

This chapter presented an insight into the concept of entropy measures and definition of entropy. The explanation of certain entropy measures and indoor mobility that are used for anomaly detection in ADLs is also provided. Also, an explanation of entropy-based thresholding techniques is presented.

Furthermore, to achieve the aim of anomaly detection in daily activities, the chapter presented the methodology adopted in this research which is based on entropy measures. This comprises gathering data from intelligent environments, extracting relevant features that can distinguish between normal and anomalous cases in daily activities, entropy measures, and detect any anomalies in the resident's activity.

In the following chapter, a description of the intelligent environments used and the data collection process are presented. Furthermore, the data pre-processing, data handling and feature extraction are explained.

Chapter 4

Data Collection and Feature Extraction

4.1 Introduction

To support the independent living of older adults in their own home, the first step is to identify when and where a specific activity has occurred in their home environment. Once the activity is recognised, then it is possible to provide appropriate support accordingly. Distinguishing and detecting anomalies in the daily activities of older adults is very important for healthcare management, as this helps carers to act early to avert prospective problems [4, 5]. For instance, if the toilet is used many times at night compared to the daily routine, or if night-time sleeping is recognised to be short compared to the usual pattern of sleep, then such activity could count as an anomaly in the resident's activity. Therefore, it is essential to investigate an appropriate approach or algorithm that can efficiently detect such anomalies based on the daily activities of the individuals who are living in home environments equipped with appropriate ambient sensory devices.

The sensor data collected from all sensors are stored in a database for further processing. Normally, such datasets include a large amount of complex sensory data representing ADLs of a person. Thus, the aim is to understand and extract the daily behaviour features of a resident from low-level sensory data. As a first step, it is essential to represent and visualise the data obtained in an appropriate format before any data processing is carried out.

In this chapter, an overview of intelligent environments, including sensor networks, is presented. The description of the procedure for data gathering from a sensor network to monitor and identify a resident's daily activities is also presented. The discussion presented in this chapter has mainly focused on the usage of data obtained using ambient sensors such as PIR or door entry sensors for anomaly detection in ADLs. The use of wearable sensors for gathering data representing ADLs is also to examined. The explanation of data pre-processing, features extraction, and data representation processes are also provided in this chapter.

This chapter is structured as follows: Section 4.2 gives an overview of ambient intelligent environment. Sensor data collection is provided in Section 4.3 where two different environments are explained in more detail. Section 4.4 outlines the data pre-processing and data handling processes for an intelligent environment, followed by feature extraction in Section 4.5. Section 4.6 explains the entropy calculation. Lastly, Section 4.7 draws conclusions to summarise the chapter.

4.2 Ambient Intelligence Environment

An Ambient Intelligent (AmI) environment is an environment equipped with appropriate sensor networks that can be utilised to monitor and identify the daily activities of the residents [178]. Information gathered from Aml environments can be used to detect and understand the occupant's activity patterns, allowing personalised care. The occupant's activity patterns can also be used to detect changes in behaviour and predict future events so that preventive action can be taken [175].

Figure 4.1 shows an overview of an AmI environment architecture. The home is first equipped with different sensors. The data is gathered by these sensors and transmitted through the communication link (either wired or wireless format) to a central hub and eventually stored in a central database. Then, a preprocessing performed for cleaning the data. This is required to get the data into an appropriate format the system can interpret. Finally, the

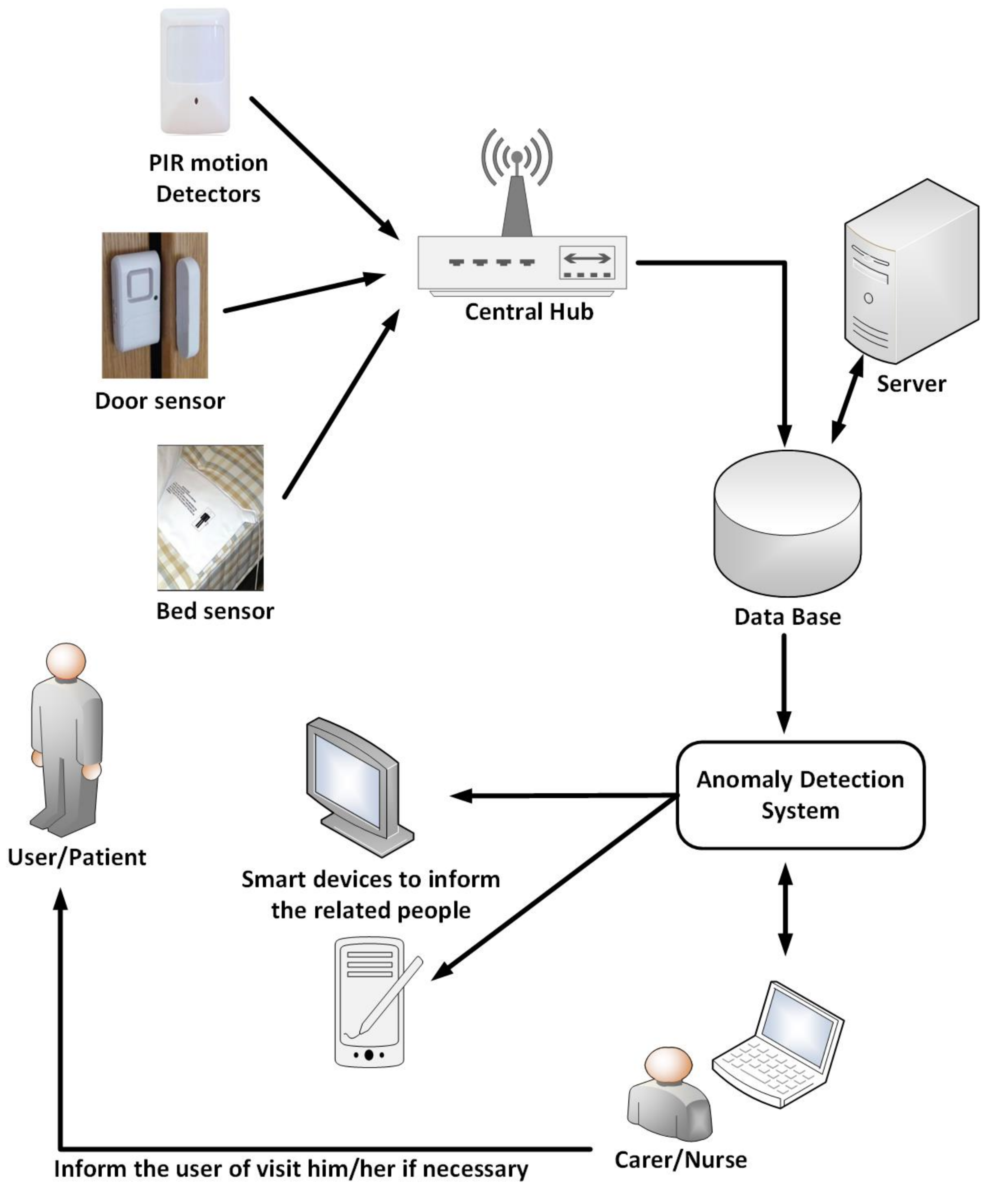


Figure 4.1: An overview of an AmI environment architecture.

results of the analysis will be shown to the responsible person in order to alert or warn the stakeholders (i.e. family members or health care workers).

The AmI technology aims to support people to have a better quality of life and ensure older adults live safely and independently in their own homes as comfortably as possible [179]. This is often the preferred solution for many older adults who want to live safely and independently in their own home. Therefore, intelligent technology is considered as one way to decrease the cost of living and care for older adults and to improve their quality of life. It has been used for many purposes, such as security and safety, speech recognition, energy saving, and activity recognition by being equipped with sensors to gather different types of data about the home and the resident [2, 3]. An automated monitoring system which could also identify abnormalities within ADLs would require an accurate recognition of human activities. Once daily human activities are recognised, the information obtained from intelligent environments or smart homes can be used to identify abnormalities in comparison with the routine activities [180]. Therefore, assistive technologies, such as automated anomaly detectors are used to help carers to act to avert prospective problems early and to improve older adults' quality of life.

4.3 Data Collection

The data collection is considered one of the most essential steps in human activity recognition [4]. The main goal of the data collection process is to collect information representing daily human activities within an intelligent environment. Different types of sensors are used for collecting the information representing various locations of a resident in a home environment. The following list outlines the detail of typical sensors:

- Passive Infra-red Sensors (PIR), also known as motion detectors, are sensitive to the movements of living objects. The PIR sensors are commonly used to track the movement of an occupant representing the occupancy of a specific area at home. They measure infrared light radiating from objects in its field of view. Hence, they can sense motion, and they are used to detect whether a human (or pet animal) has moved in or out of the sensors range. It is essential to place the PIR sensors in the right location to capture and monitor the occupant's movements in different areas.

- Door entry sensors are on/off switches which are used to detect the open and close status of a door. Door entry sensors are relatively credible as they effectually detect movement activities.

- Bed/sofa pressure sensors are utilised to detect the presence of a person in these areas.
- Electricity power usage sensors are utilised to monitor the activity of electrical devices by measuring their electrical current consumption.

There are three main steps for gathering dataset:

1. **Data collection -** The first step has to do with gathering sensor data in the smart home environment. The sensor data is captured using a dedicated sensor network and is stored in a database.
2. **Data annotation -** The second step is to annotate the activity labels into the database to recognise activities in future. This step is essential for the performance of the learning algorithms because, the annotated data is utilised in training the learning algorithms.
3. **Feature extraction -** The third step is to extract as many features as possible from the raw data. The selected features are then used as input to the proposed methods.

In most applications, only the occupancy sensors, including PIR sensor and door entry sensors, are used to track and monitor the resident in different locations in the home environment.

The data analysis provided in this thesis is based on two environments, a real and simulated home environments. For real home environments, five different datasets representing human activities are presented based on information obtained using ambient sensors. However, one of these datasets is gathered using one accelerometer sensor in order to examine whether the proposed method can be used for anomaly detection, solely based on information gathered from wearable sensors.

Details of both real and simulated home environments are provided in the following sections.

4.3.1 Real Environments

Five separate datasets gathered from a real home environment representing the ADLs are used to validate the results presented in this thesis. In these datasets, each resident is living alone in different real environments where their movement activities are different from one to another. However, some of these datasets represent multi-occupancy scenarios - situations when a visitor comes to visit the main occupant. Further details about these datasets are provided below.

4.3.1.1 Dataset A

The ADL dataset was gathered for the purpose of this research from a real home environment representing the ADLs of a single resident for a period of 72 days [68]. The dataset was gathered at the Smart NTU home facility within Nottingham Trent University. To collect the data, low-cost non-intrusive ambient sensors such as Mat pressure, PIR, and Door sensors were utilised. These kinds of sensors can be easily installed in the home environment and allow people to live normally without feeling restrained by the technology [17] used. The dataset comprises information regarding ADLs such as, preparing a meal (kitchen activity), staying in the living room, eating (dining room activity), irregular sleeping, toileting, and going out of the home. Besides, each activity is annotated including date, start time, end time, and the location of activities, as shown in Table 4.1.

4.3.1.2 Dataset B

The CASAS HH111 dataset from the CASAS repository[1] [138] is also utilised to evaluate the proposed method in this thesis. The dataset comprises information regarding the ADLs performed by a volunteer adult living alone in his home for a period of 50 days. Low-cost non-intrusive ambient sensors such as motion sensors, light sensors, temperature sensors, and door sensors were used as data collection devices. Activities recorded include eating, irregular sleeping, bathing, toileting, leaving home, etc. The dataset does not provide any information regarding whether or not the occupant's activity is abnormal.

[1] CASAS: A smart home in a box. http://casas.wsu.edu/datasets/

Table 4.1: A Sample of the gathered ADL dataset (Dataset A).

Date and Time	Sensor Status	Location
2018-05-01 17:07:07	0	Kitchen
2018-05-01 17:19:55	1	Dining room
2018-05-01 17:21:47	1	Kitchen
2018-05-01 17:29:52	0	Living room
2018-05-01 17:33:47	1	Corridor
2018-05-01 20:34:30	1	Toilet
2018-05-01 17:40:47	1	Corridor
2018-05-01 20:41:30	1	Bedroom-sleeping
2018-05-01 20:42:15	0	Living room
2018-05-01 23:01:45	1	Bedroom-sleeping
...	...	...

4.3.1.3 Dataset C

In order to evaluate the effectiveness of entropy measures for human fall detection based on dataset gathered from wearable sensors, experiments have been conducted using University of Rzeszow Fall Detection (URFD) dataset[1] [87]. It is a dataset publicly shared through the Interdisciplinary Centre for Computational Modelling, at the University of Rzeszow. This dataset was obtained using one accelerometer sensor placed near the pelvis area of the human body, and two Kinect cameras. This dataset is fully annotated. In total, the dataset contains 30 fall sequences and 40 activities of daily living sequences, such as lying on the floor, bending down, sitting down on a chair, picking an object up from the floor, and lying on the sofa/bed. In addition to this, the falls sequences contain two types of falls performed by five people, which are falling from sitting on a chair and falling from a standing position. Figure 4.2 shows examples of acceleration change curves during daily activities such as lying down on the floor, picking up an object and fall events, from the URFD dataset. In this research, only the accelerometer data is used, corresponding to 30 sequences containing human falls and 40 activities of daily living sequences.

The accelerometer obtains information in three dimensions (the x-axis, y-axis,

[1] University of Rzeszow Fall Detection (URFD) dataset. http://fenix.univ.rzeszow.pl/~mkepski/ds/uf.html

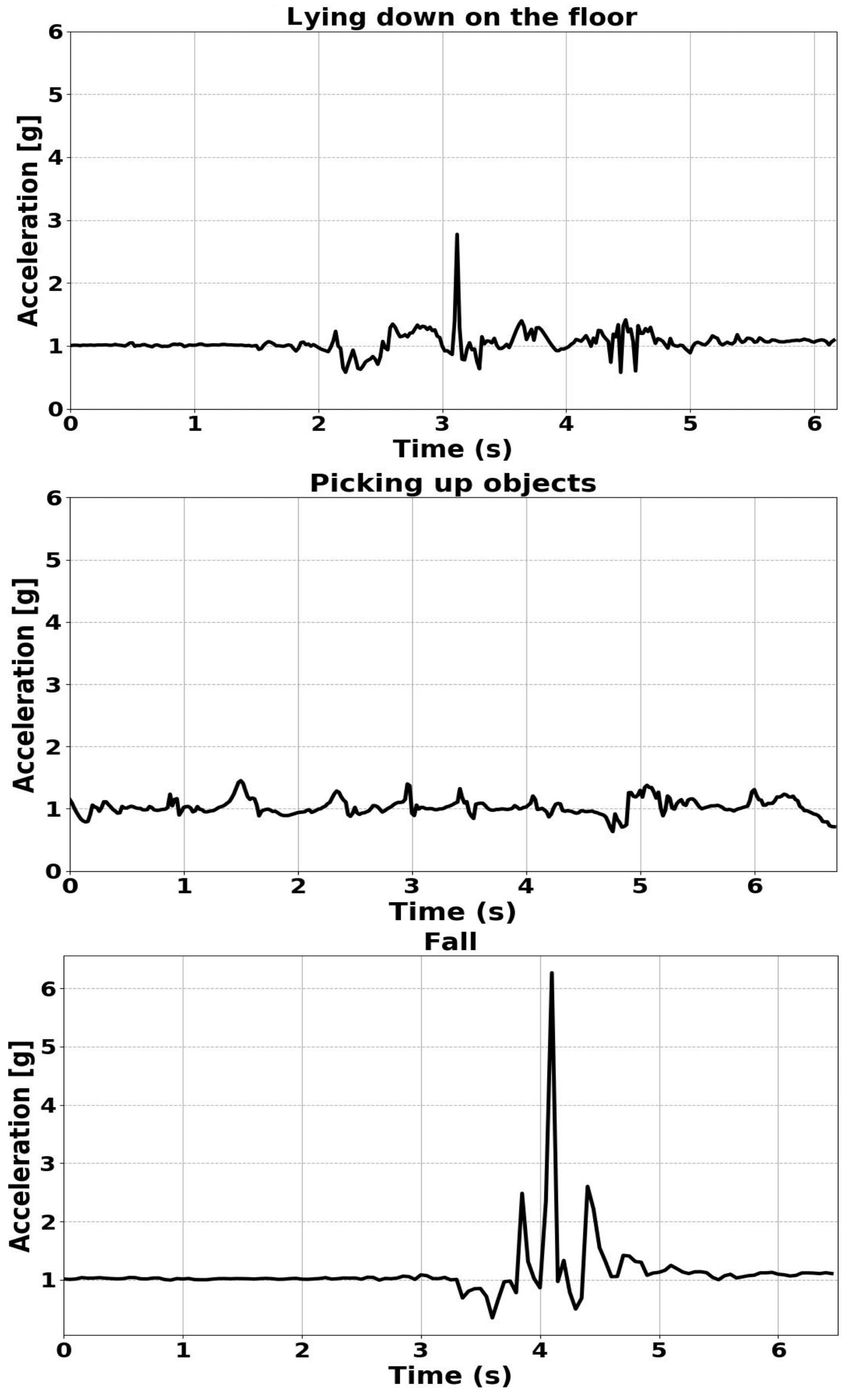


Figure 4.2: Examples of acceleration over time for URFD datasets representing; a) lying down on the floor, b) picking up an object and c) fall.

and z-axis) at time t, which are used to compute the magnitude of acceleration M as follows:

$$M(t) = \sqrt{A_x^2(t) + A_y^2(t) + A_z^2(t)} \tag{4.1}$$

where $A_x(t)$, $A_y(t)$, and $A_z(t)$ represent acceleration in the x, y, and z axes respectively at time t. Therefore, the magnitude of acceleration M is used as an input vector to entropy measure. The magnitude is converted to a set of data points equally spaced in time, and dependent on the calculation period of entropy measures.

4.3.1.4 Dataset D

The dataset used for this research is a dataset publicly shared through the University of California Irvine (UCI) Machine Learning Repository[1] [181]. To collect the data, motion sensor, pressure sensor on a sofa, a magnetic sensor on the fridge door, an electric sensor measuring microwave usage and a door entry sensor were used. The dataset comprises information regarding the ADLs performed by two users daily in their own homes. Moreover, the ADL dataset comprises 35 days of fully labelled data. It is explained by three text files. The first file is a description, which describes these data in terms of the number of rooms in the home and the number of sensors installed in the home. The second file relates to sensor events (features) and includes information in the data such as the date, start time and end time, the location of sensors in the home, and the sensor types. The final file is the activities of daily living (labels) which include activities together with the start time and the end time of each activity. To have a dataset representing multi-occupancy scenarios, a synthetic dataset simulating a visitor is injected into the datasets, which represents a visitor who comes 3 times a week and stays in the house for a couple of hours. The visitor comes around at 11:00 am and 7:00 pm. However, there are some variations within the times and periods of the visits. For example, on some days the visitor comes one hour early or late. In our investigation, only motion sensors

[1]UCI Machine Learning Repository. http://archive.ics.uci.edu/ml/datasets/Activities+of+Daily+Living+%28ADLs%29+Recognition+Using+Binary+Sensors

representing the resident in an area of the home and door sensors are utilised.

4.3.1.5 Dataset E

The dataset was obtained from a real home environment representing the ADL of a single resident for a period of 65 days. The dataset was collected at the Smart NTU home facilities within Nottingham Trent University. The house is equipped with several low-cost, non-intrusive ambient sensors such as PIR sensors, pressure sensor on sofa and bed, and door entry sensors, which are utilised as data collection devices. Due to privacy, cost issues, and ethical concerns, these sensors are the most widely used for ADL monitoring, as they allow individuals to live normally without feeling restrained by the technology [17, 182]. Moreover, these sensors track the resident's interaction in different locations in the house. A floor plan of the house and sensor locations utilised for data collection are shown in Figure 4.3.

The data gathered by these sensors are binary in the form of 1's and 0's signifying active and inactive states, respectively. In total, the dataset contains 56 normal days of ADLs, including, sleeping, eating (dining room activity), toileting, and going out of the home etc., and 9 abnormal days of ADLs. In addition to this, the abnormal days contain different abnormalities in the resident's activity, such as irregular sleep and the presence of a visitor on some days. Besides this, the information that can be obtained from the dataset is the date, start time, end time, and the location of activities. This dataset is fully annotated using self-report and visually inspecting the raw sensor data. Research team members are asked to register the information about the irregular sleep and visits they received any day.

Consideration of ethical issues prior to data collection is an important step to protect the rights of participants and inform them about the procedures. The data collection for the above experiment was conducted using a research team member, and the research was conducted according to the institutional ethical approval process.

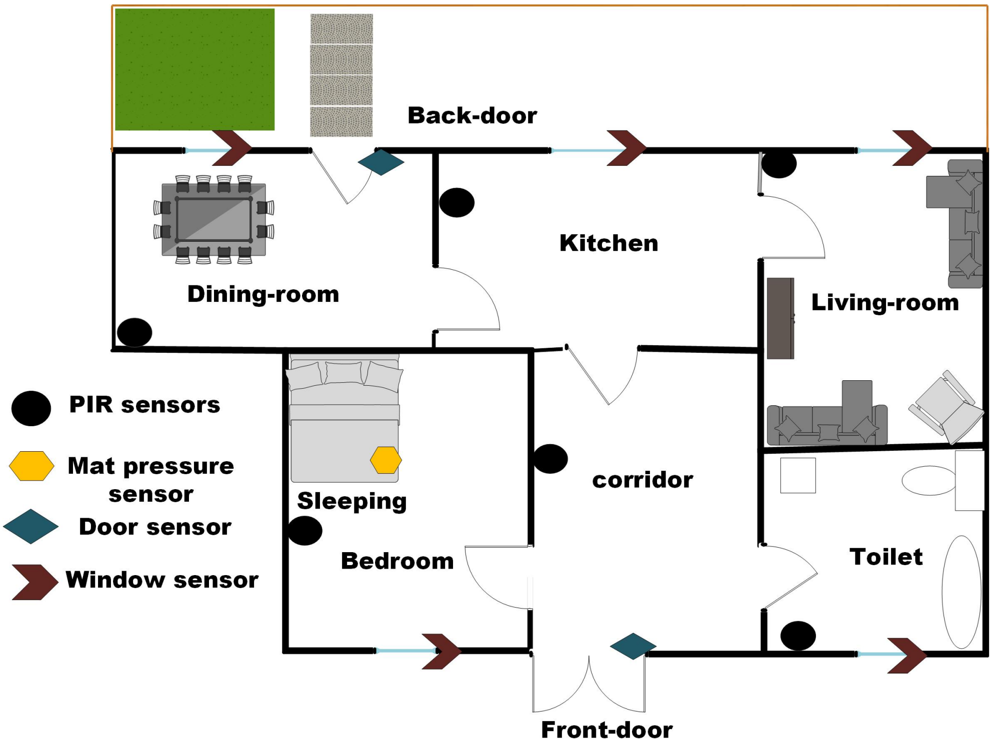


Figure 4.3: Floor plan layout and location of the installed sensors used for data collection in dataset E.

4.3.2 Simulated Environment

Conducting experiments in a real home environment could be very time consuming and expensive to run. As an alternative solution, a simulated environment could be considered to generate the required data and evaluate the research hypothesis.

In this research, an extensive data sample is required to test and evaluate better approaches for anomaly detection in an intelligent environment, which in most cases can not be collected from a real home environment. Therefore, a simulated environment is utilised to generate datasets similar to the datasets gathered from real environments without hardware costs [183, 184]. This simulated environment is equipped with different simulated sensors, such as door entry sensors, temperature sensors, humidity sensors, motion detectors, and light sensors.

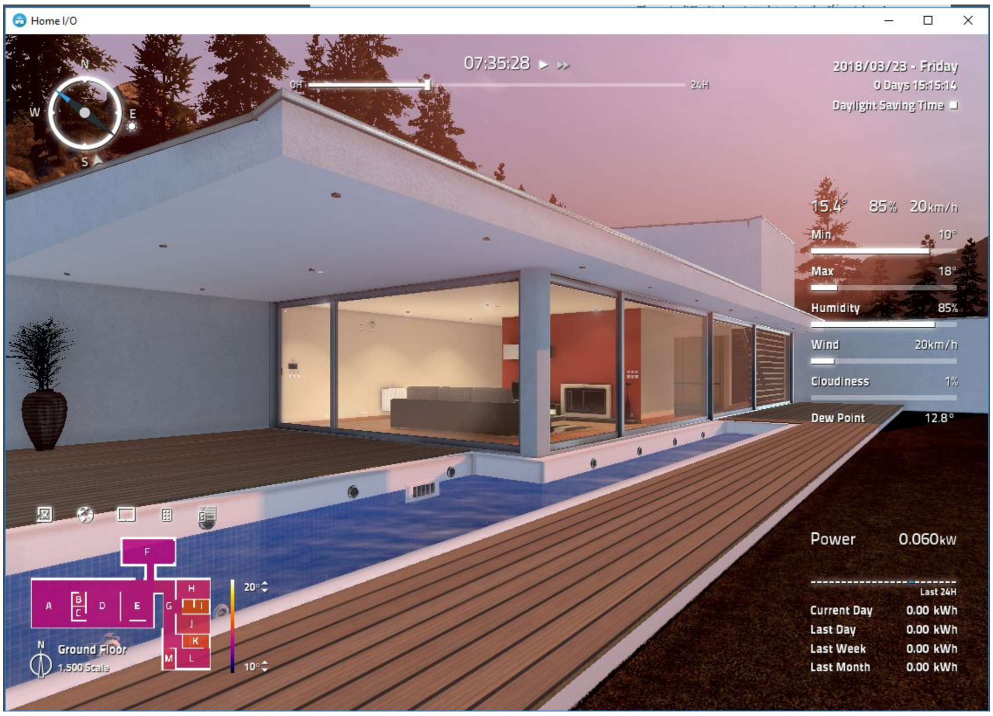


Figure 4.4: HOME I/O 3D Smart Home Simulation software package.

4.3.2.1 Dataset F

The HOME I/O simulation environment[1] [185] is used to gather sensory data representing the ADLs of a single or multiple occupancy. A screen shot of the HOME I/O simulation environment is shown in Figure 4.4. More than 400 input and output (I/O) points are provided by the simulator for collecting information representing various locations of a resident in the home. The simulator is equipped with different simulated sensors such as door entry sensors, temperature sensors, humidity sensors, PIR motion detectors, and light sensors. The occupant's interaction in different locations in the home is tracked by these sensors.

Figure 4.5 shows the floor plan and the sensor locations of the simulated environment. In this work, only PIR sensors that can track the movements of the residents within the home environment are utilised. In order to have a dataset representing multi-occupancy scenarios, the PIR data representing the movement

[1]HOME I/O 3D simulation environment. https://realgames.co/home-io/

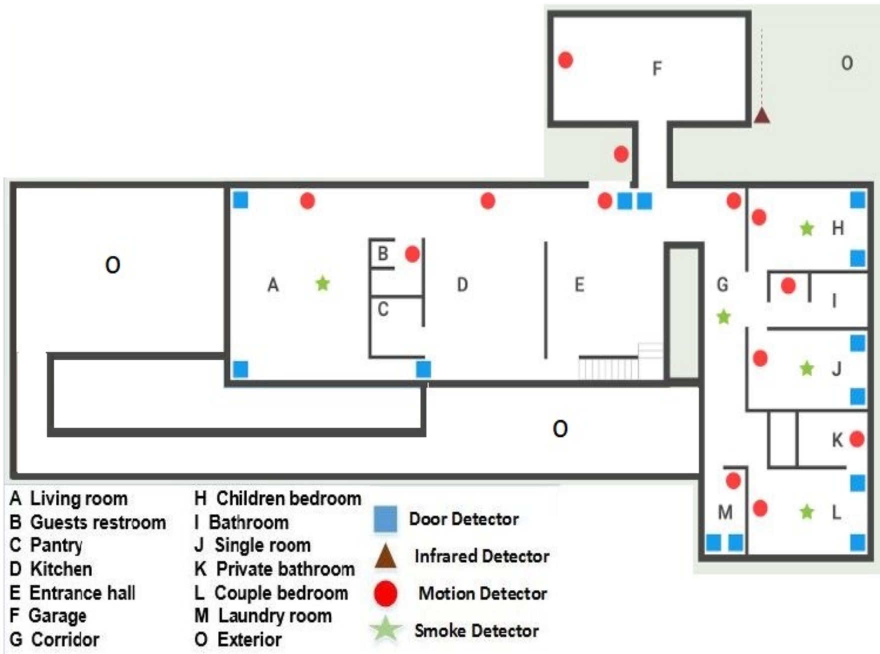


Figure 4.5: Floor plan and sensors layout in smart home Simulation software package.

of a single-occupancy within the home environment for a period of four days is gathered. Secondly, the data for an additional person, which represents the visitor entering the same environment one time during the third day and three times during the fourth day for a limited time period, is injected into the existing data. The dataset representing ADLs for a single-occupancy with the visitor is shown in Table 4.2. The information that can be obtained from the dataset are the date, time and the location of each sensor, as well as the event that has activated the sensor.

4.4 Data Preprocessing

The sensor data gathered through the acquisition process could be noisy (for instance, too many outliers), unreliable data, missing data values and sometimes false data. The false data could be a false positive or false negative. The false-positive refers to a dataset that does not include an anomaly in ADLs but is incorrectly identified as an anomaly. Moreover, The false-negative is a dataset

Table 4.2: A sample of the gathered dataset, representing ADLs for a single-occupancy with the visitor (represented as a multi-occupancy environment), from the HOME I/O simulation environment.

Date and Time	Sensors ID	Location	State
2018-11-04 09:06:22	M004	Bathroom	1
2018-11-04 09:08:26	M004	Bathroom	0
2018-11-04 09:08:28	M006	Bedroom	1
2018-11-04 09:08:28	M008	Corridor	1
2018-11-04 09:10:33	M006	Bedroom	0
2018-11-04 09:10:35	M020	Living-room	1
2018-11-04 10:08:28	M008	Corridor	1
2018-11-04 10:08:36	M018	Kitchen	1
2018-11-04 10:10:01	M018	Kitchen	0

that includes an anomaly in ADLs but is incorrectly identified as normal. If the data gathered contains false data and unrelated or not enough information, machine learning algorithms could produce less accurate and misleading results or could fail to detect anything of use at all [186]. Thereby, the purpose of data preprocessing is to convert the data gathered (raw data) into the right form required for a model. Data preprocessing comprises data cleaning, handling missing data, normalisation, feature extraction, conversion, and selection, etc. In this regard, the following steps are taken in preprocessing the data acquired:

4.4.1 Handling Missing Data

Missing data can occur due to software or hardware faults. In general, there are three types of missing data according to the mechanisms of missingness including, Missing At Random (MAR), Missing Completely At Random (MCAR), and Missing Not At Random (MNAR). In MAR cases, the missing data depends on some other observed data but is unrelated to actual values of the missing data. For example, if the sensor utilised in obtaining data is out of action for some time, it is unlikely to be related to the activity performed. Whereas data MCAR happens when the missing data are not related to either specified values to be acquired or observed. In the case where the characters of the missing data do not meet those of MAR or MCAR, then they fall into the category of MNAR. The only way to handle such cases of MNAR is to model the missing data [187].

To solve the challenge of missing data, the following steps should be taken:

- Use what is known about the data gathered and then understand how to distribute the missing data.
- Attempt to discover the reason for missing data.
- Choose the best analysis technique for handling missing data in order to obtain the least biased estimates.

There are several techniques to deal with missing data. The techniques commonly utilised include, listwise or case deletion, pairwise deletion, mean substitution, last observation carried forward, maximum likelihood, sensitivity analysis and multiple imputation [188]. The list-wise deletion is the most frequently utilised method which includes gaining repeated measurements over a time series. It is used to omit those observations with missing data and just utilise the residual data for analysis. This research depends on this technique to address missing data of observed daily human activities. Due to the data being large enough and the missing data hypothesis satisfying the MCAR, the list-wise deletion technique is the preferred solution.

4.4.2 Visualisation of Sensor Data

The majority of the dataset used in this research is mainly based on information obtained from ambient sensory devices network representing the ADLs of a resident within a smart home environment. This data include a large volume of binary string data. Therefore, data visualisation is utilised to aid in understanding the real datasets as a primary step of analysing the data. The use of data visualisation obviously can assist in facilitating the examination of large amounts of data.

There are several techniques used to visualise the binary sensor data, including visualisation based on start-time and duration. It is one of the helpful visualisation methods which can aid in understanding the binary data

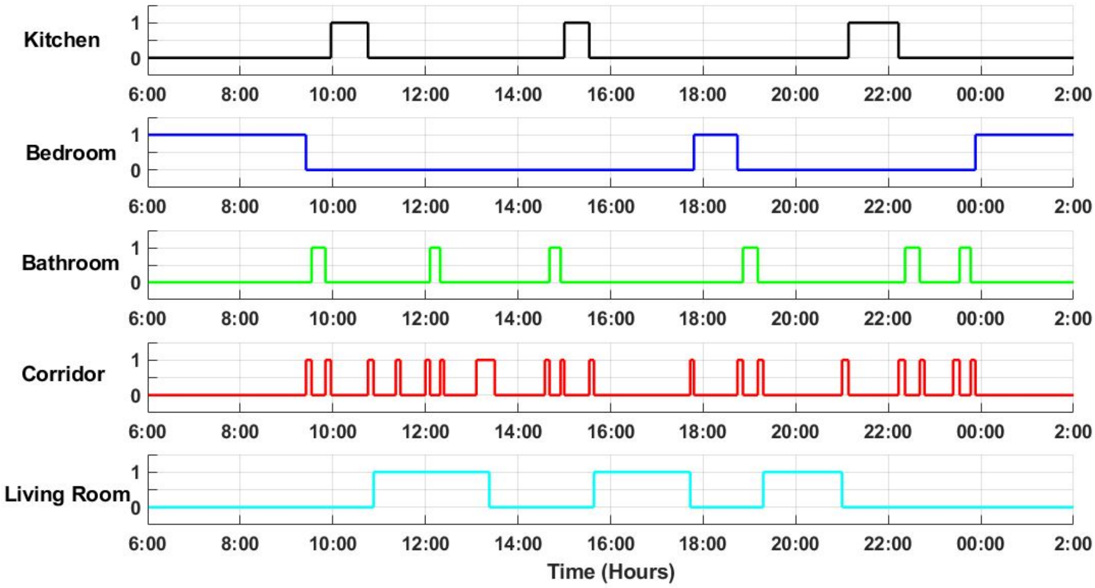


Figure 4.6: A sample of sensor data gathered from Passive Infra-Red (PIR) sensors in various locations over one-day period, where the y-axis represents the sensor status (on/off) as a binary value in different locations; and x-axis represents time in hours.

sequences. Using start-time and duration that are extracted from occupancy sensors, the occupant's movements sequences can interpret who is utilising a smart home environment, and it will be used to show the pattern of the resident. To illustrate the visualisation of the sensor data, Figure 4.6 shows sensor data gathered from 5 PIR sensors in various locations over a one-day period and the duration in hours, spent by the resident in each room. It is clear from this example that the behaviour of the resident can be more easily interpreted. For instance, in Figure 4.6, the bedroom sensor plot shows that the resident always goes to bed at midnight around 12:00 am. Moreover, the bathroom sensor plot shows that the resident goes to the bathroom six times a day. However, it is challenging to achieve this level of understanding if the visualisation and tracking of movements of this occupant represented more days (e.g. a month).

4.5 Feature Extraction

Feature extraction is a significant aspect of any activity recognition system because, as raw data obtained from activities are not able to provide enough information to permit implementing an activity recognition system. Once the sensor data is gathered from a home environment, the daily behaviour features for occupancy are computed. The selected features representing ADLs from the sensor data are:

- Start time: This is the starting hour and minutes of entering each location (room) in the house.
- Duration: This is the duration in minutes the resident spends in each room, which is obtained by subtracting the end time from the start time.
- The transition between the rooms: This is the transition from the location of the performed activity to another location inside the home.
- Encoded daily activities sequence: This is the collection of activated sensors locations at different times, in which each location (room) is encoded by replacing each activity and/or the location of the performed activity with an odd number (e.g., toilet = 1, bedroom-sleeping = 3, corridor = 5, kitchen = 9, etc.). It was considered that the higher numbers are related to rooms that were frequently utilised for shorter time periods (here, the corridor).

The above numerical features are calculated from the sequences of the input vectors extracted from the gathered data (Dataset A, Dataset B, Dataset D, Dataset E and Dataset F), solely based on information gathered from low-cost, non-intrusive ambient sensors, such as Passive Infra-Red sensors and a door entry sensor. However, regarding Dataset C which is gathered using a wearable motion-sensing device, the magnitude of the acceleration M is used as an input vector to the entropy measures as mentioned in Section (4.3.1.3).

Table 4.3: A sample activity data used to calculate the pre-processed input sequence vector for the entropy measures.

Start Time	Duration (min)	Location	Encoded number of each location
15:00:33	8	Living room	7
15:08:57	1	Corridor	11
15:09:00	3	Bathroom	3
15:09:00	3	living room	7
15:12:12	1	Corridor	11
15:13:10	15	Kitchen	1
15:28:00	1	Corridor	11
15:29:33	30	Bedroom	5
15:59:17	1	Corridor	11
16:00:00	22	Living room	7

4.6 Entropy Calculation

The input of any entropy measure should be formulated as a vector sequence (time series) as described earlier in Chapter 3. Therefore, to represent the dataset appropriate for entropy measures, the encoded dataset is converted to a set of data points equally spaced in time, which is dependent on the computational time of the entropy measures. The encoded daily activity sequence is then utilised as an input vector for entropy measures in Chapter 5 and Chapter 6. The entropy measures are utilised to measure the abnormality in the patterns of daily routines when the sample data is mostly representing normal activities.

To explain the process of how the vector sequence is obtained from the dataset, a step-by-step example is provided below. Consider the activity data sample presented in Table 4.3. Firstly, the required numerical features to be used for calculating the vector sequences are extracted from the raw dataset. Then, the daily activity sequence is encoded by replacing each location (room) with an odd number, as shown in the fourth column of the Table. Finally, the features extracted from the raw data are used with the encoded daily activities as input vector sequences to the entropy measures. The entropy measures are computed every hour, which means that there are 60 samples per hour. For the sample data presented in Table 4.3 from 15:00 to 16:00, so the activity sequence vector A_N, which consists of a 60 sample of the encoded daily activity equally spaced in time is then defined as:

$$A_N = [\underbrace{7, 7, 7, 7, 7, 7, 7, 7}_{\text{Duration}}, 11, 3, 3, 3, 11, 1, ..., 5, 5, 5, 5, 5, 5, 5, 5, 5, 11]$$

It is obvious from the given vector sequence A; the repetition of the same number reflects the time spent in each room (duration). The values of entropy measures will be computed every hour by repeating the same step. Once more than one sensor is activated at the same time, only the value of one sensor will be considered (i.e. the first activated sensor) to compute the input vector sequences as shown in Table 4.3. For instance, given that both sensors of the bathroom and living room are active at a particular time, the first activated sensor is considered to be used in the vector sequence A_N.

4.7 Discussion

This chapter presented an overview of ambient intelligence environment and data collection system employed in this research. The challenging tasks of processing the big datasets gathered from a network of sensors are also explained.

Real and simulated datasets are described in this chapter. Some datasets examples from different real environments are presented. A simulator is built to support this research by producing simulated datasets. Moreover, this chapter presented an explanation of the required numerical features to be used for calculating the sequences of the input vector for entropy measures.

The encoded daily activity sequence extracted from the raw data is used in Chapter 5 and Chapter 6, as inputs to entropy measures. In particular, they will be utilised to evaluate the proposed entropy measures for anomaly detection in activities of daily living.

Chapter 5

Anomaly Detection in Activities of Daily Living

5.1 Introduction

Anomaly detection in the ADLs of older adults is essential for healthcare management. It aids avoidance of future problems which in turn improves the quality of life. Existing methods of anomaly detection in ADLs ignore the changes in individuals' routine, thereby limiting their accuracy and reliability [52]. Hence, it is important to develop an appropriate method or algorithm that can effectively detect anomalies in older adults' daily activities.

This chapter aims to investigate the effectiveness of different entropy measures mentioned in the earlier Chapter 3 Section 3.3, in detecting and identifying various types of anomalies within the behavioural patterns of a resident in a smart home environment. Detecting anomalies in sleeping pattern, human falls, and ADLs in the presence of a visitor are the main focus of the work presented here as case studies. The entropy measures introduced earlier in Chapter 3 Section 3.3 are applied to the ADL datasets presented in Chapter 4 Section 4.3 representing the aforementioned anomalies.

As a starting point for detecting anomalies in ADLs, the investigation of the effectiveness of entropy measures initially focuses on applying one type of entropy measure (Multi-scale Fuzzy Entropy (MFE)) to investigate whether the MFE measure can be used for anomaly detection in ADLs, specifically in irregular sleep. This is described in Section 5.2. Then, the research investigates the effectiveness of another type of entropy measures (Fuzzy Entropy (FuzzyEn)) to detect and distinguish human fall from other activities, solely based on data gathered from wearable devices and this is described in Section 5.3. To evaluate the proposed method carried out in this research, the results obtained by applying the FuzzyEn entropy measure for human fall detection are compared to other methods or algorithms using the same dataset. Comparisons with other methods have also provided further support to the proposed method. Finally, all entropy measures are applied for anomaly detection in daily activities in the presence of a visitor (here, identifying visiting times and irregular sleep), solely based on information gathered from ambient sensors. Furthermore, it investigates whether entropy measures can be used effectively for anomaly detection in ADLs where anomalies are diverse and normal samples are relatively homogeneous. This is described in Section 5.4.

The remainder of this chapter is organised as follows: Section 5.2 presents a novel method based on Multi-scale Fuzzy Entropy to identify and distinguish between normal and anomalous events in ADLs, specifically in sleeping routine. In this section, the experimental results, evaluation of the performance, and comparison of the proposed method with other methods are presented. Section 5.3 investigates how Fuzzy Entropy measure can be used to detect human falls in a home environment, explains the experimental results and their evaluation. Section 5.4 presents a novel entropy-based method to detect anomalies in ADLs in the presence of a visitor, solely based on information gathered from low-cost, non-intrusive ambient sensors, the experimental results, and robust analysis. Finally, the pertinent conclusions of this chapter are drawn in Section 5.5.

5.2 Case Study 1: Irregular Sleep Detection

Several research studies have investigated methods to detect normal and abnormal human behavioural activities using different computational methods [50, 55]. However, there are some limitations to these approaches, which includes the fact that they do not take into account changes in individual routine [52]. Human behaviour is dynamic, and through an individual's life behaviour changes due to factors such as social and health influences. Reliable anomaly detection in ADLs is considered as one of the most important components of many home health care applications [154]. However, existing methods are not able to reliably detect anomalous events in ADLs and therefore generate a false alarm rate [36].

In many applications, entropy measures are utilised to quantify the concept of irregularity and the degree of randomness in a system. Nevertheless, to measure the subjective value of information under the condition of uncertainty, the Multi-scale Fuzzy Entropy measure is considered as a useful measure to discriminate between normal and anomalous cases in daily activities. One of the challenges addressed in this research is detecting anomalies in ADLs using low-cost, non-intrusive ambient sensors. This research aims to investigate whether the MFE measure can be used to detect and distinguish anomalies in ADLs, specifically in sleeping routine, which could be a sign of MCI in older adults.

5.2.1 System Overview

This study proposes a method for anomaly detection in ADLs, solely based on low-cost, non-intrusive ambient sensors such as the Passive Infrared (PIR) sensor. The research assumes that the level of changes in a resident's ADL patterns in a home environment is an indicator of normal or abnormal activities, as shown in Figure 5.1. Therefore, the MFE measure is utilised to quantify the concept of irregularity and uncertainty in the ADL data. This method can be used for detecting abnormalities in ADLs when the activity data represents normal activities for most of the time. The proposed method is based on the hypothesis that when the value of the MFE measure surpasses standard deviation boundaries, then the case should be indicated as an anomaly in ADLs. The proposed method aims to classify any values exceeding the standard deviation boundaries, as abnormal. Thus, the MFE measure enables the data to be identified as either normal or abnormal. After an extensive investigation, it was identified that MFE is the most suitable measure for discriminating between normal and anomalous cases in daily activities.

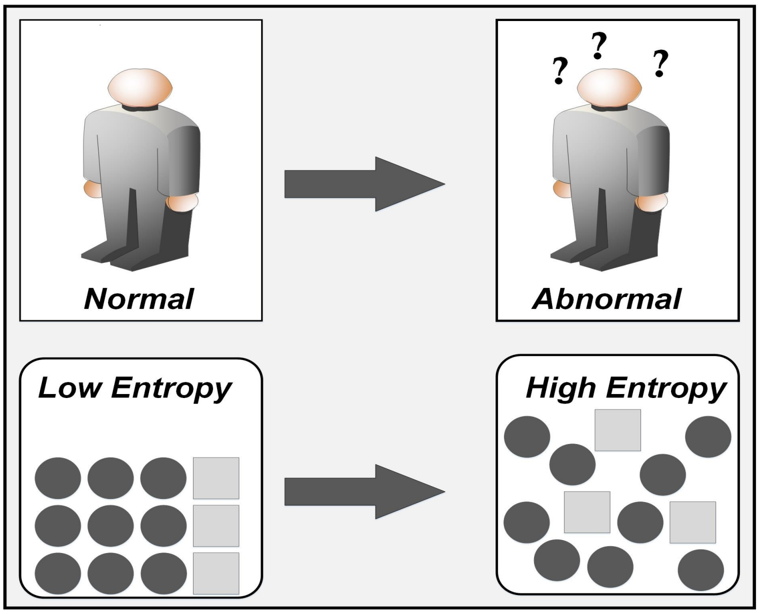


Figure 5.1: Overview of the proposed anomaly detection in activities of daily living.

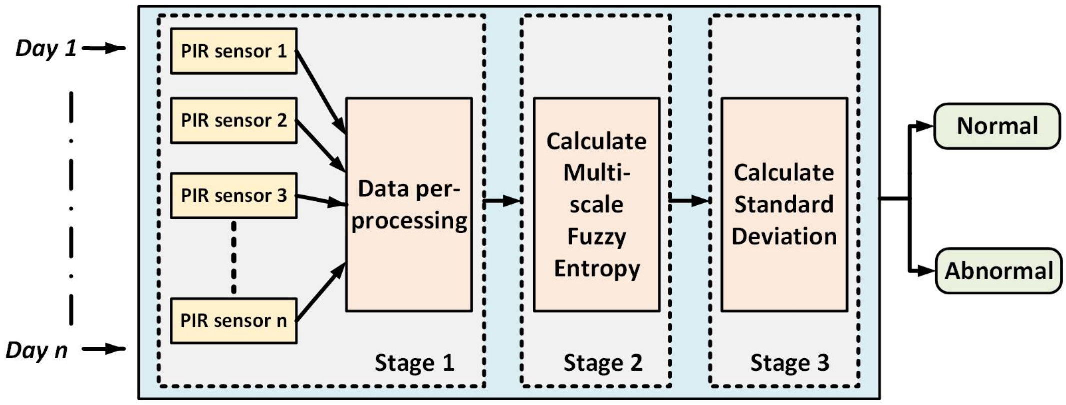


Figure 5.2: A schematic diagram of the proposed method for anomaly detection in activities of daily living.

Figure 5.2 provides a schematic diagram of the proposed method for anomaly detection in activities of daily living, which comprises three main stages:

- In the first stage, sensor data representing ADLs in a home environment are collected and pre-processed. The required numerical features to be utilised for computing the input vector sequences of the entropy measure are extracted from the raw data.

- In the second stage, the MFE measure is applied to the data collected to identify abnormalities in daily activities.
- In the third stage, the standard deviation is applied to distinguish whether there are anomalies in the resident's activity or not.

5.2.2 Experimental Setup and Results

To evaluate the proposed concept for identifying anomalies in activities of daily living, two annotated datasets; Dataset A and Dataset B described in Chapter 4 Section 4.3.1.1 and Section 4.3.1.2, respectively, are used. For this research, only the sleeping activity and the activities that occur before and after the sleeping activity are considered. The relevant features that can distinguish between normal and anomalous cases in sleeping activity are selected. The selected features representing ADLs from the sensor data are: the start time of each activity, the duration of each activity, and the transition from the location of the performed activity to another location inside the home.

The input to the MFE measure should be formulated as a vector sequence A_N (time series) as described in Chapter 3 (Section 3.3.7). Therefore, to represent the dataset suitable for MFE calculation, the dataset is transformed into a set of data points equally separated in time, which is dependent on the calculation period of the entropy measures. The MFE is used to detect abnormalities in the patterns of daily routines when the sample data mostly represents normal activities. To obtain the MFE value, the daily activity sequence is encoded by replacing each activity and/or the location of the performed activity with an odd number (e.g., bedroom-sleeping = 1, toilet = 3, kitchen = 9, etc.). The encoded daily activity sequence is then used as an input to the MFE measure.

The MFE is calculated every day, with 60 samples per hour. Therefore, the vector sequence A_N, which consists of 60 sample set equally spaced in time, is utilised as the input vector for the MFE. The MFE is dependent on three parameters that are needed for calculation: embedded dimension m, tolerance r, and the scale factor s. Thus, the algorithm for MFE is impacted by the selection of these values. The best results are obtained when the values of the parameters m, r, and s are 2, 0.2, and 1 respectively. After the MFE has been computed, the standard deviation of the average MFE values is calculated.

The threshold based on one standard deviation or one sigma is used to distinguish whether there are anomalies in the resident's activity or not. It is possible to use thresholds based on different sigma (e.g., 2 or 3) to identify anomalous days in the resident's activity. However, Our earlier work has concluded that increasing sigma is not sufficiently reliable enough to detect anomalies in ADLs. This can be justified by the fact that increasing sigma, will reduce the number of observations per time period, which will, in turn, increase the variance. As a consequence, the number of false positives will increase, which decreases the precision. The proposed method is based on the assumption that when the value of the MFE measure exceeds the standard deviation boundaries, then this indicates an anomaly in ADLs. Figure 5.3 shows the results obtained by applying the MFE method to the ADL Dataset A. The proposed method identifies 7 days as anomalous because the MFE values for these days exceed the standard deviation boundaries. After identifying 7 anomalous days, the MFE for each of these days is computed again with 30 samples every 1/2 hour to examine the possible causes of the identified anomalous days.

The results in Figure 5.4 show the possible causes of the identified anomalous days for Dataset A. From Figure 5.4(a), it can be observed that the resident has interrupted sleeping patterns (Day 39) because he has multiple transitions from bed to other locations. This reveals that the resident slept from $10:30$ pm to $12:30$ am and then from $2:30$ am to $6:00$ am. Also, Figure 5.4(b) shows that the resident also slept for a shorter period of time on day 55 (from $10:30$ pm until $4:00$ am), sleeping for approximately 5 hours. Meanwhile, Figure 5.4(c) shows that on day 59 the resident went to bed at around $1:30$ am which is late compared to the normal days.

The identified anomalous days and possible causes of these for Dataset A and Dataset B are summarised in Table 5.1 and Table 5.2 respectively.

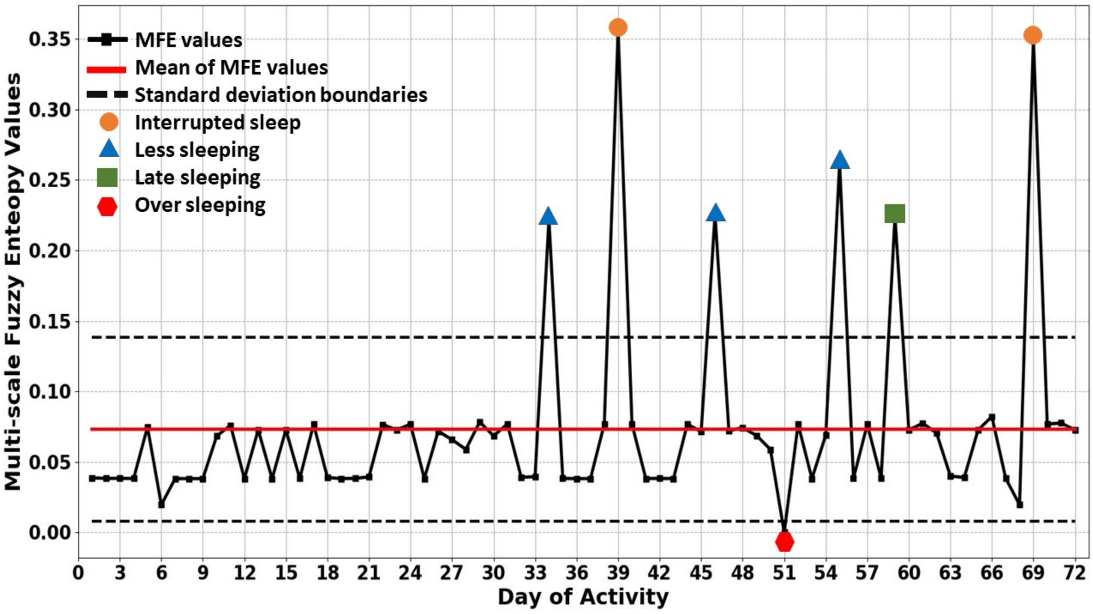


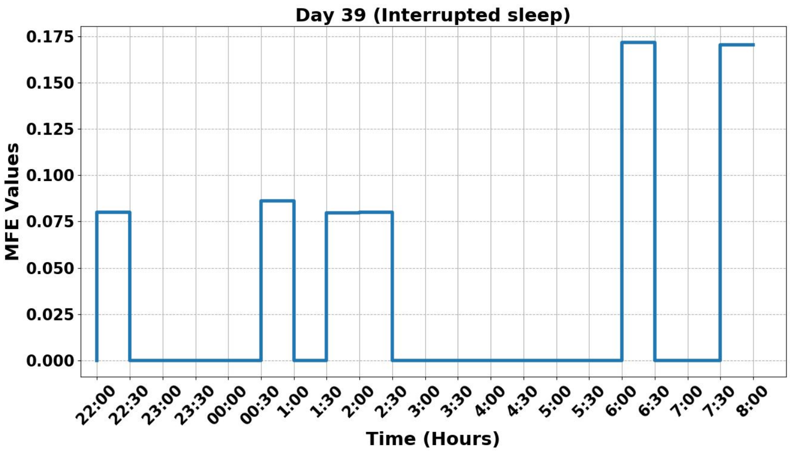
Figure 5.3: The results obtained by applying Multiscale-Fuzzy Entropy (MFE) for anomaly detection in the activities of daily living Dataset A. The figure also illustrates the standard deviation boundaries and the average value of MFE for 72 days.

Table 5.1: A summary of identified anomalies and possible causes of these for Dataset A.

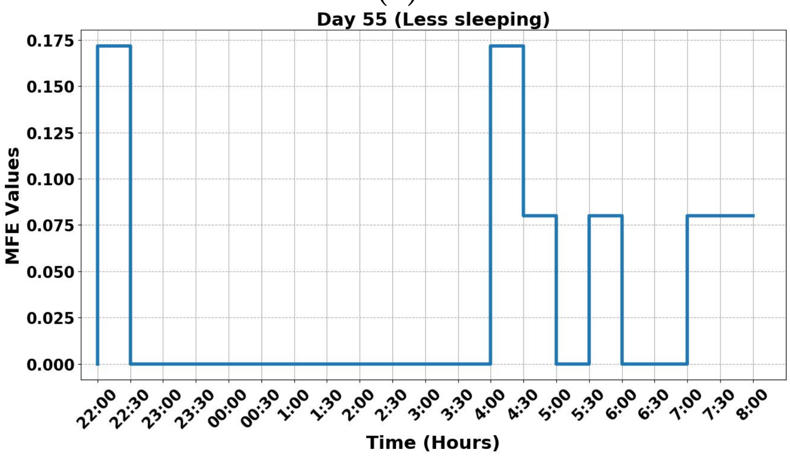
Day	Cause	Detailed description
Day 34, 46, and 55	Less sleep	The individual sleeps for a short time period compared to their usual pattern of sleep.
Day 39 and 69	Interrupted sleep	The resident has multiple transitions from the bed to the toilet and other locations in the house.
Day 59	Late sleeping	The person goes to bed late compared to the usual days.
Day 51	Over sleeping	The resident sleeps for a long period of time compared to the usual days.

5.2.3 Performance Evaluation

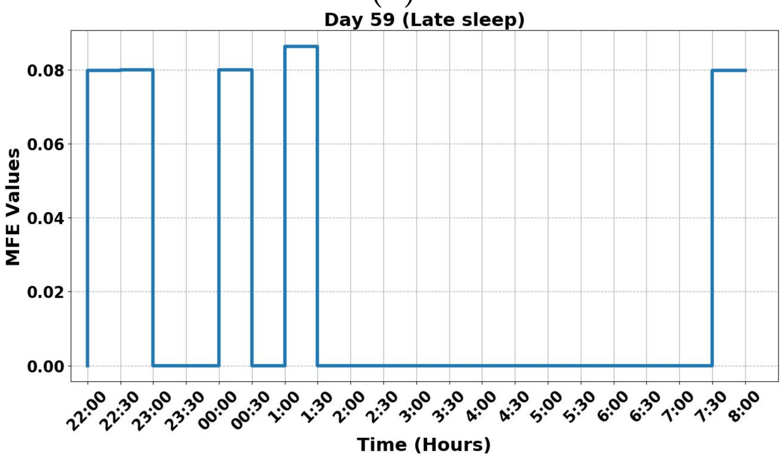
To evaluate the performance of the proposed method, the two datasets (Dataset A and Dataset B) representing ADLs of older adults, are manually labelled as normal or as abnormal in the resident's activity. Table 5.3 shows that 65 days are



(a)



(b)



(c)

Figure 5.4: Examples of identified anomalies and possible causes of these for Dataset A representing; a) interrupted sleep, b) less sleep and c) late sleeping.

Table 5.2: A summary of identified anomalies and possible causes of these for Dataset B

Day	Cause	Detailed description
Day 32, 33, and 47	Afternoon sleeping	The individual spends over $1-2$ h napping during the day compared to usual days.
Day 37, 39, 44, 46 and 49	Interrupted sleep	The user has multiple transitions from the bed to the toilet and other locations in the house.
Day 50	Less sleep	The person only sleeps for approximately 2 h compared to the usual days.
Day 26	Method error	No deviation has been identified from usual days.

indicating normal activity and 7 days are indicating abnormalities in the resident's activity, based on Dataset A. It can be observed from Table 5.4 that 41 days are indicating normal activity and 9 days are indicating an abnormality in the resident's activity based on Dataset B. The MFE method successfully identified all anomalous days for Dataset A. However, for the normal activity included in Dataset B, the MFE method identified 40 days as being normal activity out of 41 days and miss-classified only one day.

The performance evaluation is calculated automatically using a confusion matrix. There are four possible outcomes for testing anomaly detection in ADLs, which are defined as follows:

- True Positive (TP): a dataset contains an anomaly in ADLs, and this is

Table 5.3: Detection Accuracy of MFE for Dataset A.

Events	Total Days	Identified	Not identified
Normal	65	65	0
Abnormal	7	7	0

Table 5.4: Detection Accuracy of MFE for Dataset B.

Events	Total Days	Identified	Not identified
Normal	41	40	1
Abnormal	9	9	0

Table 5.5: Performance of the MFE Method for Dataset A and Dataset B.

Description	Results Obtained	
	Dataset A	Dataset B
Sensitivity	100%	100%
Specificity	100%	97.5%
False positive rate	0%	2.5%
False negative rate	0%	0%
Accuracy	100%	98%

correctly identified as an anomaly.

- False Positive (FP): a dataset does not include an anomaly in ADLs but is incorrectly identified as an anomaly.
- True Negative (TN): a dataset does not contain an anomaly in ADLs and is correctly identified as normal.
- False Negative (FN): a dataset includes an anomaly in ADLs but is incorrectly identified as normal.

The performance evaluation of the proposed entropy measure is evaluated using:

$$Sensitivity = \frac{TP}{TP + FN} \tag{5.1}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5.2}$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP + TN} \tag{5.3}$$

$$False\ Negative\ Rate\ (FNR) = \frac{FN}{FN + TP} \tag{5.4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.5}$$

The results presented in Table 5.5 show the performance of the proposed anomaly detection algorithm on Dataset A and Dataset B. The MFE method achieves 100% specificity using Dataset A, which means that all anomalous days have been correctly identified. However, the proposed method only achieves 97.5% specificity using Dataset B, and this means that one of the normal days is identified as an anomaly. The accuracy of anomaly detection in ADLs for Dataset A and Dataset B are 100% and 98%, respectively. The MFE method shows high detection rates of 100%, for Dataset A, which means that the false negative rate of anomaly detection is 0%. However, the MFE method achieves a detection rate of 97.5% for Dataset B, which means that the MFE method has a 2.5% false negative rate for anomaly detection.

Based on the results achieved, the MFE measure is a powerful tool to detect anomalies (here, anomalous sleep activity) in behaviour when the sample data mostly represents normal activities. This also confirms that the MFE measure could be utilised for anomaly detection in ADLs.

5.2.4 Comparison of the Proposed Method with Existing Methods

To evaluate the proposed method, a comparison is made with the CNDE approach [68] and the approaches proposed in [189], namely; Ensemble of Detectors with Correlated Votes (EDCV) and Ensemble of Detectors with Variability Votes (EDVV). Readers are referred to as [189] for further details about these approaches. The EDCV and EDVV are applied to the same datasets used for the research in [68]. The results obtained by applying the MFE entropy measure are compared to the other methods using the same dataset. Comparisons are made based on accuracy, as shown in Table 5.6.

Based on the presented results, it can be argued that the proposed method outperformed other approaches. The accuracy of the MFE method for anomaly detection in ADLs for Dataset A and Dataset B is 100% and 98%, respectively. It can also be confirmed that the MFE measure is considered as a useful method and can be utilised for anomaly detection in ADLs.

Table 5.6: Comparison of the Proposed Method with Other Methods Based on Accuracy

Approaches	Dataset A	Dataset B
EDCV	92.9%	83%
EDVV	90.1%	81.6%
CNDE	98.5%	95.7%
Our method (MFE)	**100%**	**98%**

5.3 Case Study 2: Human Fall Detection

To support older adults with their independent living, assistive technologies such as automated fall detectors are utilised to assist and support them to live safely in their own homes [73]. Several research studies have been carried out on detecting human falls during daily activities, using different approaches. In this study, Fuzzy Entropy (FuzzyEn) measure is investigated to detect and distinguish human fall from other activities. Distinguishing and detecting falls for older adults is essential for healthcare management [45, 88, 89]. Therefore, it is important to develop an accurate system with the ability to detect older adults' falls in their daily activities. This research aims to investigate whether Fuzzy Entropy measure can be utilised to detect human falls during daily activities.

5.3.1 Methodology

This study proposes a method for detecting human falls in the home environment, solely based on the information gathered using a wearable motion-sensing device. Since the resident's normal daily activity pattern is completely different when an abnormal event has occurred, the data recorded from accelerometer devices during daily activities is used to show abnormal (e.g. fall) patterns. The research hypothesis is that the level of changes in the resident's ADL patterns in a home environment is an indicator of normal or abnormal activities. Therefore, the entropy measure could be used as an indicator of the level of randomness in the accelerometer data. This method can be utilised for detecting abnormalities

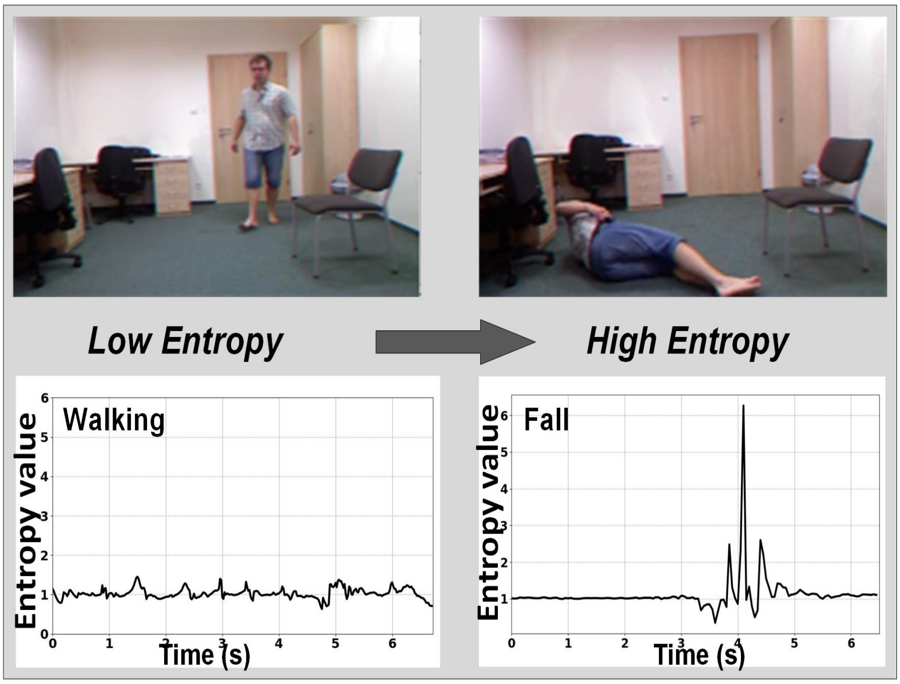


Figure 5.5: Overview of proposed human fall detection in activities of daily living.

when the sample data is mostly normal. The proposed method is based on the hypothesis that the value of entropy is high when there is a fall event, as shown in Figure 5.5. Therefore, the proposed method aims to detect a large value of the entropy. It is supposed that human falls have greater acceleration than other ADLs. Nevertheless, considering high acceleration only can lead to many false alarms during fall-like activities such as sitting down speedily [190]. Therefore, a suitable measure must be utilised to distinguish falls from other activities accurately. After an extensive investigation, it was identified that Fuzzy Entropy is the most suitable technique in distinguishing between actual falls and other daily activities.

A schematic diagram of the proposed fall detection framework is shown in Figure 5.6. It comprises three main stages.

- In the first stage, the accelerometer data representing ADLs is gathered and pre-processed.
- In the second stage, Fuzzy Entropy is applied to the data collected to detect

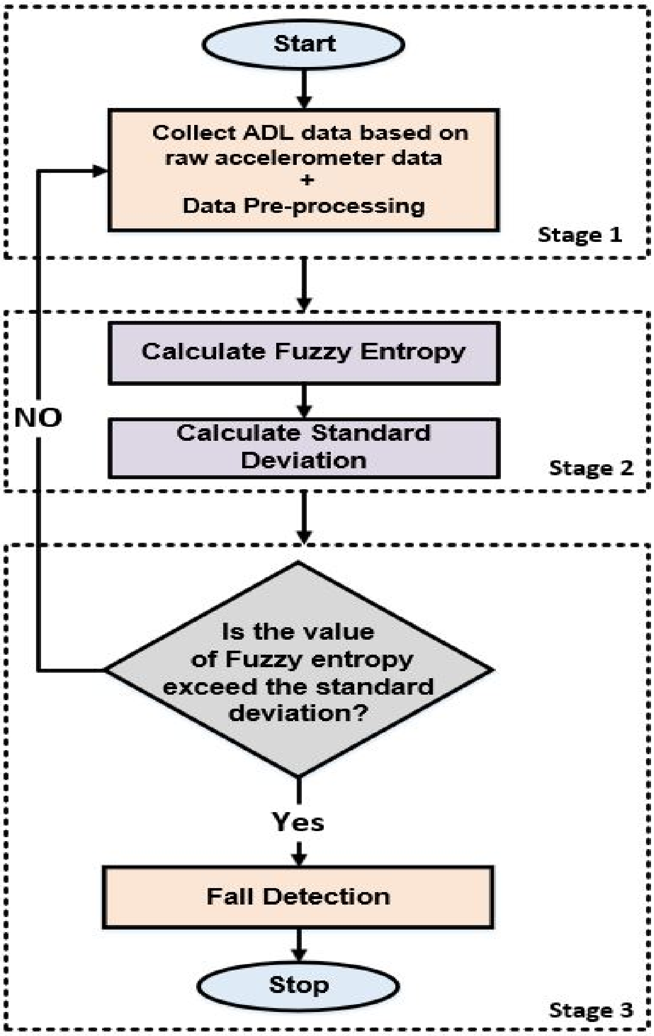


Figure 5.6: A schematic diagram of the proposed method for human fall detection.

abnormalities in daily activities. The standard deviation is then computed.

- In the third stage, the standard deviation is utilised with the Fuzzy Entropy measure to detect whether or not a fall event has occurred.

5.3.2 Experimental Setup and Results

The proposed method has been evaluated based on the annotated Dataset C mentioned in Chapter 4 Section 4.3.1.3. The aim is to determine whether FuzzyEn is a useful measure for detecting human falls in a home environment and whether it might allow the detection of changes in activities of daily living levels. To use Dataset C for FuzzyEn computation, the magnitude of acceleration M is used as an input vector to the FuzzyEn measure. The magnitude is converted to a set of data points equally spaced in time, and dependent on the calculation period of the FuzzyEn measure. The FuzzyEn is computed every second, at 60 samples per second. Therefore, the vector sequence A_N, which consists of a 60 sample set equally spaced in time, is used as the input for FuzzyEn. FuzzyEn is dependent on two parameters, which are required for its computation; embedded dimension m and tolerance r. Therefore, the algorithm for FuzzyEn is affected by choice of these parameter values. The best results are obtained when the values of the parameters m and r are 3 and 0.2 respectively. It appears that when m and r values are increased, the performance of the algorithm is decreased. After the FuzzyEn is calculated, a novel feature, namely the standard deviation of the mean of FuzzyEn values, is calculated.

The standard deviation is applied to confirm whether or not there is a fall. The proposed method is based on the hypothesis that when the value of the FuzzyEn measures exceeds the upper standard deviation boundaries, then the event is detected as a fall. Figure 5.7 shows the results obtained by applying the FuzzyEn method to Dataset C. It can be noted that the fall events were successfully detected because the value of FuzzyEn is higher than the upper standard deviation boundaries.

5.3.3 Performance Evaluation

To evaluate the performance of the proposed method, Dataset C contains 30 falls and 40 activities of daily living as observed in Table 5.7, manually labelled as a fall or non-fall event. The FuzzyEn method successfully detected all the 30 fall events. However, for the other normal activities included in the dataset, the proposed method detected 39 activities out of 40 activities and failed to classify only one activity.

The evaluation of performance is computed automatically using a confusion matrix. There are four possible results for testing a sequence as a fall event in

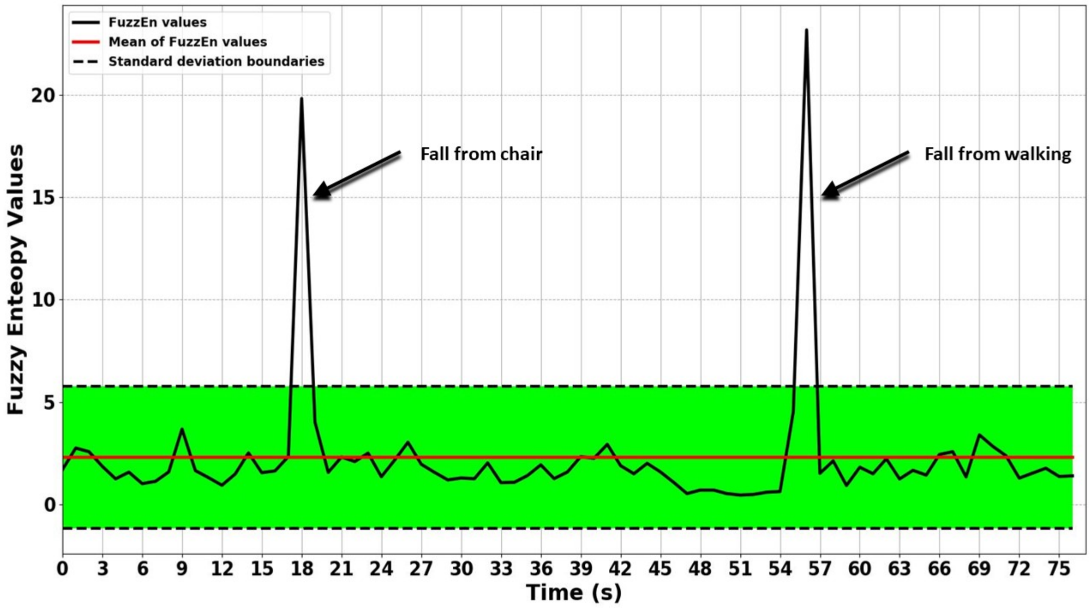


Figure 5.7: Samples detecting one fall from a chair and one fall from walking using FuzzyEn based on Dataset C.

the home environment, which are presented as follows:

- True Positive (TP): an accelerometer data contains a fall, and it is correctly detected as a fall event.
- False Positive (FP): an accelerometer data does not contain a fall but is incorrectly detected as a fall.
- True Negative (TN): an accelerometer data does not contain falls and is correctly detected as non-fall.
- False Negative (FN): an accelerometer data contains a fall but is incorrectly detected as not a fall.

Table 5.7: Detection accuracy of FuzzyEn for Dataset C.

Events	Total	Detected	Not detected
Falls	30	30	0
Other activities	40	39	1

The results presented in Table 5.8 show the classification performance of the proposed fall detection algorithm on Dataset C. The proposed method achieves 97.8% specificity, which means that one of the normal daily activities has not been detected. However, the proposed method achieves 100% sensitivity, and this means that all falls are detected as a fall event. The accuracy of human fall detection is 98.6%. The proposed method for human fall detection shows high detection rates of 100%, which means that the false negative rate of fall detection is 0%. Based on the results achieved, FuzzyEn is a powerful measure to detect abnormality (here, falls) in behaviour when the sample data mostly represents normal activities. This also confirms that the FuzzyEn measure could be used to detect human falls.

5.3.4 Comparison of the Proposed Method with Existing Methods

Considering the literature review conducted for this research, the most commonly used methods for detecting human falls are SVM, RNN, and DNN. Therefore, to evaluate the proposed method carried out in this research, the results obtained by applying the FuzzyEn entropy measure are compared to other methods using the same Dataset C. The comparisons were made in terms of sensitivity and specificity, as shown in Table 5.9.

Considering the achieved results, the FuzzyEn measure is considerably better for human fall detection compared to other approaches. The FuzzyEn produces 100% sensitivity and 97.8% specificity. This also confirms that the FuzzyEn

Table 5.8: The classification performance of FuzzyEn using Dataset C.

Description	Obtained Result
Sensitivity	100%
Specificity	97.8%
False positive rate	2.2%
False negative rate	0%
Positive predictive value	97.2%
Negative predictive vale	100%
Accuracy	98.6%

Table 5.9: Comparison of the proposed method with other methods based on Dataset C.

Methods	Sensitivity (%)	Specificity (%)
Extended CORE9[191]	93.3	95
SVM [87]	100	96.6
DNN [48]	75	92.1
RNN [89]	100	96.67
FuzzyEn (Proposed method)	**100**	**97.8**

measure could be used to detect human falls during ADLs in a home environment.

5.4 Case Study 3: Anomaly Detection in Activities of Daily Living in the Presence of a Visitor

Anomaly detection aims to detect and identify any abnormal patterns in activities of daily living. Most of the current research in detecting an anomaly in ADLs focuses on a single-occupant environment where only one individual is monitored. The hypothesis that home environments are occupied by one resident all the time is not usually the case. It is common for the resident to receive visits from family members or health care workers. Visiting is considered as one of the most significant activities for older adults living alone at home [23]. Therefore, the resident's activity pattern is expected to be different when there is a visitor in the same environment (represented as a multi-occupancy environment), which can also be considered as an abnormal pattern in the resident's activities. The behaviour of a person could vary due to some personal factors such as visits and the influence of health conditions. Reliable anomaly detection in ADLs, or identifying visiting times (e.g. visits made by healthcare workers) is considered one of the most important components of many home health care applications [5]. Thus, existing methods are not able to reliably detect anomalous events in the resident's activities in the presence of a visitor and identify the time of visits, therefore generating a high false alarm rate.

The entropy measures mentioned in Chapter 3 Section 3.3 have been applied in two repetitions. In the first iteration, they are used to reveal days with abnormal behaviours, leading to the detection of days on which abnormality occurred. In the second iteration, they are utilised to detect days with anomalies as well as identify the potential causes of an anomaly by computing entropy measures. The distinction between normal and abnormal entropy values is achieved by finding the maximum entropy value on normal days, to be used as a threshold to detect any anomalies in ADLs. When the entropy values exceed the threshold, then this indicates an anomaly in ADLs. This means that by finding the maximum entropy value on normal days of ADLs, it is possible to detect abnormal behaviours in human ADLs in completely unseen data.

5.4.1 Methodology

This study proposes a novel entropy-based method to detect anomalies in ADLs in the presence of a visitor, solely based on information gathered from low-cost, non-intrusive ambient sensors, which include Passive Infra-Red sensors and a door entry sensor. Since the normal daily activity patterns of the resident are expected to be different when there is a visitor in the same environment or when there are conditions which affect normal behaviour, such as disrupted sleeping pattern. The aim is to collect the ADL data from ambient sensors to detect the anomalies in ADLs (here, identifying visiting times and irregular sleep). The challenge addressed in this paper is to avert the need to utilise a camera vision-based approach or wearable sensor to detect the anomalies in a resident's activities, and also to identify visiting times when there is a visitor.

The research hypothesis is that the level of changes in the occupant's activity patterns in a home environment is an indicator of normal or abnormal behaviours in ADLs. Therefore, the proposed entropy measures are based on finding the maximum entropy value in normal daily activities, which will be used as a threshold to detect abnormal behaviours in ADLs in completely unseen data. This means that any value that surpasses the computed maximum value for entropy on normal days will be indicated as an anomaly behaviour in the ADLs. Furthermore, the entropy measures are not only used to detect

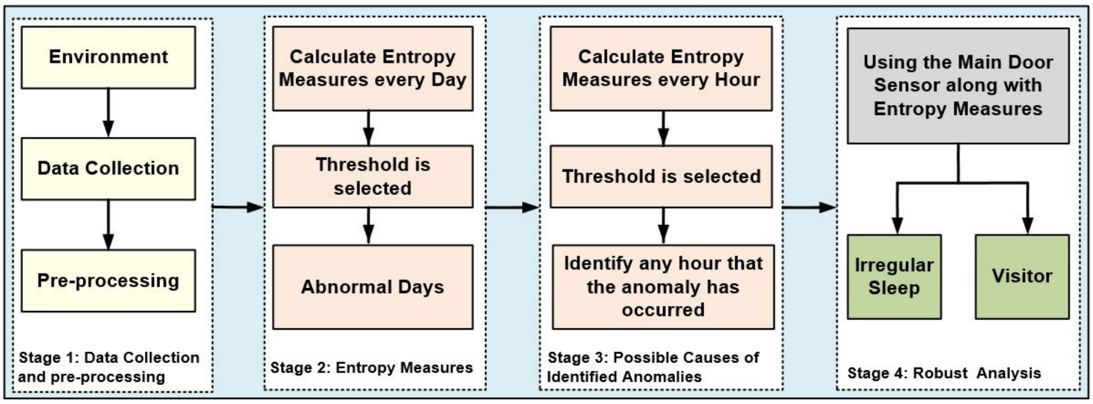


Figure 5.8: A schematic diagram of the proposed anomaly detection in activities of daily living in the presence of a visitor.

anomalies in ADLs, but also to identify the potential causes of anomalies. This is achieved by distinguishing whether the anomaly was the result of abnormal behaviour (e.g., sleeping disorder) or when there is a visitor to the same environment which naturally disturbs the normal activity.

A schematic diagram of the proposed entropy measures for anomaly detection in ADL in the presence of a visitor is illustrated in Figure 5.8, which consists of four processing stages.

- In the first stage, sensor data representing ADLs in a home environment is gathered based on PIR motion detectors and door entry sensors and then pre-processed. The required numerical features to be used for computing the input vector sequences of the entropy measures are extracted from the raw data. The values of this vector are then utilised as an input vector for entropy measures.

- In the second stage, the entropy measures are applied to the extracted vector sequence from the raw data and are calculated every day. Then, the threshold is selected as the maximum entropy value of normal days to be used for detecting any anomalous days.

- In the third stage, the entropy measures for each of the anomalous days are computed again every hour to examine the possible causes of the detected anomalous days and to identify any hour in which an anomaly has occurred.

- In the fourth stage, the main door entry sensor along with entropy measures, is used to distinguish between the irregular pattern in the resident's activity and visitors. The door entry sensor is also utilised to confirm the time of visits in a home environment and, in particular, for identifying exact visiting times.

5.4.2 Experimental Setup and Results

The dataset utilised for the validation of the proposed method is annotated Dataset E, explained earlier in Chapter 4 Section 4.3.1.5. The dataset includes different ADLs. For this work, only PIR sensors representing the resident in an area of the house and door sensor are selected and used. The relevant features that can distinguish between normal and anomalous cases in daily activities are selected, as explained in Chapter 4 Section 4.5. The selected features representing ADLs from the sensor data are the start time of entering each location (room), the time spent in each room, the transitions from one room to another inside the house, and the encoded daily activities sequence. The example provided in Chapter 4 Section 4.6, elaborated on the details about the process of how the required numerical features are obtained from the raw dataset. The input of any entropy measure should be formulated as a vector sequence (time series). Thus, to represent the dataset appropriate for entropy measures, the encoded dataset is converted to a set of data points equally spaced in time, which is dependent on the computational time of the entropy measures. The encoded daily activity sequence is then utilised as an input vector for entropy measures.

The entropy measures mentioned earlier are applied to the encoded data vector sequence to measure normal/abnormal patterns and detect anomalies in ADLs, and specifically in an irregular sleeping routine and identifying visiting times. The entropy measures are computed every day at 60 samples per hour (60×24) to identify anomalous days. This means that the vector sequence, A_N, consists of 1440 equally spaced samples. The vector sequence, A_N, is used as the input vector for the entropy measures to reveal days with abnormal behaviours, leading to the detection of days on which an abnormality occurred. To compute the ApEn, SampEn, FuzzyEn, and MFE, the parameters of embedded dimension, m, and tolerance, r, are required to be defined. Thus, the algorithm for these entropy measures is impacted by the selection of these values. The best results were obtained when the values of the parameters m, and r are 2, and 1 respectively. Whereas the values of the parameters m and time delay τ, which are required to compute PerEn and MPE are set as 2 and 1, respectively. After the entropy measures have been computed, the threshold is selected as the maximum entropy value of normal days to be used for detecting anomalous days. When the entropy value of each day goes beyond the calculated maximum value for entropy on normal days, it is treated as anomalous days in the resident's activity.

It is possible to compute entropy measures at different time scales (e.g., 15, 30, 60, or 120 minutes) to identify anomalous days in the resident's activity. However, Our earlier work has concluded that when the calculation period of entropy measures is less than one hour, it is not sufficiently reliable enough to detect anomalies in ADLs. This can be justified by the fact that decreasing the computational period of entropy measures will reduce the number of observations per time period, which will, in turn, increase the variance. Consequently, the number of false positives will increase, which reduces precision. Therefore, the best performance is obtained when the computational time of entropy measures is based on a one-hour time period.

The proposed method is based on the assumption that when the entropy value of each day exceeds the threshold value, then this indicates that there is an abnormality in the resident's activity on these days. Thereby, the proposed method can detect anomalous behaviour in unseen human abnormality data, which means the proposed method is capable of adapting to detect abnormal behaviour in ADLs in completely unseen data. Figure 5.9 and Figure 5.10 show the results obtained from applying the ShEn and FuzzyEn measures for identifying any anomaly in ADLs in the presence of a visitor, respectively. The results in Figure 5.9 shows that the proposed ShEn method identifies only 7 days (days 16, 29, 33, 38, 49, 52, and 63) as anomalous days in the resident's activity out of 9 anomalous days, and failed to detect 2 of the anomalous days (days 25 and 42). This can be justified by the fact that ShEn is strongly dependent on the length of the time series and the need to discriminate the

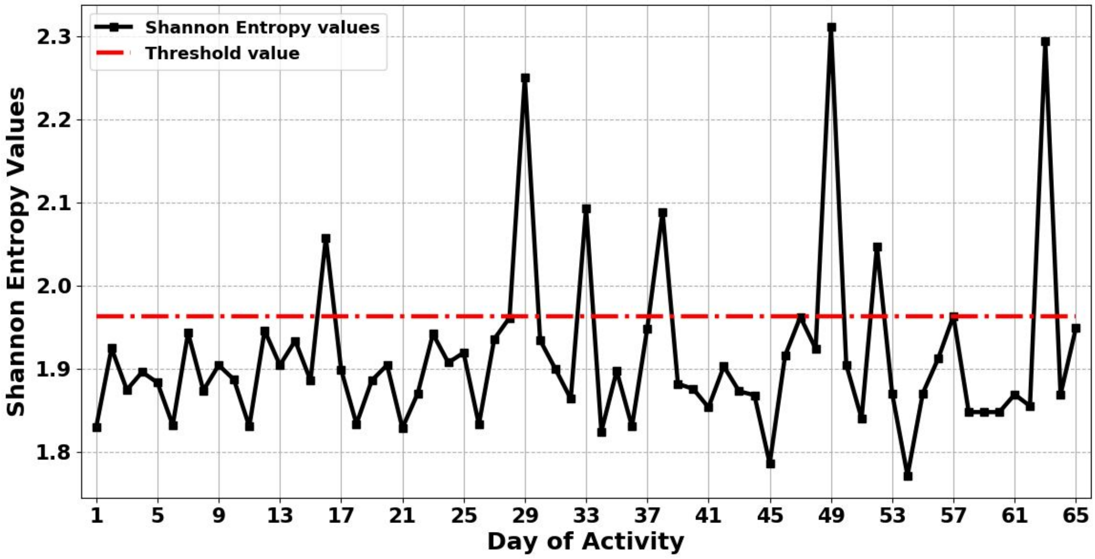


Figure 5.9: The results obtained by applying Shannon Entropy (ShEn) for anomaly detection in the activities of daily living in the presence of a visitor. The figure also illustrates the threshold value for 65 days.

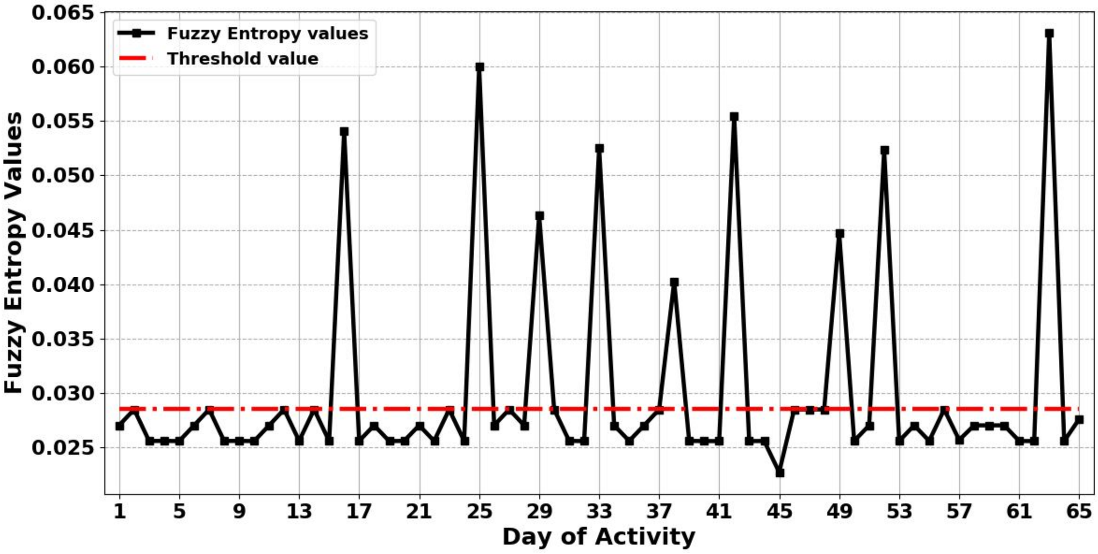


Figure 5.10: The results obtained by applying Fuzzy Entropy (FuzzyEn) for anomaly detection in the activities of daily living in the presence of a visitor. The figure also illustrates the threshold value for 65 days.

nature of the generating systems [159]. However, from Figure 5.10, it can be seen that the proposed FuzzyEn method detects 9 days as anomalous days

(days 16, 25, 29, 33, 38, 42, 49, 52, and 63) as the FuzzyEn values of these days overrode the threshold. This means that the proposed FuzzyEn method successfully identified all anomalous days in the resident's activity, based on Dataset E.

After detecting 9 anomalous days in the resident's activity, the entropy measures for each of these days are calculated again every hour at 60 samples per hour to examine the possible causes of the detected anomalous days and identify any hour that the anomaly had occurred. This means that the input vector sequence to entropy measures, A_N, consists of a 60 equally spaced samples. Besides, the threshold is selected as the maximum entropy value of normal days, which is also calculated again every hour, to be used for detecting any hour the anomaly has occurred on anomalous days. Therefore, when the entropy values of each hour on a given day goes beyond this threshold value, this then indicates an anomaly in ADLs at that hour. This means that by finding the maximum entropy value in normal daily activities, it is possible to detect abnormal behaviour in ADLs in completely unseen data.

The results in Figure 5.11 and Figure 5.12 indicate the hour that the anomaly has occurred in the detected anomalous days by applying ShEn and FuzzyEn for Dataset E based on one-hour time periods. The threshold value for this experiment is chosen by calculating the maximum entropy value on normal days based on one-hour time periods. To detect the hour that the anomaly has occurred in the detected anomalous days, ShEn and FuzzyEn values for each hour in Figure 5.11 and Figure 5.12 were compared with the threshold value to determine when the entropy value has passed the threshold value. From Figure 5.11, it can be observed that the proposed ShEn method identified only one hour in each day of 5 detected anomalous days out of 13 hours in 9 identified anomalous days, and failed to detect any hours the anomaly had occurred on 4 anomalous days, which are days 25, 29, 38, and 49. Nevertheless, all anomaly hours in ADLs are correctly detected in all identified anomalous days by applying the FuzzyEn method to the ADL dataset because the FuzzyEn values for these days exceed the threshold, as shown in Figure 5.12.

To identify potential causes of the hours that the anomaly has occurred in the detected anomalous days, the main entry door sensor is used to distinguish

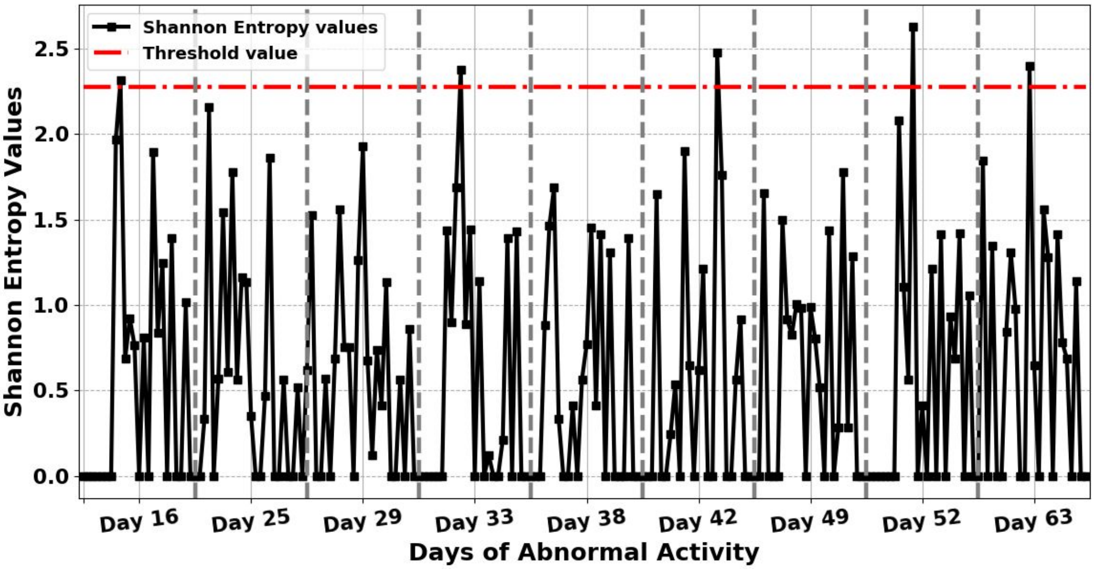


Figure 5.11: The results obtained when applying Shannon Entropy (ShEn) for 9 days of abnormal activity to examine the possible causes of the identified anomalous days based on one-hour time periods. The figure also shows the threshold value for entropy on normal days, which will be used for detecting any hour the anomaly has occurred on anomalous days.

between the entropy changes caused by irregular sleep in the resident's activity and a visitor. The finer-grained analysis provided in Section 5.4.4 will elaborate on the details of identifying potential causes of the hours that the anomaly has occurred in the detected anomalous days.

Table 5.10: Detection accuracy of ShEn, ApEn, SampEn, PerEn, MPE, FuzzyEn, and MFE for Dataset E.

		ShEn	ApEn	SampEn	PerEn	MPE	FuzzyEn	MFE
Events	Total Samples	Detected						
Normal	203	203	203	203	202	202	203	203
Abnormal	13	5	10	8	13	13	13	13
Events	Total Samples	Not Detected						
Normal	203	0	0	0	1	1	0	0
Abnormal	13	8	3	5	0	0	0	0

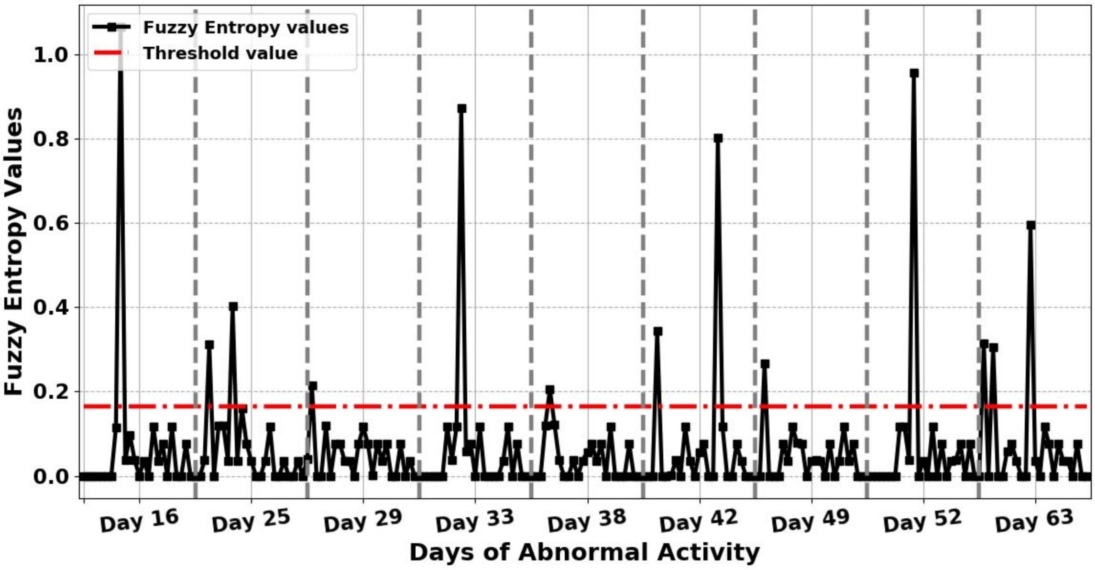


Figure 5.12: The results obtained from applying Fuzzy Entropy (FuzzyEn) for 9 days of abnormal activity to examine the possible causes of the identified anomalous days based on one-hour time periods. The figure also shows the selected threshold value for entropy on normal days, which will be used for detecting any hour the anomaly has occurred on anomalous days.

5.4.3 Performance Evaluation

The proposed method is based on the hypothesis that the values of entropy measures are higher than a threshold value when there are anomalies in a resident's activity. Therefore, to evaluate the performance of the proposed ShEn, ApEn, SampEn, PerEn, MPE, FuzzyEn, and MFE measures, first, annotated Dataset E representing the ADLs of a single user are manually classified as normal or as abnormal in the resident's activity based on periods of one hour. As can be seen from Table 5.10, there are 203 events indicated as normal activities of the resident, and 13 events are fixed as indicating abnormalities in the resident's activity. The first row indicates that the proposed ShEn, ApEn, SampEn, FuzzyEn, and MFE measures successfully detected all normal activity included in Dataset E. However, both the proposed PerEn and MPE measures identified 202 events as being normal activity out of 203 events and miss-classified only one event. The second row demonstrates that the proposed PerEn, MPE, FuzzyEn, and MFE measures successfully identified all anomalous events, while the proposed ShEn, ApEn, and SampEn measures detected only 5, 10, and 8 anomaly events out of a total of 13 anomaly events, respectively. Based on the results shown in Table 5.10, it can be argued that the proposed FuzzyEn and MFE measure correctly identified all normal and anomalous events in ADLs and outperformed other entropy measures.

The results presented in Table 5.11 represents the performance of the proposed entropy measures for anomaly detection in ADLs in the presence of a visitor when they are computed over a one-hour time period. The results based on specificity indicate that all the proposed entropy measures achieve a high specificity of 100%, which means that all normal daily activities are correctly detected as normal in a resident's activity. In contrast, the results related to sensitivity show that the proposed PerEn, MPE, FuzzyEn, and MFE measures perform better than the ShEn, ApEn, and SampEn measures, since they indicate a perfect sensitivity of 100%, which means that all anomalous events in ADLs have been correctly identified. Besides, the proposed ShEn, ApEn, and SampEn measures achieve a detection rate of 38.4.5%, 69.2%, and 61.5%, respectively, which means that they have a 61.6%, 30.8%, and 38.5% false-negative rate for identifying anomalies in ADLs, respectively. However, the proposed PerEn, MPE, FuzzyEn, and MFE measures show high detection rates of 100%, which means that the false-negative rate of anomaly detection in ADLs is 0%.

Based on the results achieved, the proposed PerEn, MPE, FuzzyEn and MFE measures are better indices than the proposed ShEn, ApEn, and SampEn

Table 5.11: The performance results of the proposed entropy measures for Dataset E when the computational time is based on one-hour time periods.

Entropy Measures	Sensitivity	Specificity	FPR	FNR	Accuracy
ShEn	38.4%	100%	0%	61.6%	96.2%
ApEn	69.2%	100%	0%	30.8%	98.1%
SampEn	61.5%	100%	0%	38.5%	97.6%
PerEn	100%	99.5%	0.5%	0%	99.5%
MPE	100%	99.5%	0.5%	0%	99.5%
FuzzyEn	**100%**	**100%**	**0%**	**0%**	**100%**
MFE	**100%**	**100%**	**0%**	**0%**	**100%**

measures to detect anomalies (here, detecting visitor and irregular sleep) in behaviour when the sample data mostly represents normal activities. This also confirms that the proposed entropy measure could be used for anomaly detection in ADLs in the presence of a visitor.

5.4.4 Robust Analysis

The entropy measures are used not only to detect anomalies in ADLs, but also to identify potential causes of anomalies by calculating entropy measures on an hourly basis and then by distinguishing between irregular sleep in the resident's activity and visitors. The distinction between irregular sleep and visitors was achieved by using the main door sensor along with entropy measures. Moreover, The door entry sensor is used along with entropy measures to confirm the present time of the visitor in the home environment. As the visitors enter and exit the home through the main door, the door sensor is used to confirm the time of visits. This will increase the performance evaluation of the proposed entropy measures. In general, door opening or closing does not necessarily mean that a visitor is present in the home environment, as the door might be opened by the main occupant; e.g., in response to a postman or a neighbour. Therefore, the presence of visitors cannot be identified only by utilising the main door sensor. Thus, entropy measures are utilised to detect anomalies in ADLs in the presence of the visitor, and then the door sensor is utilised to confirm the time of the visit.

Figure 5.13 shows the distinction between irregular sleep in the resident's activity and the visitor using a door entry sensor with entropy measures. As the best results are obtained when the computational time of entropy measures is performed based on one-hour intervals, the door entry sensor is used to confirm the visiting time on each day. As can be seen in Figure 5.13(a), the door was opened six times on day 16, but the visitor came once on that day, at 09:03 am, and stayed in the house until 09:50 am. This means that the main resident might have caused the other door events. On this day, it can be confirmed from the door sensor data that the type of anomalies included only the visitor because the entropy value on this day exceeds the threshold value when the door is opened. Whilst Figure 5.13(b) shows that the door was opened four times on day 49, but

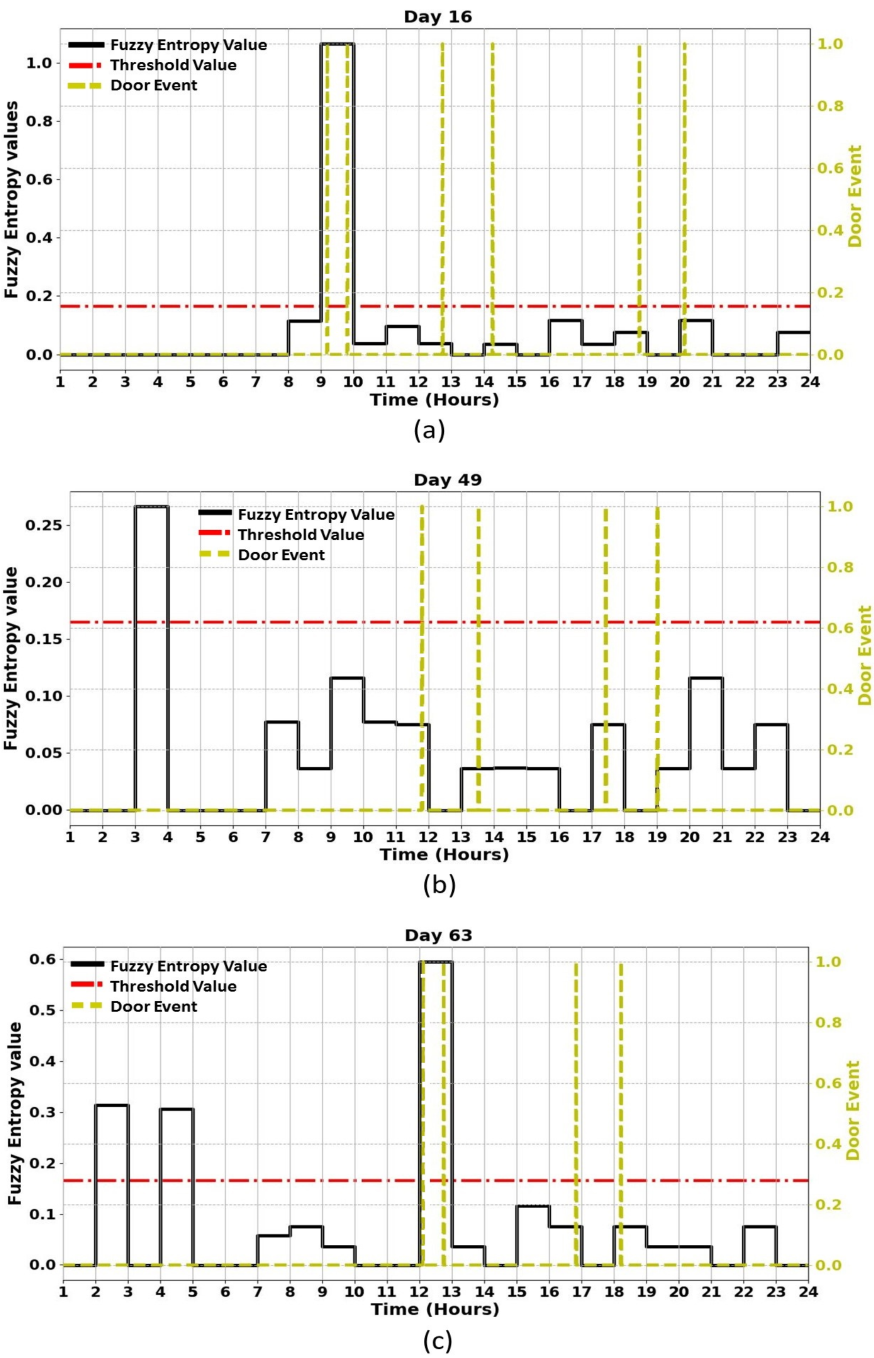


Figure 5.13: Examples of an identified visitor and irregular sleep using a door sensor with entropy measures for Dataset E representing: a) visiting time on day 16, with the time confirmed using the door sensor; b) irregular sleep on day 49; and, c) visitor and irregular sleep on day 63.

the entropy values were higher than the threshold value when no door is opened or closed. This means that the resident has an irregular sleeping pattern, and this is confirmed by the time (03:00 am). Meanwhile, Figure 5.13(c) shows that the door was opened four times on day 63, but the entropy values were higher than the threshold value in three different positions. On this day, it can be seen that the entropy values were higher than the threshold value when no door is opened or closed. This means that the resident has an irregular sleeping pattern at 02:00 am and 04:00 am compared to the usual days because it cannot be confirmed from the door sensor data. However, it can be confirmed from the door sensor data that the resident had a visitor, and the visitor came 12:08 pm and stayed in the home until 12:46 pm. This means that on this day (day 63), there was irregular sleep in resident's activity, and the occupant had a visitor on this day. The identified anomalous days and possible causes of these for the ADL dataset are summarised in Table 5.12.

In summary, the entropy measures are useful and relevant tools to detect abnormality (here, irregular sleep and a visitor) in behaviour when the sample data mostly represents normal activities. This also confirms the possibility that the entropy measures are used to distinguish between different causes of anomalies when they are used in conjunction with data gathered from a secondary sensor.

Table 5.12: A summary of identified anomalies days and possible causes of these for Dataset E.

Day	Cause	Detailed description
Day 16, 33, and 52	Visitor	The resident receives visits on these days, which might be from family members or health care workers.
Day 25, 42 and 63	Irregular sleep and Visitor	The resident has an irregular sleeping pattern and also receives a visitor on these days, and this is confirmed by using the main door sensor.
Day 29, 38, and 49	Irregular sleep	The resident has an irregular sleeping pattern compared to the usual days.

5.5 Discussion

In this chapter, different entropy measures are applied to investigate the effectiveness of these methods in identifying and detecting various types of anomalies in ADLs. The goal is to investigate whether entropy measures can be used to detect and distinguish anomalies in ADLs, specifically, in sleeping, human falls, and in identifying visiting times. Due to the dynamic nature of human behaviour, there are uncertainties associated with identifying and detecting their anomalous activities in a single-occupancy or multi-occupancy environment. This work focused on proposing different entropy measures capable of detecting and identifying different anomalies in daily activities in a single-occupancy or multi-occupancy environment, specifically in sleeping routine, human falls, and anomalies in ADLs in the presence of a visitor.

The research assumption is that the level of changes in a resident's activity patterns in a home environment is an indicator of normal or abnormal activities. The threshold, based on the standard deviation of the occupancy data in conjunction with several entropy measures, is applied to identify and detect whether there is an anomaly in the resident's activity or not. Hence, when the value of entropy measures exceeds the threshold value, then the case is indicated as an anomaly in ADLs. Real home environments, including three case studies, are used to show the effectiveness of proposed entropy measures for anomaly detection in ADLs. The results show that the performance of the proposed entropy measures is better than the other approaches. This also confirms that the proposed entropy measures are a promising technique to distinguish between normal and anomalous events in a resident's activity in the home environment.

The direction of research following this chapter is to investigate the effectiveness of different entropy measures in identifying visitors in a multi-occupancy home environment, solely based on the information gathered from motion detectors and door entry sensors. This is presented in the following chapter.

Chapter 6

Visitor Detection in Multi-Occupancy Environments

6.1 Introduction

Most research works related to recognising ADLs have focused only on single occupancy environments, wherein, it is assumed that only one person (i.e. the prominent resident) is present in the home [16, 27, 28]. Nevertheless, the real home environments are likely to be occupied by more than one person [18, 29, 30, 31]. For example, it is likely that older adults will receive visits from family members or healthcare workers (referred to as a multi-occupancy environment). Visiting is considered as one of the most important activities for older adults living alone at home [23], which makes multi-occupancy scenarios are far more realistic [13, 14]. Therefore, it is essential to identify human activities in the presence of visitors without the visitors putting on any specialised devices to distinguish their activities.

Many current research works acknowledge the challenges of multi-occupancy in HAR [13, 32, 33]. Such challenges are, finding suitable models to represent the data association problem (i.e., the detection of a visitor) and finding an activity recognition system that captures different interactions among residents [14, 34]. Previous studies report that detecting and identifying a visitor in a home environment using only binary sensors is a primary challenge, as binary sensors are not able to provide any information about the personal identity of who triggered the sensor [18, 35]. Some previous studies have used wearable sensors to overcome the problem of detecting and identifying multi-occupancy in a home environment [29, 119].

Researchers within this area of study have focused on diverse challenges, although the activity recognition in a multi-occupancy smart home environment is still considered the primary challenge [18]. While most of the activities can be appropriately recognised when there is only one occupant in the home, the activities of multiple occupants living together in the same environment cannot be easily separated and recognised, since ambient sensors are not able to provide any information about the personal identity of who triggered the sensor. Therefore, to accurately recognise human activities within multi-occupancy environments, a method is required to distinguish the ADL when the data represents multi-occupancy in the same environment.

Identifying visitors and the time of their visits (such as healthcare visitors) are essential for healthcare management [23]. It is crucial to develop a system with the ability to identify the exact time of a visit without the need for visitors to be asked to carry a tag or wearable device to identify them. The challenge of this study is to avoid using human tracking devices or any tagging sensors. To overcome the challenge of detecting and identifying multi-occupancy in a home environment, an unsupervised method is proposed in this study, using entropy measures to investigate their effectiveness in identifying visitors (visiting time). The research aims to investigate whether entropy measures introduced in the earlier chapters (in Chapter 3 Section 3.3) can be used to identify multi-occupancy in a home environment. Furthermore, the research investigates the impact of changing the values of an embedded dimension, m, and tolerance, r, as parameters required to calculate some named entropy measures.

The remaining sections in this chapter are structured as follows. In Section 6.2, the proposed method for identifying visitors (time of visit) in a home environment based on different entropy measures is presented. Section 6.3 presents the experimental results and results. Section 6.4 presents the impact of changing the values of parameters m and r required to calculate some entropy measures, followed by robust analysis in Section 6.5. In Section 6.6, the performance of the proposed visitor detection is compared with existing modelling techniques and conclusions of the work are drawn in Section 6.7.

6.2 Methodology

It can be argued that the ADLs of a single user in a home environment are different from the ADLs representing multi-users in the same environment. The pattern of activities when a visitor comes to visit an individual (represented as a multi-occupancy environment) will be different from when only the primary occupant is in that environment. When the environment is occupied by one person, it is possible to recognise different activities and develop a method representing the normal activities. Once a newly perceived activity differs from the routine of a specific person, it will be represented as an abnormality in the behaviour. However, when there is more than one person in the same environment, the activities of the primary occupant cannot be easily distinguished from simultaneous activities. This research seeks to identify the activities of the primary occupant without introducing any new hardware (or monitoring devices) to the environment or using tagging systems (such as pendant or wristband with RFID).

Standard statistical measures such as activity count and sensors activation can be used as a measure of multiple occupants. However, they are incapable of distinguishing the level of activities and visitors. It can also be argued that when different types of sensors such as pressure sensor on beds, or sofa or door entry sensors are used, then the data collected are not comparable and the activity count is meaningless. Therefore, techniques such as entropy measures show potential in terms of indicating changes and/or disorders in a resident's activity pattern in a home environment. Entropy can be utilised as a measure of disorder or irregularity in data since the level of disorder in a multi-occupancy environment is expected to be higher than the single-occupancy case. Therefore, entropy measures can be used to identify and detect a visitor in a home environment with a single occupant.

The proposed approach is based on the hypothesis that the presence of a visitor can be detected when the entropy value is greater than a nominal value.

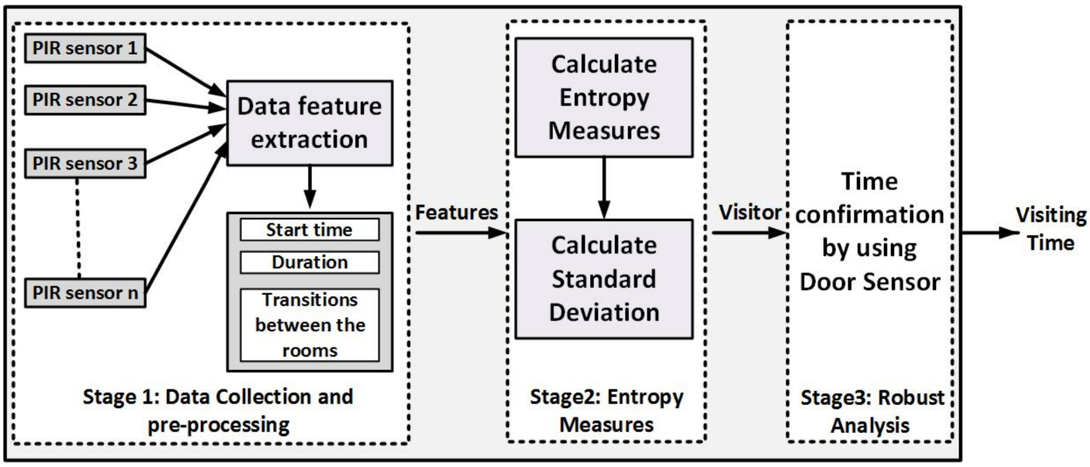


Figure 6.1: A schematic diagram of the proposed visit detection (time of visit) framework.

A large value of entropy does not exclusively signify the presence of a visitor in a home environment. For example, a large value of entropy may be influenced by other factors, such as house-cleaning duties, which are different from having a visitor. Having a visitor is considered a deviation in the normal pattern of daily activities for a person living alone.

A schematic diagram of the proposed framework in this work for detecting a visitor is illustrated in Figure 6.1. There are three distinct phases to identify the multi-occupancy.

- In the first phase, the sensor data representing ADLs in a multi-occupancy environment is gathered based on PIR motion detectors. This work primarily concentrates on the motion data representing the occupancy of different areas in a home environment. Data gathered from other sensors, including door entry sensors, can also be used. The required numerical features to be used for calculating the sequences of the input vector are extracted from the raw data. The values of this vector are used as inputs to the entropy measures. The selected features representing the ADLs from the sensor data are the start time of entering each location (room), the time spent in each room, the transitions from one room to another inside the house, and the encoded daily activities sequence as explained in Section 4.5. Provided examples in Section 4.6 have elaborated on the details of these features.

- In the second phase of the proposed process, different entropy measures are applied to the extracted vector sequences from the raw data to detect the presence of a visitor in a single home environment, represented as an abnormality in the extracted activity patterns. Then, the standard deviation of entropy measures is calculated and used to detect and identify whether there is a visitor in the home environment.

- In the third phase, the opening and closing of the main door to the home environment is used to confirm the time of the visit.

6.3 Experimental Setup and Results

To evaluate the performance of the entropy measures, two experiments are conducted using two different datasets - Dataset D and Dataset F - as described in Chapter 4 Section 4.3.1.4 and Section 4.3.2.1, respectively. These datasets comprise information regarding the ADLs performed by two users daily in their own homes, solely based on information gathered from low-cost, non-intrusive ambient sensors. The challenge of this study is to avoid using human tracking devices or any tagging sensors that can be used for visitors detection.

6.3.1 Experiment and Results with Dataset D

In this experiment, ShEn, ApEn, SampEn, PerEn, MPE, and FuzzyEn measures are applied to the Dataset D. These entropy measures are applied to the generated vector sequence from the data to measure the normal/abnormal patterns and the degree of variance between the measurements in consecutive days, to detect the multi-occupancy patterns. A Comparison of ShEn, PerEn, and MPE measures based on the activity of daily living for Dataset D is shown in Figure 6.2.

Figure 6.3 illustrates the results obtained by applying FuzzyEn on Dataset D based on one-hour time periods, as well as the FuzzyEn values for each day and the mean value of FuzzyEn for seven days. The mean of FuzzyEn for seven

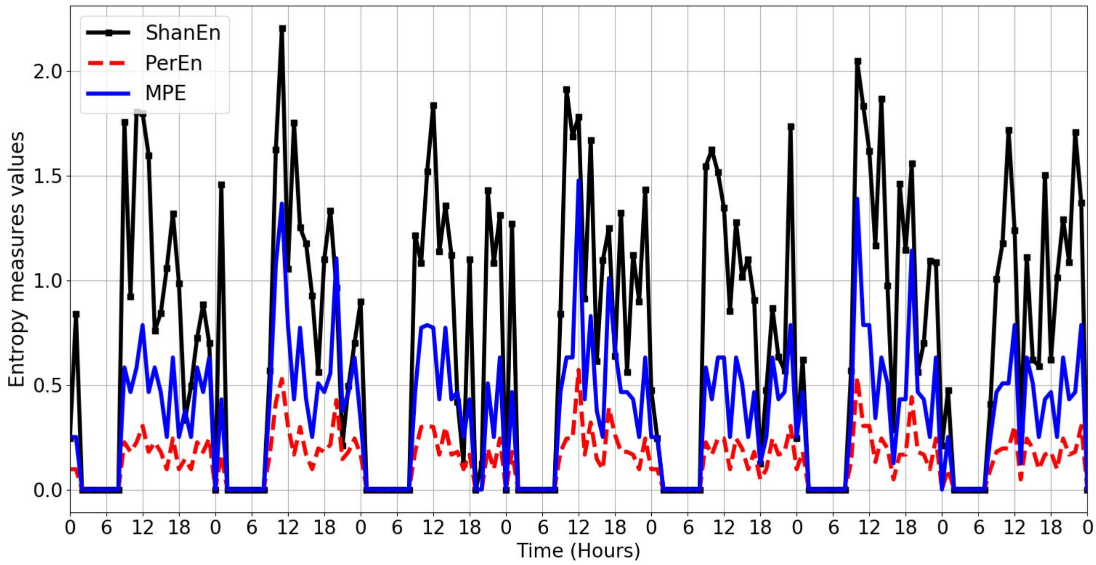


Figure 6.2: Comparison of Shannon Entropy (ShEn), Permutation Entropy (PerEn), and Multiscale-Permutation Entropy (MPE) measures based on one-hour time periods using Dataset D. The figure also shows that ShEn, PerEn, and MPE present similar patterns.

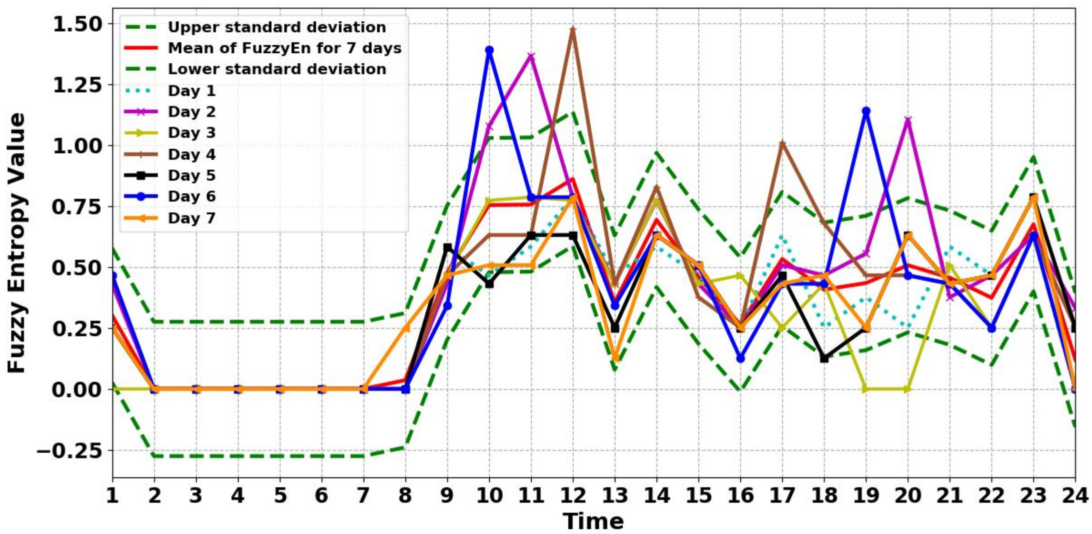


Figure 6.3: The results obtained by applying fuzzy entropy for seven days based on Dataset D to identify visiting time based on one-hour time periods. The figure also shows the mean value of fuzzy entropy for seven days and standard deviation boundaries.

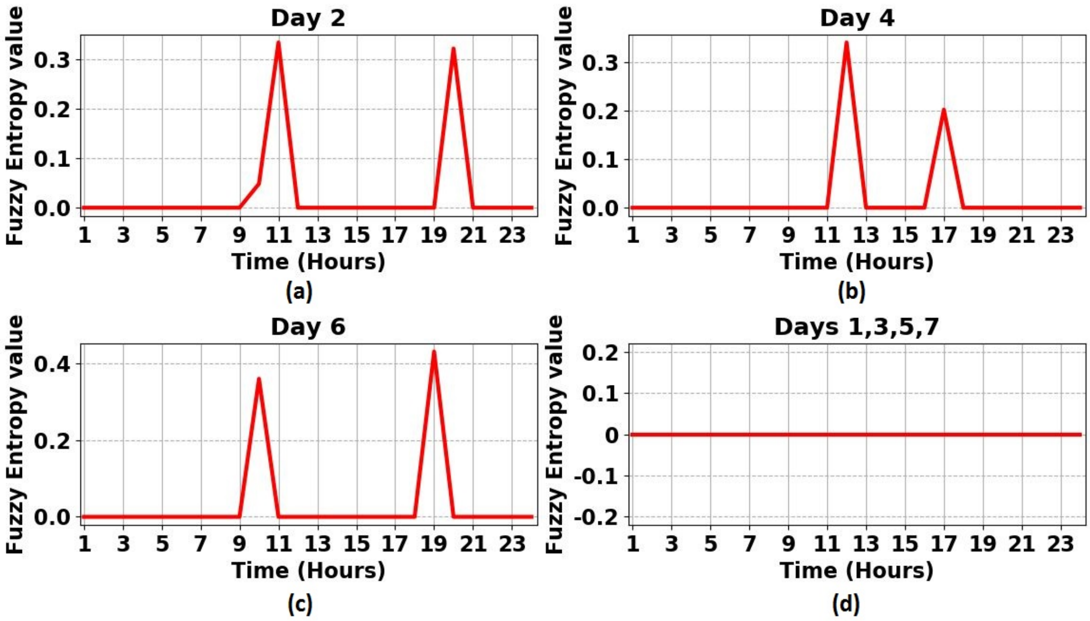


Figure 6.4: Fuzzy entropy values representing visiting time in each day compared with the standard deviation using Dataset D; (**a**), (**b**) and (**c**) show that the visitor came twice in day 2, 4, and 6 because there are two bumps in the fuzzy entropy values; and (**d**) shows that no visitor came in day 1, 3, 5 and 7 because the fuzzy entropy values are zero (no bumps in the entropy values).

days and threshold based on standard deviation are calculated for a period of 24 hours.

The threshold value is chosen based on the standard deviation, $\sigma = 1$. It is possible to use thresholds based on different values for σ (e.g., 2 or 3) to detect and identify whether there is a visitor in the home environment. It was observed that when the threshold is increased the number of observations per time period will be reduced, increasing the calculated variance. Consequently, the number of false positives (detected as visits) increases, reducing the calculated precision.

The threshold value is chosen based on the standard deviation that varies over time and not as a constant value. This threshold value was 0.074 changing over time. The standard deviation is depicted by the dotted line in Figure 6.3. To detect the visitor in a home environment, FuzzyEn values for each day in Figure 6.3 were compared with the upper standard deviation boundaries to see which days go beyond the upper boundary of standard deviation. Figure 6.4 shows the visiting time in each day based on entropy values using Dataset D after they are compared with the upper boundary of the standard deviation in Figure 6.3. As can be seen in Figure 6.4(a), Figure 6.4(b) and Figure 6.4(c), there are two bumps in the entropy values, which indicate that the visitor came at those times. This also confirms that the visitor came twice in day 2, 4, and 6. However, Figure 6.4(d) shows that no visitor came in day 1, 3, 5, and 7 because there are no bumps in the entropy values. In summary, it can be confirmed that the visitor came twice a day, three days a week, which means that the visitor was identified accurately in all instances.

The classification performance of ShEn, ApEn, SampEn, PerEn, MPE, and FuzzyEn measures are evaluated by a confusion matrix that includes accuracy, recall, and precision. There are four possible results for testing the detection of a visitor in the home environment, which are presented as follows:

- True Positive (TP) is a set of data that contains a visitor event and was correctly classified as a visitor event.
- False Positive (FP) is a set of data that does not contain a visitor event, but it was incorrectly classified as a visitor event.
- True Negative (TN) is a set of data that does not contain a visitor event, and it was correctly classified as a non-visitor event.
- False Negative (FN) is a set of data that contains a visitor event, and it was incorrectly classified as a non-visitor event.

The accuracy, precision, and recall are computed for each entropy measure. The accuracy is defined as the percentage of correctly identified events (visitor and non-visitor). Precision indicates the percentage of the positive visitor events that are correctly identified, while recall indicates the percentage of true activity labels which were correctly identified. The accuracy of the entropy measures would be high even if the visitor was not well identified. However, the recall and precision would be low. In this case, to show the classification performance of the named entropy, the precision and recall are chosen as the best choice rather than accuracy to demonstrate the entropy measure's performance.

Table 6.1: The classification performance of ShEn, ApEn, SampEn, PerEn, MPE, and FuzzyEn using Dataset D when they are calculated at 120-, 60-, and 15-min time period.

Entropy Measure	Calculation Period of 120 Min			Calculation Period of 60 Min			Calculation Period of 15 Min		
	Accu.	Prec.	Rec.	Accu.	Prec.	Rec.	Accu.	Prec.	Rec.
ShEn	79	18	21	85	20	25	68	15	20
ApEn	96.5	100	66.6	**100**	**100**	**100**	85.3	35	69.2
SampEn	86.6	50	66.6	96.4	75	75	86	29.3	66.6
PerEn	93.2	79.6	65.4	98.8	87.5	87.5	84.7	34.3	68.5
MPE	95.2	93.1	64	**99.1**	**95**	**87.5**	86.4	36.7	68.9
FuzzyEn	96.5	100	66.6	**100**	**100**	**100**	87.5	38.2	73.6

Table 6.1 represents the classification performance of ApEn, SampEn, and FuzzyEn using Dataset D when they are computed at 120, 60, and 15 mins time period. When the period of calculation for the entropy measures was two hours, the precision results show that the proposed ApEn and FuzzyEn perform much better than SampEn; while the results related to the recall demonstrates that the FuzzyEn performs much better than ApEn and SampEn with a difference of 28% and 14% respectively. However, it is noted that the best performance is obtained when the computational time is performed based on one-hour time periods. The precision results indicate that the ApEn and FuzzyEn are outperformed by SampEn by approximately 38%; whereas the results related to the recall show that ApEn achieves a very low performance compared to FuzzyEn and SampEn by approximately 12.5%. On the other hand, the results show that all the entropy measures achieved very low performance when they are calculated at 15 min time periods. The results related to precision illustrate that all the entropy measures show a very low performance, which means that the number of false positives increased. To explain what led to these results, it was observed that when the computational time of entropy measures is decreased, the number of observations per time period will be reduced, increasing the calculated variance. Consequently, the number of false positives (detected as visits) increases, reducing the calculated precision.

6.3.2 Experiment and Results with Dataset F

The aim of this experiment is to determine whether entropy measures, including ApEn, SampEn, and FuzzyEn, can be used to identify multi-occupancy in a home environment using Dataset F. To perform the experiment, the dataset is transformed into notional values as described in Chapter 4 Section 4.6. The ApEn, SampEn, and FuzzyEn are computed at time intervals for a set of data with different patterns of ADLs. The computational period of entropy measures is divided into time slices of lengths 120, 60, and 15 min. The reason for limiting these time slices is that the period of the visits (by the carer) is one hour or less. The values of the parameters, m and r, which are needed for entropy calculations, are 2 and 1 respectively.

ApEn, SampEn, and FuzzyEn entropy measures present similar patterns. In addition to entropy measures, the average values of daily pattern and threshold based on the standard deviation of the occupancy data is used in conjunction with the entropy measures for a period of 24 h to decide whether or not there is a visitor in the home environment. For example, the threshold value for this experiment is chosen based on the standard deviation that varies over time, and it is not a constant value (it was 0.04). Therefore, when the entropy value of each day goes beyond this value, it means the event is detected as a visitor in the home environment.

Figure 6.5 illustrates the results obtained by applying FuzzyEn on Dataset F based on one-hour time periods, as well as the FuzzyEn values for each day and the mean value of FuzzyEn for four days. To detect and identify the visiting time of the visitor in a home environment, the FuzzyEn values for each day in Figure 6.5 were compared with the upper standard deviation boundaries to see which days exceed the upper boundary of standard deviation. Figure 6.6 shows the visiting time in each day based on fuzzy entropy values for Dataset E after they are compared with the upper boundary of the standard deviation in Figure 6.5. It is depicted in Figure 6.6(a) and 6.6(b) that no visitor came on day 1 and 2, as there are no bumps in the entropy values. In contrast, Figure 6.6(c) and 6.6(d), there is a bump in the FuzzyEn values for day 3, and there are three bumps in FuzzyEn values for day 4. This also confirms that the visitor came once, on day

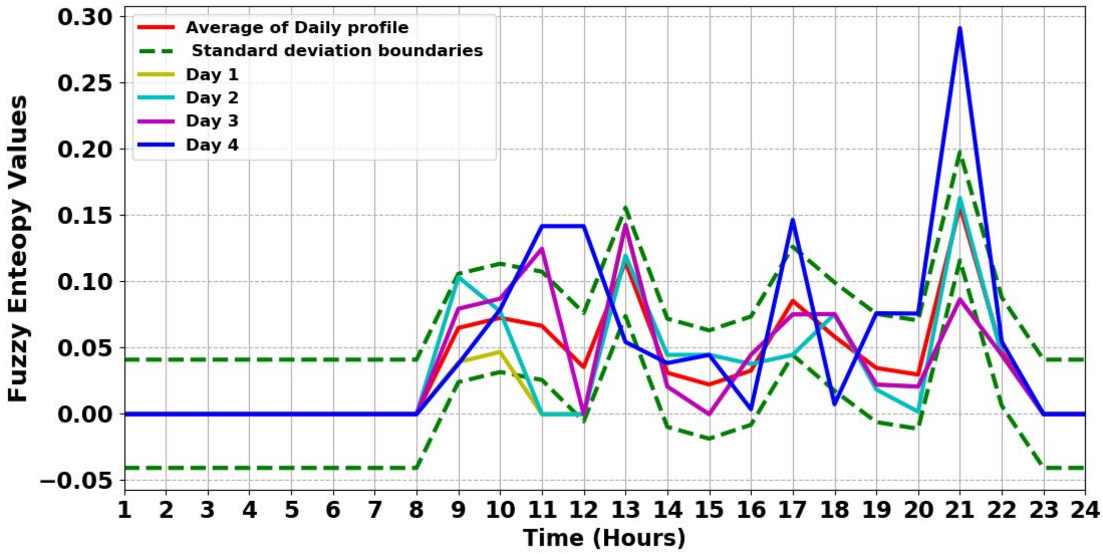


Figure 6.5: The result obtained when FuzzyEn is applied on the Dataset F using one hour as calculation time. The figure also shows the mean value of fuzzy entropy for four days and standard deviation boundaries.

Table 6.2: The classification performance of ApEn, SampEn, and FuzzyEn using Dataset F when they are calculated at 120-, 60-, and 15-min time period.

Entropy Measure	Calculation Period of 120 Min			Calculation Period of 60 Min			Calculation Period of 15 Min		
	Accu.	Prec.	Rec.	Accu.	Prec.	Rec.	Accu.	Prec.	Rec.
ApEn	96.5	100	66.6	**100**	**100**	**100**	85.3	35	69.2
SampEn	86.6	50	66.6	96.4	75	75	86	29.3	66.6
FuzzyEn	96.5	100	66.6	**100**	**100**	**100**	87.5	38.2	73.6

3, and three times on day 4. In conclusion, it can be confirmed that the visitor came once, and three times on day 3 and day 4, respectively, which means that the visitor was accurately detected in all cases.

Table 6.2 represents the classification performance of ApEn, SampEn, and FuzzyEn using Dataset F when calculated at 120, 60, and 15 minutes time periods. The accuracy, precision, and recall results show that ApEn and FuzzyEn perform much better than SampEn. It should also be noted that the best performance is obtained when the computational time is performed based on a one-hour time period. Furthermore, the results demonstrate that ApEn

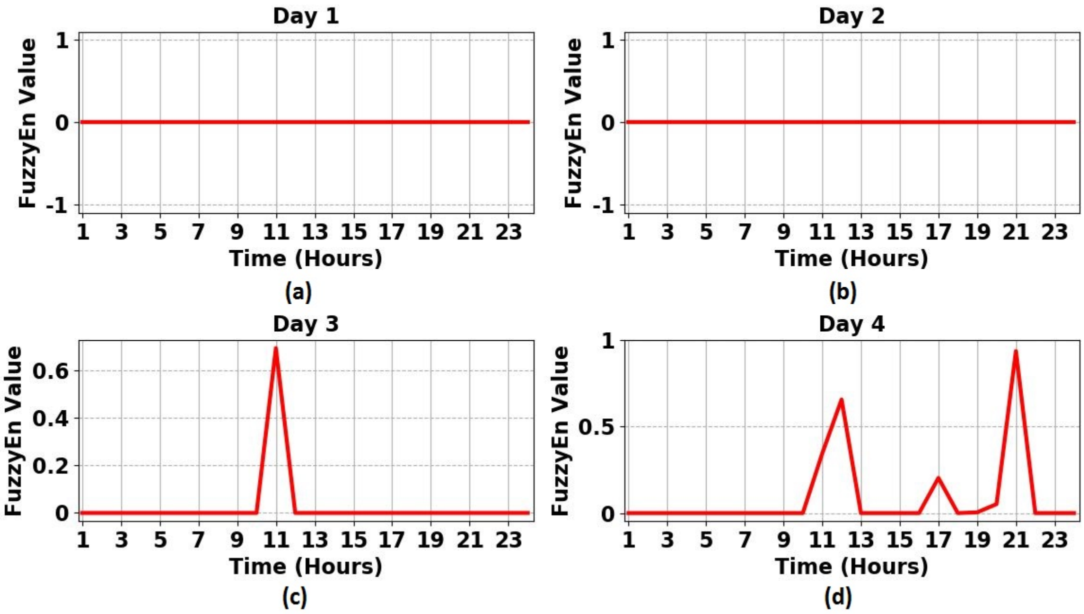


Figure 6.6: Fuzzy entropy values representing visiting time in each day compared with the standard deviation using Dataset F; (**a**) and (**b**) show that no visitor came on day 1 and 2 because the fuzzy entropy values are zero (no bumps in the entropy values); (c) shows that the visitor came once on day 3, as there are one bumps in the fuzzy entropy values; and (d) shows that the visitor came three times on day 4 because there are three bumps in the fuzzy entropy values.

and FuzzyEn produce similar results for accuracy, precision, and recall, which means that the presence of a visitor was accurately identified in the home environment. Therefore, the results of accuracy, precision, and recall indicate that ApEn and FuzzyEn are the best measures for identifying multi-occupancy in a home environment with relatively high accuracy. In contrast, when the period used in the calculation of the entropy measures was 15 minutes, the results show that all three entropy measures achieve very low performance. This can be justified by the fact that decreasing the calculation period will reduce the number of observations per time period, which will increase the variance. Consequently, the number of false positives will increase, which reduces precision.

6.4 Parameters Impact Assessment

To evaluate the robustness of the proposed method described in this work, the impacts of the parameters m and r on the classification performance of the entropy measures are investigated. The selection of parameters m and r needed for the computations of the named entropy measures may be different when they are applied to the ADLs datasets. To investigate the impact of changing the values of these parameters, the performance of the algorithm is examined using Dataset D.

Table 6.3 shows the results of the experiment in terms of the effect of changing the parameter values m and r required for the computation of ApEn, SampEn, and FuzzyEn measures using Dataset D. Clearly, the result of precision and recall shows that the best results are obtained when the value of m is 2 and r ranges from 0.2 to 1.8 respectively. Based on the current results, it appears that when m and r values are increased, the performance of the algorithm is decreased. To summarise, the algorithm of ApEn, SampEn, and FuzzyEn are affected by choice of parameter values m and r.

Based on the results obtained from both experiments, the best performance is obtained when the computational time is performed based on one hour time periods. Therefore, it is helpful to evaluate how a visitor can be identified when a different shifting of computational time is considered. Table 6.4 represents

Table 6.3: The classification results of the effect of changing the parameter values m and r required for the computation of ApEn, SampEn, and FuzzyEn measures using Dataset D.

m⇒	1		2		3		6		10	
r ⇓	Prec. (%)	Rec. (%)	Prec. (%)	Rec. (%)	Prec. (%)	Rec. (%)	Prec. (%)	Rec. (%)	Prec. (%)	Rec. (%)
0.2	100	83	**100**	**100**	33.3	83.3	16.6	50	10	50
0.6	100	83	**100**	**100**	33.3	83.3	16.6	50	10	50
1	100	83	**100**	**100**	33.3	83.3	16.6	50	10	50
1.8	100	83	**100**	**100**	33.3	83.3	16.6	50	10	50
2	23.5	66.6	25	66.6	23.5	66.6	14	50	8	33
3	23.5	66.6	25	66.6	23.5	66.6	14	50	8	33
5	14.2	50	11.1	50	10.7	50	7.5	33	6	33

the classification performance of ApEn, SampEn, and FuzzyEn using dataset B when they are computed at different shifting times. It is observed that the best performance of ApEn, SampEn, and FuzzyEn is obtained when the computational time is performed based on one hour time periods with no shifting time and overlapping, $x \geq 30\%$. The percentage of overlapping, x, is calculated as:

$$x = \frac{\text{Overlap of the visitor period}}{\text{Calculation period}} \times 100 \tag{6.1}$$

This means that by using the calculation period of one hour without shifting, the visitor can be accurately identified. On the other hand, when the shifting time of 15, 30, and 45 minutes are used to calculate the entropy measures, the results show that all entropy measures achieved very low performance with less precision. This can be justified by the fact that when the value of x is decreased, the number of false positives will be increased, which reduces the precision. This means that the proposed methods can be used for identifying the visitor if the time period of one hour and overlapping of $\geq 30\%$ are used.

According to the results shown in Tables 6.1 and 6.2, the best performance is obtained when the computational time is performed based on one hour time

Table 6.4: The classification results of Entropy measures using different shifting time when the computational time is performed based on one-hour time period.

Results	ApEn	SampEn	FuzzyEn	Shifting time	Overlapping $(x)\%$
Accuracy	**99.4**	97	**100**	0 minute	$x \geq 30$
Precision	**100**	61.5	**100**		
Recall	**87.5**	100	**100**		
Accuracy	94	95.5	96.4	15 minutes	$16 \leq x < 29$
Precision	50	56.2	62.5		
Recall	80	90	100		
Accuracy	88	90.4	86.9	30 minutes	$11 \leq x < 15$
Precision	30	35.7	27.2		
Recall	50	41.6	50		
Accuracy	87.5	90.5	91.6	45 minutes	$0 \leq x < 10$
Precision	26.2	33.3	39		
Recall	60	60	70		

periods. Moreover, the results related to the precision demonstrate that ApEn, PerEn, MPE, and FuzzyEn perform much better than ShEn and SampEn as 100%, 87.5%, 95%, 100%, 20% and 61.5% respectively. It can be summarised that the ApEn, PerEn, MPE, and FuzzyEn are relatively better indices to identify multi-occupancy in a home environment. This also confirms that entropy measures could be used to distinguish occupancy data in the presence of a visitor in a home environment.

6.5 Robust Analysis

To evaluate the robustness of the proposed entropy measures for visitor detection, the main door entry sensor is used along with entropy measures to confirm the visitor's presence and time in the home environment. As the visitor enters and exits the home through the main door, the door sensor is utilised to confirm the time of visits. This will increase the performance evaluation of the proposed entropy measures. In general, the door opening or closing does not necessarily mean that a visitor is present in the home environment, as the door might be opened by the main occupant, e.g. in response to a postman or a neighbour. Thus, the presence of a visitor cannot be detected only by using the main door sensor. Therefore, entropy measures are used to detect the visitor, and then the door sensor is used to confirm the time of the visit.

Figure 6.7 shows the confirmation of the visiting time each day based on fuzzy entropy values and using a door sensor. As the best results are obtained when the computational time of entropy measures is performed based on one-hour intervals, the door entry sensor is used to confirm the time of visits in this case. It is depicted in Figure 6.7(a), that the main door was opened six times on day 2, but the visitor came twice on that day, first at 9:25 am and stayed in the home until 10:57 am, and second at 7:04 pm until 7:53 pm. This means that the main occupant might have caused the other door events. Whilst the entropy measures can detect the visitor based on one-hour periods, they do not specify the exact time of the visit. For example, in Figure 6.7(c), the visitor came twice on the day 6, at around 09:00 am and 6:00 pm, without knowing the specific time of the visit. Therefore, the door sensor is used to confirm the time of the visits.

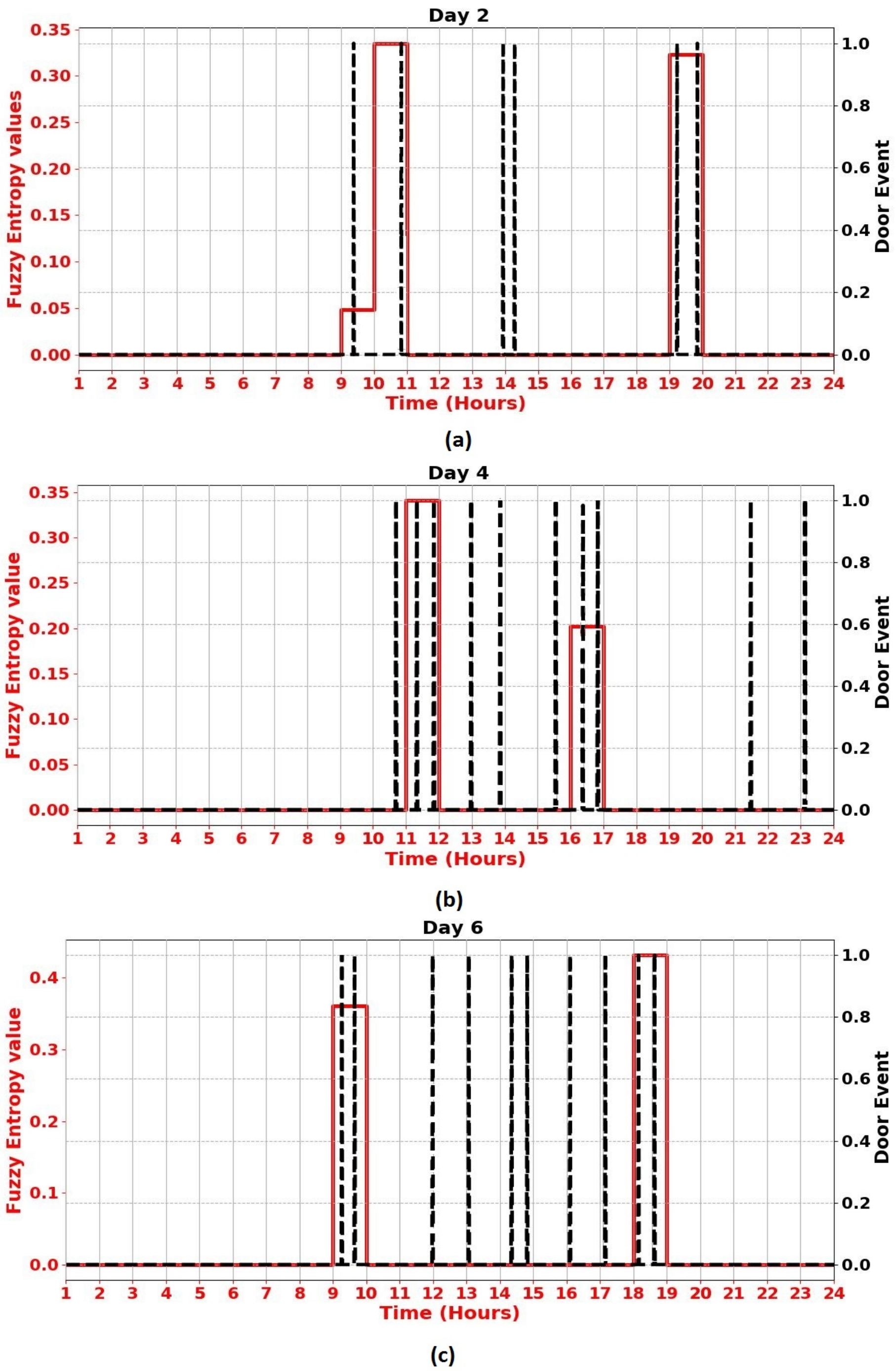


Figure 6.7: Examples of the time confirmation of visits using a door sensor with Fuzzy Entropy (FuzzyEn) measures for the Dataset D representing: (a-c) fuzzy entropy values representing visiting times on days 2, 4, and 6, and the time is confirmed using the door sensor.

On this day, it can be confirmed from the door sensor data that the visitor came at 9:12 am, stayed in the home until 9:44 am, and then came at 6:07 pm until 6:38 pm.

In summary, the entropy measures are powerful tools to detect abnormality (here, multi-occupancy) in behaviour when the sample data is mostly representing normal activities (here, single-occupancy). This also confirms the possibility that abnormality detection by entropy measures can be confirmed with door sensors data, particularly for identifying the exact visiting times.

6.6 Comparison with Existing Modelling Techniques

In order to evaluate the proposed method described in this chapter, the performance of the proposed entropy measures is compared to other methods that achieve the same goal. Considering the literature review carried out for this research as mentioned in Chapter 2, the most commonly utilised approaches for detecting a visitor in a home environment are SVM [23, 110] and MMPP [12]. Therefore, to evaluate the proposed methods carried out in this research, the results obtained by applying ShEn, ApEn, SampEn, PerEn, MPE, and FuzzyEn entropy measures are compared to other approaches that achieve the same goal, such as SVM, Indoor Mobility (IM), and MMPP.

The Dataset D was applied to the SVM model, as well as the Indoor Mobility (IM) measure, and the results were compared with the proposed entropy measures. The IM is defined as the frequency of the transition from room to room in a home environment. Readers are referred to[175] for more details about this measure. The features used as input to the SVM are the start time of entering into each location (room), the time spent in each room (duration), the encoded number of each room and the transitions from one room to another inside the house. The final preprocessing step is to divide the data into two subsets, one with about 70% of the instances for training, and another with around the remaining 30% of instances for testing. Furthermore, the proposed entropy measures were compared with another study that used

Table 6.5: Comparison of the accuracy, precision, and recall for ShEn, ApEn, SampEn,PerEn, MPE, and FuzzyEn entropy measures with the existing methods SVM, IM, and MMPP.

Approach	Accuracy	Precision	Recall
ShEn	85%	20%	25%
ApEn	**99.4%**	**100%**	**87.5%**
SampEn	97%	61.5%	100%
PerEn	**98.8%**	**87.5%**	**87.5%**
MPE	**99.1%**	**95%**	**87.5%**
FuzzyEn	**100%**	**100%**	**100%**
SVM	82.2%	70.8%	72.8%
IM	93.5%	84%	83%
MMPP	78.6	75.2%	78.4%

MMPP to detect visits in a home environment [12]. Multiple datasets were used in their research based on the data gathered from binary sensors, which were collected by the authors (note that the authors did not use a public dataset). The results presented in Table 6.5 show the classification performance of ShEn, ApEn, SampEn, PerEn, MPE and FuzzyEn compared with the existing SVM, IM, and MMPP in terms of accuracy, precision, and recall.

According to the results achieved in Table 6.5, the ApEn, PerEn, MPE, and FuzzyEn entropy measures are considerably better for visitor detection in a home environment compared to other approaches. The ApEn, PerEn, MPE, and FuzzyEn produce an overall accuracy of 99.4%, 98.8%, 99.1%, and 100% respectively. This also confirms that that entropy measures could be used to detect visitors in a home environment.

6.7 Discussion

In this chapter, a novel method based on different entropy measures is proposed to identify visitors and the time of their visits based on non-intrusive sensors data. The proposed entropy measures are employed to investigate their effectiveness in identifying visitors in a home environment. The proposed method is based on the hypothesis that the values of entropy measures are higher than a nominal value when a visitor is present in the home environment, which is represented as an abnormality in behaviour when the sample data mostly represents normal activities. Therefore, when the entropy values of each day exceed the standard deviation, then the event is associated with the presence of a visitor.

In this work, simulated and real home environments, including two experiments, are used to show the effectiveness of the proposed entropy measures for visitor detection in a home environment. The results obtained from both experiments show that the visitor could be identified with a high degree of accuracy based on the data collected from the PIR sensors. The impact of changing the values of embedded dimension m and tolerance r on the classification performance of the entropy measures were also investigated. The experimental results show that the proposed method obtained a high identification rate of 100% when $m = 2$ and $r = 1$. It should be noted that the values of ApEn, SampEn, and FuzzyEn are affected by the choice of parameter values m and r. To evaluate the robustness of the proposed entropy measures, a door entry sensor with entropy measures was utilised to confirm the presence time of the visitor in the home environment.

According to the results shown in Table 6.5, the ApEn, PerEn, MPE, and FuzzyEn entropy measures are considerably better for visitor detection in a home environment compared to other approaches. The ApEn, PerEn, MPE, and FuzzyEn produce an overall accuracy of 99.4%, 98.8%, 99.1%, and 100% respectively. This also confirms that that entropy measures could be used to detect the visitor in a home environment. The conclusion for this investigation is that ApEn, PerEn, MPE, and FuzzyEn are shown to be the best entropy measures in identifying multi-occupancy in a home environment. This is a preferred alternative solution compared with using wearable sensors or visual cameras with associated privacy concern.

Although several attempts have been made to address the challenge of detecting and identifying multi-occupancy in a home environment solely based on the information collected from ambient sensors, the proposed entropy measures for visitor detection proves to be efficient in achieving the goal of identifying visitors and the time of their visits in multi-occupancy environments.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

The work presented in this thesis is a novel entropy-based approach for anomaly detection in activities of daily living to support independent living, particularly for older adults living alone within a home environment. Based on the results obtained from this research, it can be concluded that the ability to detect and distinguish anomalies in ADLs using entropy measures depends mainly on the level of changes in a resident's ADLs pattern. This research hypothesises that the level of changes in a resident's activity patterns in a home environment is an indicator of normal or abnormal activities. Hence, when the entropy value for a day exceeds the threshold value, this could be an indication that there is an abnormality in the resident's activity.

This research aimed to investigate the effectiveness of different entropy measures in detecting and identifying various types of anomalies in daily activities. The motivation for the work is to find an acceptable solution that can be used to detect and identify anomalies in ADLs in a single-occupancy and multi-occupancy environment. As a starting point for detecting anomalies in ADLs, the investigation of the effectiveness of entropy measures initially focused on a single-occupant environment, when only one individual is monitored, and their activities are detected as normal or abnormal (e.g., irregular sleep and human falls). Then, the research investigated the effectiveness of entropy measures for anomaly detection in a multi-occupancy environment. Furthermore, the entropy measures were not only used to detect anomalies in ADLs but also to identify potential causes of anomalies, and to distinguish anomalies in ADLs data (here, irregular sleep in the resident's activity and visitors).

The datasets provided for this investigation were gathered based on two environments, real and simulated home environments, as well as from publicly available datasets. For real home environments, five different datasets representing human activities were presented based on information obtained using ambient sensors. However, one of these datasets was gathered using one accelerometer sensor to examine whether the entropy measures can be used for human fall detection, solely based on information collected from wearable sensors.

In summary, throughout this research, original knowledge on anomaly detection in ADLs in a single-occupancy and multi-occupancy environment has been presented. In the remaining part of this chapter, the research conclusions with a summary of major contributions and the direction of future work are presented.

7.2 Summary of Major Contributions

The proposed method employed to achieve the aim set out in this thesis led to significant contributions. These contributions are discussed as follows:

7.2.1 A Novel Entropy-Based Method for Irregular Sleep Detection.

This thesis presented a novel method based on a Multi-scale Fuzzy Entropy (MFE) measure for distinguishing between normal and anomalous cases in ADLs, specifically in sleeping routine, from data obtained using ambient sensory-based devices, such as the Passive Infra-red (PIR) sensor. A novel feature, namely the standard deviation of MFE values, was applied to identify whether there is an anomaly in the resident's activity or not. The proposed method was based on the assumption that when the value of the MFE measure overrides the standard deviation boundaries, the case is indicated as an anomaly in ADLs. Furthermore, the entropy measures were not only used to detect anomalies in ADLs but also to examine the possible causes of the identified anomalous days (e.g., Less sleep, interrupted sleep, and late sleep). Detecting such anomalies will assist carers in acting to avert prospective problems early and to improve older adults' quality of life. Experiments are conducted based on two different datasets to demonstrate the effectiveness of the proposed method.

The results obtained based on the data gathered solely from ambient sensory devices, show the ability of the proposed method to distinguish between normal and anomalous cases in ADLs with a high degree of accuracy. Comparisons with other methods have also offered support to the proposed method. This also confirms that the Multi-scale Fuzzy Entropy is a promising technique to distinguish between normal and anomalous events in a resident's activity in the home environment.

7.2.2 A Novel Entropy-Based Method for Human Fall Detection

The work presented in the thesis also proposed a novel method based on Fuzzy Entropy (FuzzyEn) measure to detect and distinguish human fall from other activities, solely based on the information gathered from a wearable motion-sensing device. The aim of the research was to investigate whether the FuzzyEn measure can be used to detect human falls during daily activities. Since the resident's normal daily activity pattern is completely different when an abnormal event has occurred, the data recorded from accelerometer devices during daily activities is used to show abnormal (e.g. fall) patterns. The proposed method is based on the hypothesis that the value of entropy is high when there is a fall event. Therefore, the proposed method aims to detect a large value of the entropy. It is supposed that human falls have greater acceleration than other ADLs. A novel feature, namely the standard deviation of the mean of FuzzyEn values is used to confirm whether or not there is a fall. Therefore, when the value of the FuzzyEn measure exceeds the upper standard deviation boundaries, the event is detected as a fall.

Considering the results achieved from the conducted experiments, it is shown that FuzzyEn obtained a high detection rate, of 100%, and a low false-positive rate, of 2.2%. The proposed FuzzyEn entropy measure is considerably better for human fall detection compared to other approaches. The FuzzyEn produces 100% sensitivity and 97.8% specificity. This also confirms that the FuzzyEn measure can be used to detect human falls during ADLs in a home environment based on data acquired from an accelerometer device.

7.2.3 A Novel Entropy-Based Method for Anomaly Detection in Activities of Daily Living in the Presence of a Visitor

This thesis presented a novel entropy-based approach for anomaly detection in ADLs in the presence of a visitor, solely based on information gathered from low-cost, non-intrusive ambient sensors, which include Passive Infra-Red (PIR) sensors and a door entry sensor. The entropy measures, including Shannon Entropy (ShEn), Approximate Entropy (ApEn), Sample Entropy (SampEn), Permutation Entropy (PerEn), Multi-scale Permutation Entropy (MPE), FuzzyEn, and MFE, have been applied in two scenarios. The first case was to reveal days with abnormal behaviours, leading to the identification of the days in which abnormalities occurred. In the second case, the entropy measures were used to detect anomalies in ADLs and also identify potential causes of those anomalies (here, an irregular sleep pattern and detecting a visitor) by calculating the entropy values. The distinction between normal and abnormal entropy values was achieved in the second case by finding the maximum entropy value on normal days around the clock. This meant that any value that exceeded the calculated maximum value for entropy on normal days was treated as an abnormal behaviour point.

When the entropy values for each hour on a given day exceed the threshold value, the entropy measures indicate an anomaly in ADLs at that hour. This means that by finding the maximum entropy value on normal days of ADLs, it is possible to detect abnormal behaviour in ADLs in completely unseen data.

To distinguish between the entropy changes caused by irregular sleep in the resident's activity and a visitor, the main door entry sensor along with entropy measures are used to confirm the time of the visitor's presence in the home environment. The experimental results show that the PerEn, MPE, FuzzyEn, and MFE measures perform much better than the ShEn, ApEn, and SampEn measures to detect anomalies in behaviour when the sample data mostly represents normal activities. The PerEn, MPE, FuzzyEn, and MFE measures show high detection rates of 100%, which means that the false-negative rate of anomaly detection in ADLs is 0%. The conclusion drawn from this research is that the PerEn, MPE, FuzzyEn, and MFE measures are considerably better than ShEn, ApEn, and SampEn measures for anomaly detection in ADLs based on data gathered only from ambient sensors.

7.2.4 A Novel Entropy-Based Method for visitor detection in Multi-Occupancy Environments

This thesis has investigated a means of detecting a visitor in a single-occupancy home environment (represented as a multi-occupancy environment) based on different entropy measures using ambient sensors. The proposed method is based on the hypothesis that the values of entropy measures are higher than a nominal value when a visitor is present in the home environment, which is represented as an abnormality in behaviour when the sample data mostly represents normal activities. The threshold, based on the standard deviation of the resident data in conjunction with entropy measures (ShEn, ApEn, SampEn, PerEn, MPE, and FuzzyEn), is applied to detect when the visitor is present in the home environment.

When the entropy values of each day exceed the standard deviation, the event is associated with the presence of a visitor. To evaluate the robustness of the proposed entropy measures, a door entry sensor is used along with the entropy values to confirm the time and duration of the visitor in the home environment. Experiments are conducted on two different datasets to investigate the effectiveness of entropy measures to identify visitors and the time of their visits without employing extra wearable sensors to tag the visitors.

The results obtained show that ApEn, PerEn, MPE, and FuzzyEn perform much better than ShEn and SampEn as 100%, 87.5%, 95%, 100%, 20% and 61.5% respectively. The conclusion for this investigation is that the ApEn, PerEn, MPE, and FuzzyEn measures are shown to be the best entropy measures in detecting visitors in a home environment based on data gathered from the ambient sensor. Furthermore, it shows that the ApEn, PerEn, MPE, and FuzzyEn measures outperform ShEn and SampEn measures, which confirms that entropy measures could be used to detect the visitor in a home environment. This is a preferred alternative solution compared with using wearable sensors or visual cameras with associated privacy concerns.

7.3 Future Work and Recommendations

Similar to any research, the need for future work for the improvement of the proposed framework is recommended. This section identifies the directions for future research and recommendations for improvement of the framework for anomaly detection in activities of daily living.

- Application of several other entropy measures.

  The possibility of using entropy to determine the degree of disorder or uncertainty in a system resulted in the definition of different types of entropy [37]. Although the framework proposed in this thesis applied seven different entropy measures that are more relevant and adapted to work with binary series information, other entropy measures can be applied and tested for their effectiveness in identifying and detecting anomalies in daily activities, such as Spectral Entropy, Dispersion Entropy, and Multi-scale Dispersion Entropy.

- Extension of entropy measures for a multi-model system.

  The framework developed in this thesis is proposed for anomaly detection in a single-occupancy and a multi-occupancy environment from data obtained using either ambient sensors or wearable sensors. Future work would be the fusion of ambient sensors with wearable sensors to investigate the effectiveness of entropy measures for anomaly detection in ADLs. This could also be used to provide information about separating the activities of one person from another person living in the same environment. The only restriction is that one person will have a wristband worn at all times. The system could also be extended for human fall detection in the presence of a visitor based on information obtained using both ambient and wearable sensors.

- Evaluation of entropy based on visual sensors.

  The datasets used in this thesis were gathered using ambient sensory based-devices and wearable sensory based-devices. A future study is recommended to investigate the effectiveness of entropy measures to detect and identify various types of anomalies in daily activities in single-occupancy or multi-occupancy environments, solely based on information gathered from visual sensors. It will be interesting to investigate whether entropy measures can be used to detect and distinguish anomalies in ADLs using vision-based sensors.

- Extension to other applications.

  The entropy measures framework presented in this research focused on identifying and detecting four different anomalies in ADLs, including irregular sleep, human falls, irregular sleep in the presence of a visitor, and visitor detection in a multi-occupancy environment. Future work should consider other applications that were not considered in this research, such as gait recognition systems.

- The proposed entropy measures are based on identifying a resident's normal daily pattern to detect any anomalies in the resident's activity. Since the normal daily activity patterns are completely different from one person to another. It will be interesting to extend the work to develop the proposed method to be used to detect any anomalies in any resident's activity without identifying the occupancy normal patterns.

References

[1] Saisakul Chernbumroong, Shuang Cang, Anthony Atkins, and Hongnian Yu, "Elderly activities recognition and classification for applications in assisted living", *Expert Systems with Applications*, vol. 40, no. 5, pp. 1662–1674, 2013. 1

[2] Ercilia Dini and Shayla Goldring, "Estimating the changing population of the'oldest old"', *Population Trends*, , no. 132, pp. 8, 2008. 1, 61

[3] Qin Ni, Ana Belén García Hernando, De la Cruz, and Iván Pau, "The elderly's independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development", *Sensors*, vol. 15, no. 5, pp. 11312–11362, 2015. 1, 61

[4] Jinghuan Guo, Yiming Li, Mengnan Hou, Shuo Han, and Jianxun Ren, "Recognition of daily activities of two residents in a smart home based on time clustering", *Sensors*, vol. 20, no. 5, pp. 1457, 2020. 1, 32, 36, 58, 61

[5] Abdulhamit Subasi, Kholoud Khateeb, Tayeb Brahimi, and Akila Sarirete, "Human activity recognition using machine learning methods in a smart healthcare environment", in *Innovation in Health Informatics*, pp. 123–144. Elsevier, 2020. 1, 4, 38, 58, 94

[6] Tan-Hsu Tan, Munkhjargal Gochoo, Fu-Rong Jean, Shih-Chia Huang, and Sy-Yen Kuo, "Front-door event classification algorithm for elderly people living alone in smart house using wireless binary sensors", *IEEE Access*, vol. 5, pp. 10734–10743, 2017. 1, 15

[7] Tan-Hsu Tan, Munkhjargal Gochoo, Shih-Chia Huang, Yi-Hung Liu, Shing-Hong Liu, and Yun-Fa Huang, "Multi-resident activity recognition in a

smart home using rgb activity image and dcnn", *IEEE Sensors Journal*, vol. 18, no. 23, pp. 9718–9727, 2018. 1, 29, 36

[8] Damla Arifoglu and Abdelhamid Bouchachia, "Detection of abnormal behaviour for dementia sufferers using convolutional neural networks", *Artificial intelligence in medicine*, vol. 94, pp. 88–95, 2019. 1, 5, 15, 17, 34, 38

[9] Maoguo Gong, Huimin Zeng, Yu Xie, Hao Li, and Zedong Tang, "Local distinguishability aggrandizing network for human anomaly detection", *Neural Networks*, vol. 122, pp. 364–373, 2020. 1, 15

[10] Hyunjun Ju, Dongha Lee, Junyoung Hwang, Junghyun Namkung, and Hwanjo Yu, "Pumad: Pu metric learning for anomaly detection", *Information Sciences*, 2020. 1, 5, 19, 34, 38

[11] Meriem Zerkouk and Belkacem Chikhaoui, "Spatio-temporal abnormal behavior prediction in elderly persons using deep learning models", *Sensors*, vol. 20, no. 8, pp. 2359, 2020. 1

[12] Ahmed Nait Aicha, Gwenn Englebienne, and Ben Kröse, "Unsupervised visit detection in smart homes", *Pervasive and Mobile Computing*, vol. 34, pp. 157–167, 2017. 2, 8, 26, 36, 124, 125

[13] Hande Alemdar and Cem Ersoy, "Multi-resident activity tracking and recognition in smart environments", *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, no. 4, pp. 513–529, 2017. 2, 3, 4, 26, 36, 108

[14] Asma Benmansour, Abdelhamid Bouchachia, and Mohammed Feham, "Modeling interaction in multi-resident activities", *Neurocomputing*, vol. 230, pp. 133–142, 2017. 2, 3, 4, 27, 32, 108

[15] Ifat Afrin Emi and John A Stankovic, "Sarrima: smart adl recognizer and resident identifier in multi-resident accommodations", in *Proceedings of the conference on Wireless Health*. ACM, 2015, p. 4. 2, 29, 32, 36

[16] Jianguo Hao, Abdenour Bouzouane, and Sébastien Gaboury, "Recognizing multi-resident activities in non-intrusive sensor-based smart homes by

formal concept analysis", *Neurocomputing*, vol. 318, pp. 75–89, 2018. 2, 3, 32, 108

[17] Raihani Mohamed, Thinagaran Perumal, Md Nasir Sulaiman, Norwati Mustapha, and Syaifulnizam Abd Manaf, "Tracking and recognizing the activity of multi resident in smart home environments", *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 9, no. 2-11, pp. 39–43, 2017. 2, 32, 63, 67

[18] Ghassem Mokhtari, Amjad Anvari-Moghaddam, Qing Zhang, and Mohanraj Karunanithi, "Multi-residential activity labelling in smart homes with wearable tags using ble technology", *Sensors*, vol. 18, no. 3, pp. 908, 2018. 2, 3, 4, 25, 30, 108, 109

[19] Ferhat Attal, Samer Mohammed, Mariam Dedabrishvili, Faicel Chamroukhi, Latifa Oukhellou, and Yacine Amirat, "Physical human activity recognition using wearable sensors", *Sensors*, vol. 15, no. 12, pp. 31314–31338, 2015. 2, 27, 31, 36

[20] Mei Han, Wei Xu, Hai Tao, and Yihong Gong, "An algorithm for multiple object trajectory tracking", in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* IEEE, 2004, vol. 1, pp. I–I. 2, 27, 36

[21] Raihani Mohamed, Thinagaran Perumal, Md Nasir Sulaiman, and Norwati Mustapha, "Multi resident complex activity recognition in smart home: A literature review", *Int. J. Smart Home*, vol. 11, no. 6, pp. 21–32, 2017. 2, 31

[22] Brad Schumitsch, Heiko Schwarz, and Thomas Wiegand, "Optimization of transform coefficient selection and motion vector estimation considering interpicture dependencies in hybrid video coding", in *Image and Video Communications and Processing 2005*. International Society for Optics and Photonics, 2005, vol. 5685, pp. 327–335. 2, 32, 36

[23] Rui Hu, Hieu Pham, Philipp Buluschek, and Daniel Gatica-Perez, "Elderly people living alone: Detecting home visits with ambient and wearable

sensing", in *Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care*. ACM, 2017, pp. 85–88. 2, 3, 25, 28, 36, 38, 94, 108, 109, 124

[24] Shyamal Patel, Hyung Park, Paolo Bonato, Leighton Chan, and Mary Rodgers, "A review of wearable sensors and systems with application in rehabilitation", *Journal of neuroengineering and rehabilitation*, vol. 9, no. 1, pp. 21, 2012. 2

[25] Ari Yair Barrera-Animas, Luis A Trejo, Miguel Angel Medina-Pérez, Raúl Monroy, J Benito Camiña, and Fernando Godínez, "Online personal risk detection based on behavioural and physiological patterns", *Information Sciences*, vol. 384, pp. 281–297, 2017. 3

[26] Marko Borazio, Eugen Berlin, Nagihan Kücükyildiz, Philipp Scholl, and Kristof Van Laerhoven, "Towards benchmarked sleep detection with wrist-worn sensing units", in *2014 IEEE International Conference on Healthcare Informatics*. IEEE, 2014, pp. 125–134. 3, 16

[27] Arindam Ghosh, Amartya Chakraborty, Joydeep Kumbhakar, Mousumi Saha, and Sujoy Saha, "Humansense: a framework for collective human activity identification using heterogeneous sensor grid in multi-inhabitant smart environments", *PERSONAL AND UBIQUITOUS COMPUTING*, 2020. 3, 108

[28] Fatma Najar, Sami Bourouis, Nizar Bouguila, and Safya Belghith, "Unsupervised learning of finite full covariance multivariate generalized gaussian mixture models for human activity recognition", *Multimedia Tools and Applications*, pp. 1–23, 2019. 3, 108

[29] Liang Wang, Tao Gu, Xianping Tao, Hanhua Chen, and Jian Lu, "Recognizing multi-user activities using wearable sensors in a smart home", *Pervasive and Mobile Computing*, vol. 7, no. 3, pp. 287–298, 2011. 3, 27, 108, 109

[30] Nirmalya Roy, Archan Misra, and Diane Cook, "Ambient and smartphone sensor assisted adl recognition in multi-inhabitant smart environments",

Journal of ambient intelligence and humanized computing, vol. 7, no. 1, pp. 1–19, 2016. 3, 31, 36, 108

[31] Geetika Singla, Diane J Cook, and Maureen Schmitter-Edgecombe, "Recognizing independent and joint activities among multiple residents in smart environments", *Journal of ambient intelligence and humanized computing*, vol. 1, no. 1, pp. 57–63, 2010. 3, 108

[32] Qimeng Li, Raffaele Gravina, Ye Li, Saeed H Alsamhi, Fangmin Sun, and Giancarlo Fortino, "Multi-user activity recognition: Challenges and opportunities", *Information Fusion*, 2020. 4, 25, 108

[33] Son N Tran, Qing Zhang, and Mohan Karunanithi, "On multi-resident activity recognition in ambient smart-homes", *arXiv preprint arXiv:1806.06611*, vol. 1806, pp. 10, 2018. 4, 108

[34] Patrick Lapointe, Kévin Chapron, Kévin Bouchard, et al., "A new device to track and identify people in a multi-residents context", *Procedia Computer Science*, vol. 170, pp. 403–410, 2020. 4, 25, 32, 108

[35] Asma Benmansour, Abdelhamid Bouchachia, and Mohammed Feham, "Multioccupant activity recognition in pervasive smart home environments", *ACM Computing Surveys (CSUR)*, vol. 48, no. 3, pp. 34, 2016. 4, 25, 109

[36] Enamul Hoque, Robert F Dickerson, Sarah M Preum, Mark Hanson, Adam Barth, and John A Stankovic, "Holmes: A comprehensive anomaly detection system for daily in-home activities", in *2015 International Conference on Distributed Computing in Sensor Systems*. IEEE, 2015, pp. 40–51. 4, 19, 34, 37, 39, 79

[37] Marta Borowska, "Entropy-based algorithms in the analysis of biomedical signals", *Studies in Logic, Grammar and Rhetoric*, vol. 43, no. 1, pp. 21–32, 2015. 4, 5, 42, 43, 45, 132

[38] Abdur Rahim Mohammad Forkan, Ibrahim Khalil, Zahir Tari, Sebti Foufou, and Abdelaziz Bouras, "A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient

assisted living", *Pattern Recognition*, vol. 48, no. 3, pp. 628–641, 2015. 4, 16, 17, 34

[39] Shehroz S Khan, Michelle E Karg, Dana Kulić, and Jesse Hoey, "Detecting falls with x-factor hidden markov models", *Applied Soft Computing*, vol. 55, pp. 168–177, 2017. 4, 21, 35

[40] Huorong Ren, Zhixing Ye, and Zhiwu Li, "Anomaly detection based on a dynamic markov model", *Information Sciences*, vol. 411, pp. 52–65, 2017. 4, 16, 17, 34

[41] Jens Lundström, Eric Järpe, and Antanas Verikas, "Detecting and exploring deviating behaviour of smart home residents", *Expert Systems with Applications*, vol. 55, pp. 429–440, 2016. 4, 16, 34

[42] Quentin Bramas and Sébastien Tixeuil, "The complexity of data aggregation in static and dynamic wireless sensor networks", *Information and Computation*, vol. 255, pp. 369–383, 2017. 5

[43] Narayanan C Krishnan and Diane J Cook, "Activity recognition on streaming sensor data", *Pervasive and mobile computing*, vol. 10, pp. 138–154, 2014. 5

[44] Eunju Kim, Sumi Helal, and Diane Cook, "Human activity recognition and pattern discovery", *IEEE pervasive computing*, vol. 9, no. 1, pp. 48–53, 2009. 5, 37

[45] Guto Leoni Santos, Patricia Takako Endo, Kayo Henrique de Carvalho Monteiro, Elisson da Silva Rocha, Ivanovitch Silva, and Theo Lynn, "Accelerometer-based human fall detection using convolutional neural networks", *Sensors*, vol. 19, no. 7, pp. 1644, 2019. 5, 20, 22, 23, 35, 88

[46] Carl Chalmers, Paul Fergus, Casimiro Adays Curbelo Montanez, Sudip Sikdar, Freya Ball, and Bryony Kendall, "Detecting activities of daily living and routine behaviours in dementia patients living alone using smart meter load disaggregation", *IEEE Transactions on Emerging Topics in Computing*, 2020. 5, 16, 18, 34

[47] Leila Panahi and Vahid Ghods, "Human fall detection using machine vision techniques on rgb–d images", *Biomedical Signal Processing and Control*, vol. 44, pp. 146–153, 2018. 5, 20, 22, 35

[48] Sankalp Garg, Bijaya Ketan Panigrahi, and Deepak Joshi, "An accelerometer based fall detection system using deep neural network", in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*. IEEE, 2019, pp. 1–7. 5, 22, 94

[49] Jin Cui, Khaled Boussetta, and Fabrice Valois, "Classification of data aggregation functions in wireless sensor networks", *Computer Networks*, p. 107342, 2020. 5

[50] Soon-Chang Poh, Yi-Fei Tan, Xiaoning Guo, Soon-Nyean Cheong, Chee-Pun Ooi, and Wooi-Haw Tan, "Lstm and hmm comparison for home activity anomaly detection", in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. IEEE, 2019, pp. 1564–1568. 5, 15, 18, 34, 38, 39, 78

[51] Ahmed Nait Aicha, Gwenn Englebienne, and Ben Kröse, "How busy is my supervisor?: Detecting the visits in the office of my supervisor using a sensor network", in *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2012, p. 12. 8, 25, 28

[52] Salisu Wada Yahaya, Ahmad Lotfi, Mufti Mahmud, Pedro Machado, and Naoyuki Kubota, "Gesture recognition intermediary robot for abnormality detection in human activities", 2019. 8, 39, 77, 79

[53] Paolo Barsocchi, Mario GCA Cimino, Erina Ferro, Alessandro Lazzeri, Filippo Palumbo, and Gigliola Vaglini, "Monitoring elderly behavior via indoor position-based stigmergy", *Pervasive and Mobile Computing*, vol. 23, pp. 26–42, 2015. 15

[54] Shehroz S Khan, Michelle E Karg, Jesse Hoey, and Dana Kulic, "Towards the detection of unusual temporal events during activities using hmms", in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 1075–1084. 15

[55] Ane Alberdi Aramendi, Alyssa Weakley, Asier Aztiria Goenaga, Maureen Schmitter-Edgecombe, and Diane J Cook, "Automatic assessment of functional health decline in older adults based on smart home data", *Journal of biomedical informatics*, vol. 81, pp. 119–130, 2018. 15, 39, 78

[56] Damla Arifoglu and Abdelhamid Bouchachia, "Activity recognition and abnormal behaviour detection with recurrent neural networks", *Procedia Computer Science*, vol. 110, pp. 86–93, 2017. 16

[57] Junshui Ma and Simon Perkins, "Time-series novelty detection using one-class support vector machines", in *Proceedings of the International Joint Conference on Neural Networks, 2003.* IEEE, 2003, vol. 3, pp. 1741–1745. 16, 18

[58] Shiri Gaber, Omer Sagi, and Avitan Gefen, "Identifying anomalies in user internet of things activity profile using analytic engine", Apr. 30 2020, US Patent App. 16/173,291. 16

[59] Manish Sainani, Adam Jamison Oliner, Jacob Barton Leverich, Leonid Alekseyev, and Sonal Barton Maheshwari, "Anomaly detection", Mar. 17 2020, US Patent 10,592,093. 16

[60] Krasimir Tonchev, Pavlina Koleva, Agata Manolova, Georgi Tsenov, and Vladimir Poulkov, "Non-intrusive sleep analyzer for real time detection of sleep anomalies", in *2016 39th International Conference on Telecommunications and Signal Processing (TSP).* IEEE, 2016, pp. 400–404. 17, 34

[61] Sowmya Ramapatruni, Sandeep Nair Narayanan, Sudip Mittal, Anupam Joshi, and Karuna Joshi, "Anomaly detection models for smart home security", in *2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS).* IEEE, 2019, pp. 19–24. 17

[62] Sulaiman Alhaidari and Mohamed Zohdy, "Network anomaly detection using two-dimensional hidden markov model based viterbi algorithm",

in *2019 IEEE International Conference On Artificial Intelligence Testing (AITest)*. IEEE, 2019, pp. 17–18. 17

[63] Diane J Cook, "Learning setting-generalized activity models for smart spaces", *IEEE intelligent systems*, vol. 2010, no. 99, pp. 1, 2010. 18, 34

[64] Wael Khreich, Babak Khosravifar, Abdelwahab Hamou-Lhadj, and Chamseddine Talhi, "An anomaly detection system based on variable n-gram features and one-class svm", *Information and Software Technology*, vol. 91, pp. 186–197, 2017. 18

[65] Mohammad Saleh Sadooghi and Siamak Esmaeilzadeh Khadem, "Improving one class support vector machine novelty detection scheme using nonlinear features", *Pattern Recognition*, vol. 83, pp. 14–33, 2018. 18

[66] Zaruhi Alaverdyan, Julien Jung, Romain Bouet, and Carole Lartizien, "Regularized siamese neural network for unsupervised outlier detection on brain multiparametric magnetic resonance imaging: application to epilepsy lesion screening", *Medical Image Analysis*, vol. 60, pp. 101618, 2020. 18, 34

[67] Andrew B Gardner, Abba M Krieger, George Vachtsevanos, and Brian Litt, "One-class novelty detection for seizure analysis from intracranial eeg", *Journal of Machine Learning Research*, vol. 7, no. Jun, pp. 1025–1044, 2006. 18

[68] Salisu Wada Yahaya, Ahmad Lotfi, and Mufti Mahmud, "A consensus novelty detection ensemble approach for anomaly detection in activities of daily living", *Applied Soft Computing*, vol. 83, pp. 105613, 2019. 19, 34, 39, 63, 87

[69] Habtamu Fanta, Zhiwen Shao, and Lizhuang Ma, "Sitgru: Single-tunnelled gated recurrent unit for abnormality detection", *Information Sciences*, 2020. 19, 34

[70] Labiba Gillani Fahad and Muttukrishnan Rajarajan, "Anomalies detection in smart-home activities", in *2015 IEEE 14th International Conference on*

Machine Learning and Applications (ICMLA). IEEE, 2015, pp. 419–422. 19

[71] Earnest Paul Ijjina, "Human fall detection using temporal templates and convolutional neural networks", in *Proceedings of the Third International Conference on Computational Intelligence and Informatics*. Springer, 2020, pp. 763–772. 20, 22

[72] Jin Zhang, Cheng Wu, and Yiming Wang, "Human fall detection based on body posture spatio-temporal evolution", *Sensors*, vol. 20, no. 3, pp. 946, 2020. 20

[73] Panagiotis Kostopoulos, Tiago Nunes, Kevin Salvi, Michel Deriaz, and Julien Torrent, "F2d: A fall detection system tested with real data from daily life of elderly people", in *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*. IEEE, 2015, pp. 397–403. 20, 88

[74] Xiuyi Fan, Huiguo Zhang, Cyril Leung, and Zhiqi Shen, "Robust unobtrusive fall detection using infrared array sensors", in *2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2017, pp. 194–199. 20

[75] Behzad Mirmahboub, Shadrokh Samavi, Nader Karimi, and Shahram Shirani, "Automatic monocular system for human fall detection based on variations in silhouette area", *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 2, pp. 427–436, 2012. 20

[76] Qing Zhang and Mohan Karunanithi, "Feasibility of unobstrusive ambient sensors for fall detections in home environment", in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 566–569. 20

[77] Kripesh Adhikari, Hamid Bouchachia, and Hammadi Nait-Charif, "Deep learning based fall detection using simplified human posture", *International Journal of Computer and Systems Engineering*, vol. 13, no. 5, pp. 255–260, 2019. 20

[78] Umar Asif, Benjamin Mashford, Stefan Von Cavallar, Shivanthan Yohanandan, Subhrajit Roy, Jianbin Tang, and Stefan Harrer, "Privacy preserving human fall detection using video data", in *Machine Learning for Health Workshop*, 2020, pp. 39–51. 20

[79] Ahmad Lotfi, Suad Albawendi, Heather Powell, Kofi Appiah, and Caroline Langensiepen, "Supporting independent living for older adults; employing a visual based fall detection through analysing the motion and shape of the human body", *IEEE Access*, vol. 6, pp. 70272–70282, 2018. 20

[80] Vaishna S Kumar, Kavan Gangadhar Acharya, B Sandeep, T Jayavignesh, and Ashvini Chaturvedi, "Wearable sensor-based human fall detection wireless system", in *Wireless Communication Networks and Internet of Things*, pp. 217–234. Springer, 2019. 20

[81] Xiaoqun Yu, Hai Qiu, and Shuping Xiong, "A novel hybrid deep neural network to predict pre-impact fall for older people based on wearable inertial sensors", *Frontiers in bioengineering and biotechnology*, vol. 8, 2020. 20

[82] Pranesh Vallabh and Reza Malekian, "Fall detection monitoring systems: a comprehensive review", *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 6, pp. 1809–1833, 2018. 20

[83] Adrián Núñez-Marcos, Gorka Azkune, and Ignacio Arganda-Carreras, "Vision-based fall detection with convolutional neural networks", *Wireless communications and mobile computing*, vol. 2017, 2017. 20, 22

[84] Thathupara Subramanyan Kavya, Young-Min Jang, Erdenetuya Tsogtbaatar, and Sang-Bock Cho, "Fall detection system for elderly people using vision-based analysis", *SCIENCE AND TECHNOLOGY*, vol. 23, no. 1, pp. 69–83, 2020. 20

[85] Yixiao Yun and Irene Yu-Hua Gu, "Human fall detection in videos by fusing statistical features of shape and motion dynamics on riemannian manifolds", *Neurocomputing*, vol. 207, pp. 726–734, 2016. 20

[86] S-G Miaou, Pei-Hsu Sung, and Chia-Yuan Huang, "A customized human fall detection system using omni-camera images and personal information",

in *1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare, 2006. D2H2.* IEEE, 2006, pp. 39–42. 20

[87] Bogdan Kwolek and Michal Kepski, "Human fall detection on embedded platform using depth maps and wireless accelerometer", *Computer methods and programs in biomedicine*, vol. 117, no. 3, pp. 489–501, 2014. 20, 22, 35, 64, 94

[88] Patricia Bet, Paula C Castro, and Moacir A Ponti, "Fall detection and fall risk assessment in older person using wearable sensors: a systematic review", *International journal of medical informatics*, 2019. 20, 88

[89] T Theodoridis, V Solachidis, N Vretos, and P Daras, "Human fall detection from acceleration measurements using a recurrent neural network", in *Precision Medicine Powered by pHealth and Connected Health*, pp. 145–149. Springer, 2018. 20, 22, 23, 35, 88, 94

[90] Eduardo Casilari, José-Antonio Santoyo-Ramón, and José-Manuel Cano-García, "Analysis of public datasets for wearable fall detection systems", *Sensors*, vol. 17, no. 7, pp. 1513, 2017. 20

[91] Nicolas Thome and Serge Miguet, "A hhmm-based approach for robust fall detection", in *2006 9th International Conference on Control, Automation, Robotics and Vision*. IEEE, 2006, pp. 1–8. 21, 35

[92] Claudio Coppola, Serhan Cosar, Diego R Faria, and Nicola Bellotto, "Automatic detection of human interactions from rgb-d data for social activity classification", in *2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 2017, pp. 871–876. 21

[93] Shuo Yu, Hsinchun Chen, and Randall A Brown, "Hidden markov model-based fall detection with motion sensor orientation calibration: A case for real-life home monitoring", *IEEE journal of biomedical and health informatics*, vol. 22, no. 6, pp. 1847–1853, 2017. 22

[94] Huiqiang Cao, Shuicai Wu, Zhuhuang Zhou, Chung-Chih Lin, Chih-Yu Yang, Shih-Tseng Lee, and Chieh-Tsai Wu, "A fall detection method

based on acceleration data and hidden markov model", in *2016 IEEE International Conference on Signal and Image Processing (ICSIP)*. IEEE, 2016, pp. 684–689. 22

[95] Nabil Zerrouki, Fouzi Harrou, Ying Sun, and Amrane Houacine, "Accelerometer and camera-based strategy for improved human fall detection", *Journal of medical systems*, vol. 40, no. 12, pp. 284, 2016. 22

[96] Na Lu, Yidan Wu, Li Feng, and Jinbo Song, "Deep learning for fall detection: Three-dimensional cnn combined with lstm on video kinematic data", *IEEE journal of biomedical and health informatics*, vol. 23, no. 1, pp. 314–323, 2018. 22

[97] Weidong Min, Hao Cui, Hong Rao, Zhixun Li, and Leiyue Yao, "Detection of human falls on furniture using scene analysis based on deep learning and activity characteristics", *IEEE Access*, vol. 6, pp. 9324–9335, 2018. 22

[98] Pramod Kumar Soni and Ayesha Choudhary, "Automated fall detection from a camera using support vector machine", in *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*. IEEE, 2019, pp. 1–6. 23, 35

[99] Chadia Khraief, Faouzi Benzarti, and Hamid Amiri, "Convolutional neural network based on dynamic motion and shape variations for elderly fall detection", *International Journal of Machine Learning and Computing*, vol. 9, no. 6, 2019. 23, 35

[100] Caroline Rougier, Jean Meunier, Alain St-Arnaud, and Jacqueline Rousseau, "Fall detection from human shape and motion history using video surveillance", in *21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07)*. IEEE, 2007, vol. 2, pp. 875–880. 23, 35

[101] Lourdes Martínez-Villaseñor and Hiram Ponce, "Design and analysis for fall detection system simplification", *Journal of visualized experiments*, , no. 158, 2020. 24

[102] Paulo Vitor C Souza, Augusto J Guimaraes, Vanessa S Araujo, Lucas O Batista, and Thiago S Rezende, "An interpretable machine learning model for human fall detection systems using hybrid intelligent models", in *Challenges and Trends in Multimodal Fall Detection for Healthcare*, pp. 181–205. Springer, 2020. 24

[103] Yoosuf Nizam, Mohd Norzali Haji Mohd, and M Mahadi Abdul Jamil, "Human fall detection from depth images using position and velocity of subject", *Procedia Computer Science*, vol. 105, pp. 131–137, 2017. 24

[104] Kaibo Fan, Ping Wang, and Shuo Zhuang, "Human fall detection using slow feature analysis", *Multimedia Tools and Applications*, vol. 78, no. 7, pp. 9101–9128, 2019. 24

[105] Edouard Auvinet, Caroline Rougier, Jean Meunier, Alain St-Arnaud, and Jacqueline Rousseau, "Multiple cameras fall dataset", *DIRO-Université de Montréal, Tech. Rep*, vol. 1350, 2010. 24, 35

[106] Xin Ma, Haibo Wang, Bingxia Xue, Mingang Zhou, Bing Ji, and Yibin Li, "Depth-based human fall detection via shape features and improved extreme learning machine", *IEEE journal of biomedical and health informatics*, vol. 18, no. 6, pp. 1915–1922, 2014. 24

[107] Umar Asif, Stefan Von Cavallar, Jianbin Tang, and Stefan Harre, "Sshfd: Single shot human fall detection with occluded joints resilience", *arXiv preprint arXiv:2004.00797*, 2020. 24, 35

[108] Imen Charfi, Johel Miteran, Julien Dubois, Mohamed Atri, and Rached Tourki, "Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and adaboost-based classification", *Journal of Electronic Imaging*, vol. 22, no. 4, pp. 041106, 2013. 24, 35

[109] Ahmed Nait Aicha, Gwenn Englebienne, and Ben Kröse, "How lonely is your grandma?: detecting the visits to assisted living elderly from wireless sensor network data", in *Proceedings of the ACM conference on Pervasive and ubiquitous computing adjunct publication*, 2013, pp. 1285–1294. 25, 27

[110] Johanna Petersen, Nicole Larimer, Jeffrey A Kaye, Misha Pavel, and Tamara L Hayes, "Svm to detect the presence of visitors in a smart home environment", in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 5850–5853. 25, 28, 36, 124

[111] Son N Tran and Qing Zhang, "Towards multi-resident activity monitoring with smarter safer home platform", in *Smart Assisted Living*, pp. 249–267. Springer, 2020. 25

[112] Daniele Riboni and Flavia Murru, "Unsupervised recognition of multi-resident activities in smart-homes", *IEEE Access*, 2020. 25

[113] Christopher M Bishop et al., *Neural networks for pattern recognition*, Oxford university press, 1995. 26

[114] Mohamed Eldib, Francis Deboeverie, Dirk Van Haerenborgh, Wilfried Philips, and Hamid Aghajan, "Detection of visitors in elderly care using a low-resolution visual sensor network", in *Proceedings of the 9th International Conference on Distributed Smart Cameras*. ACM, 2015, pp. 56–61. 26

[115] Kevin P Murphy, *Machine learning: a probabilistic perspective*, MIT press, 2012. 26

[116] Ahmed Nait Aicha, Gwenn Englebienne, and Ben Kröse, "Modeling visit behaviour in smart homes using unsupervised learning", in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 2014, pp. 1193–1200. 26

[117] "W.s. university: Casas smart home project", `http://ailab.eecs.wsu.edu/casas/`, Accessed on 18 April 2019. 27, 33

[118] Bryan David Minor, Janardhan Rao Doppa, and Diane J Cook, "Learning activity predictors from sensor data: Algorithms, evaluation, and applications", *IEEE transactions on knowledge and data engineering*, vol. 29, no. 12, pp. 2744–2757, 2017. 29

[119] Tao Gu, Zhanqing Wu, Liang Wang, Xianping Tao, and Jian Lu, "Mining emerging patterns for recognizing activities of multiple users in pervasive computing", in *Mobile and Ubiquitous Systems: Networking & Services, MobiQuitous, 2009. MobiQuitous' 09. 6th Annual International*. IEEE, 2009, pp. 1–10. 30, 109

[120] Alaa Alhamoud, Arun Asokan Nair, Christian Gottron, Doreen Böhnstedt, and Ralf Steinmetz, "Presence detection, identification and tracking in smart homes utilizing bluetooth enabled smartphones", in *Local Computer Networks Workshops (LCN Workshops), 2014 IEEE 39th Conference on*. IEEE, 2014, pp. 784–789. 30

[121] Ji Xiong, Fangmin Li, Ning Zhao, and Na Jiang, "Tracking and recognition of multiple human targets moving in a wireless pyroelectric infrared sensor network", *Sensors*, vol. 14, no. 4, pp. 7209–7228, 2014. 30

[122] Ghassem Mokhtari, Qing Zhang, Ghavameddin Nourbakhsh, Stephen Ball, and Mohanraj Karunanithi, "Bluesound: A new resident identification sensor—using ultrasound array and ble technology for smart home platform", *IEEE Sensors Journal*, vol. 17, no. 5, pp. 1503–1512, 2017. 30

[123] Duy Nguyen, Luong Nguyen, and Son Nguyen, "A novel approach of ontology-based activity segmentation and recognition using pattern discovery in multi-resident homes", in *Frontiers in Intelligent Computing: Theory and Applications*, pp. 167–178. Springer, 2020. 32, 36

[124] Nirmalya Roy, Archan Misra, and Diane Cook, "Infrastructure-assisted smartphone-based adl recognition in multi-inhabitant smart environments", in *2013 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2013, pp. 38–46. 32

[125] Tinghui Wang and Diane J Cook, "smrt: Multi-resident tracking in smart homes with sensor vectorization", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 32

[126] Kuo-Chung Hsu, Yi-Ting Chiang, Gu-Yang Lin, Ching-Hu Lu, Jane Yung-Jen Hsu, and Li-Chen Fu, "Strategies for inference mechanism of

conditional random fields for multiple-resident activity recognition in a smart home", in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, 2010, pp. 417–426. 32

[127] Diane J Cook, Aaron Crandall, Geetika Singla, and Brian Thomas, "Detection of social interaction in smart spaces", *Cybernetics and Systems: An International Journal*, vol. 41, no. 2, pp. 90–104, 2010. 33

[128] Markus Prossegger and Abdelhamid Bouchachia, "Multi-resident activity recognition using incremental decision trees", in *Adaptive and Intelligent Systems*, pp. 182–191. Springer, 2014. 33, 36

[129] Hande Alemdar, Halil Ertan, Ozlem Durmaz Incel, and Cem Ersoy, "Aras human activity datasets in multiple homes with multiple residents", in *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2013, pp. 232–235. 33

[130] Rong Chen and Yu Tong, "A two-stage method for solving multi-resident activity recognition in smart environments", *Entropy*, vol. 16, no. 4, pp. 2184–2203, 2014. 33

[131] Yi-Ting Chiang, Kuo-Chung Hsu, Ching-Hu Lu, Li-Chen Fu, and Jane Yung-Jen Hsu, "Interaction models for multiple-resident activity recognition in a smart home", in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 3753–3758. 33

[132] Can Tunca, Hande Alemdar, Halil Ertan, Ozlem Durmaz Incel, and Cem Ersoy, "Multimodal wireless sensor network-based ambient assisted living in real homes with multiple residents", *Sensors*, vol. 14, no. 6, pp. 9692–9719, 2014. 33, 36

[133] Weixin Li, Vijay Mahadevan, and Nuno Vasconcelos, "Anomaly detection and localization in crowded scenes", *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 1, pp. 18–32, 2013. 34

[134] Cewu Lu, Jianping Shi, and Jiaya Jia, "Abnormal event detection at 150 fps in matlab", in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 2720–2727. 34

[135] Arthur Asuncion and David Newman, "Uci machine learning repository", 2007. 34

[136] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. 34

[137] Diane J Cook and Maureen Schmitter-Edgecombe, "Assessing the quality of activities in a smart environment", *Methods of information in medicine*, vol. 48, no. 05, pp. 480–485, 2009. 34

[138] Diane J Cook, Aaron S Crandall, Brian L Thomas, and Narayanan C Krishnan, "Casas: A smart home in a box", *Computer*, vol. 46, no. 7, pp. 62–69, 2012. 34, 63

[139] Yixin Chen, Xin Dang, Hanxiang Peng, and Henry L Bart, "Outlier detection with the kernelized spatial depth function", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 288–305, 2008. 34

[140] R Thörner, M Persson, H Eriksson, A Isaksson, and J Lundström, "Trygg om natten", *Tech. Rep.*, 2011. 34

[141] Eduardo Casilari, Raúl Lora-Rivera, and Francisco García-Lagos, "A study on the application of convolutional neural networks to fall detection evaluated with multiple public datasets", *Sensors*, vol. 20, no. 5, pp. 1466, 2020. 35

[142] Korbinian Frank, Maria Josefa Vera Nadales, Patrick Robertson, and Tom Pfeifer, "Bayesian recognition of motion related activities with inertial sensors", in *Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing-Adjunct*, 2010, pp. 445–446. 35

[143] Carlos Medrano, Raul Igual, Inmaculada Plaza, and Manuel Castro, "Detecting falls as novelties in acceleration patterns acquired with smartphones", *PloS one*, vol. 9, no. 4, 2014. 35

[144] Ahmet Turan Özdemir, "An analysis on sensor locations of the human body for wearable fall detection devices: Principles and practice", *Sensors*, vol. 16, no. 8, pp. 1161, 2016. 35

[145] Taylor R Mauldin, Marc E Canby, Vangelis Metsis, Anne HH Ngu, and Coralys Cubero Rivera, "Smartfall: A smartwatch-based fall detection system using deep learning", *Sensors*, vol. 18, no. 10, pp. 3363, 2018. 35

[146] Vera Nadales and Maria Josefa, *Recognition of human motion related activities from sensors*, PhD thesis, University of Malaga, 2010. 35

[147] George Vavoulas, Matthew Pediaditis, Emmanouil G Spanakis, and Manolis Tsiknakis, "The mobifall dataset: An initial evaluation of fall detection algorithms using smartphones", in *13th IEEE International Conference on BioInformatics and BioEngineering*. IEEE, 2013, pp. 1–4. 35

[148] Manan Jethanandani, Thinagaran Perumal, Jieh-Ren Chang, Abhishek Sharma, and Yipeng Bao, "Multi-resident activity recognition using multi-label classification in ambient sensing smart homes", in *2019 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*. IEEE, 2019, pp. 1–2. 36

[149] Chun Zhu, Qi Cheng, and Weihua Sheng, "Human activity recognition via motion and vision data fusion", in *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*. IEEE, 2010, pp. 332–336. 36

[150] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou, "Isolation-based anomaly detection", *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 1, pp. 1–39, 2012. 37

[151] Guansong Pang, Chunhua Shen, and Anton van den Hengel, "Deep anomaly detection with deviation networks", in *Proceedings of the 25th*

ACM SIGKDD international conference on knowledge discovery & data mining, 2019, pp. 353–362. 37

[152] Ahmad Jalal, Kibum Kim, et al., "Wearable inertial sensors for daily activity analysis based on adam optimization and the maximum entropy markov model", *Entropy*, vol. 22, no. 5, pp. 579, 2020. 38

[153] Samih Eisa and Adriano Moreira, "A behaviour monitoring system (bms) for ambient assisted living", *Sensors*, vol. 17, no. 9, pp. 1946, 2017. 39

[154] Denis Elbert, Holger Storf, Michael Eisenbarth, Özgür Ünalan, and Mario Schmitt, "An approach for detecting deviations in daily routine for long-term behavior analysis", in *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. IEEE, 2011, pp. 426–433. 39, 79

[155] R Clausius, "On the motive power of heat and the laws which can be deduced therefrom regarding the theory of heat", *Annalen der Physik*, vol. 155, no. 4, pp. 500–524, 1850. 42

[156] Alfréd Rényi et al., "On measures of entropy and information", in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California, 1961. 42

[157] Claude Elwood Shannon, "A mathematical theory of communication", *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948. 42, 45

[158] Mostafa Rostaghi and Hamed Azami, "Dispersion entropy: A measure for time-series analysis", *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 610–614, 2016. 42, 44

[159] Steven M Pincus, "Approximate entropy as a measure of system complexity.", *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991. 43, 45, 46, 99

[160] Joshua S Richman and J Randall Moorman, "Physiological time-series analysis using approximate entropy and sample entropy", *American Journal*

of Physiology-Heart and Circulatory Physiology, vol. 278, no. 6, pp. H2039–H2049, 2000. 43, 47

[161] Manish N Tibdewal, Himanshu R Dey, Manjunatha Mahadevappa, AjoyKumar Ray, and Monika Malokar, "Multiple entropies performance measure for detection and localization of multi-channel epileptic eeg", *Biomedical Signal Processing and Control*, vol. 38, pp. 158–167, 2017. 43, 46, 50

[162] Weiting Chen, Zhizhong Wang, Hongbo Xie, and Wangxin Yu, "Characterization of surface emg signal based on fuzzy entropy", *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 15, no. 2, pp. 266–272, 2007. 43, 51

[163] Christoph Bandt and Bernd Pompe, "Permutation entropy: a natural complexity measure for time series", *Physical review letters*, vol. 88, no. 17, pp. 174102, 2002. 44, 48, 49

[164] Hamed Azami, Alberto Fernández, and Javier Escudero, "Refined multiscale fuzzy entropy based on standard deviation for biomedical signal analysis", *Medical & biological engineering & computing*, vol. 55, no. 11, pp. 2037–2052, 2017. 44

[165] Madalena Costa, Ary L Goldberger, and C-K Peng, "Multiscale entropy analysis of biological signals", *Physical review E*, vol. 71, no. 2, pp. 021906, 2005. 44, 50

[166] Jörgen Bruhn, Lutz E Lehmann, Heiko Röpcke, Thomas W Bouillon, and Andreas Hoeft, "Shannon entropy applied to the measurement of the electroencephalographic effects of desflurane", *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 95, no. 1, pp. 30–35, 2001. 45

[167] Alfonso Delgado-Bonal and Alexander Marshak, "Approximate entropy and sample entropy: A comprehensive tutorial", *Entropy*, vol. 21, no. 6, pp. 541, 2019. 47

[168] Wen Li, Dong Ming, Rui Xu, Hao Ding, Hongzhi Qi, and Baikun Wan, "Research on visual attention classification based on eeg entropy parameters", in *World Congress on Medical Physics and Biomedical Engineering May 26-31, 2012, Beijing, China*. Springer, 2013, pp. 1553–1556. 47

[169] Francesco Carlo Morabito, Domenico Labate, Fabio La Foresta, Alessia Bramanti, Giuseppe Morabito, and Isabella Palamara, "Multivariate multi-scale permutation entropy for complexity analysis of alzheimer's disease eeg", *Entropy*, vol. 14, no. 7, pp. 1186–1202, 2012. 48

[170] Duan Li, Zhenhu Liang, Yinghua Wang, Satoshi Hagihira, Jamie W Sleigh, and Xiaoli Li, "Parameter selection in permutation entropy for an electroencephalographic measure of isoflurane anesthetic drug effect", *Journal of clinical monitoring and computing*, vol. 27, no. 2, pp. 113–123, 2013. 49

[171] Wajid Aziz and Muhammad Arif, "Multiscale permutation entropy of physiological time series", in *2005 Pakistan Section Multitopic Conference*. IEEE, 2005, pp. 1–6. 50

[172] V Vakharia, VK Gupta, and PK Kankar, "A multiscale permutation entropy based approach to select wavelet for fault diagnosis of ball bearings", *Journal of Vibration and Control*, vol. 21, no. 16, pp. 3123–3131, 2015. 50

[173] Shuen-De Wu, Po-Hung Wu, Chiu-Wen Wu, Jian-Jiun Ding, and Chun-Chieh Wang, "Bearing fault diagnosis based on multiscale permutation entropy and support vector machine", *Entropy*, vol. 14, no. 8, pp. 1343–1356, 2012. 50

[174] Jinde Zheng, Junsheng Cheng, Yu Yang, and Songrong Luo, "A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination", *Mechanism and Machine Theory*, vol. 78, pp. 187–200, 2014. 52

[175] Saisakul Chernbumroong, Ahmad Lotfi, and Caroline Langensiepen, "Prediction of mobility entropy in an ambient intelligent environment", in

Intelligent Agents (IA), 2014 IEEE Symposium on. IEEE, 2014, pp. 65–72. 53, 59, 124

[176] R Jansi and R Amutha, "Sparse representation based classification scheme for human activity recognition using smartphones", *Multimedia Tools and Applications*, vol. 78, no. 8, pp. 11027–11045, 2019. 54

[177] Yonggang Lu, Ye Wei, Li Liu, Jun Zhong, Letian Sun, and Ye Liu, "Towards unsupervised physical activity recognition using smartphone accelerometers", *Multimedia Tools and Applications*, vol. 76, no. 8, pp. 10701–10719, 2017. 54

[178] Carlos Ramos, Juan Carlos Augusto, and Daniel Shapiro, "Ambient intelligence—the next step for artificial intelligence", *IEEE Intelligent Systems*, vol. 23, no. 2, pp. 15–18, 2008. 59

[179] Emile Aarts and Reiner Wichert, "Ambient intelligence", in *Technology guide*, pp. 244–249. Springer, 2009. 60

[180] Ahmad Lotfi, Caroline Langensiepen, Sawsan M Mahmoud, and Mohammad Javad Akhlaghinia, "Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour", *Journal of ambient intelligence and humanized computing*, vol. 3, no. 3, pp. 205–218, 2012. 61

[181] Dua Dheeru and Efi Karra Taniskidou, "Uci machine learning repository. available online", `https://archive.ics.uci.edu/ml/index.php`, 2017, (Accessed on 18 January 2019). 66

[182] Abubaker Elbayoudi, Ahmad Lotfi, and Caroline Langensiepen, "The human behaviour indicator: A measure of behavioural evolution", *Expert Systems with Applications*, vol. 118, pp. 493 – 505, 2019. 67

[183] Fabien Cardinaux, Simon Brownsell, David Bradley, and Mark S Hawley, "A home daily activity simulation model for the evaluation of lifestyle monitoring systems", *Computers in biology and medicine*, vol. 43, no. 10, pp. 1428–1436, 2013. 68

[184] Balázs Kormányos and Béla Pataki, “Multilevel simulation of daily activities: Why and how?”, in *2013 IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA)*. IEEE, 2013, pp. 1–6. 68

[185] “Realgames. home i/o smart home automation simulation. available online:”, https://realgames.co/home-io/, Accessed: 2018-06-24. 69

[186] SB Kotsiantis, Dimitris Kanellopoulos, and PE Pintelas, “Data preprocessing for supervised leaning”, *International Journal of Computer Science*, vol. 1, no. 2, pp. 111–117, 2006. 71

[187] Hyun Kang, “The prevention and handling of the missing data”, *Korean journal of anesthesiology*, vol. 64, no. 5, pp. 402, 2013. 72

[188] Tamal Kumar De, Bart Michiels, René Tanious, and Patrick Onghena, “Handling missing data in randomization tests for single-case experiments: A simulation study”, *Behavior Research Methods*, pp. 1–16, 2020. 72

[189] José Ramón Pasillas-Díaz and Sylvie Ratté, “An unsupervised approach for combining scores of outlier detection techniques, based on similarity measures.”, *Electr. Notes Theor. Comput. Sci.*, vol. 329, pp. 61–77, 2016. 87

[190] Suad Albawendi, Ahmad Lotfi, Heather Powell, and Kofi Appiah, “Video based fall detection using features of motion, shape and histogram”, in *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*. ACM, 2018, pp. 529–536. 89

[191] Shobhanjana Kalita, Arindam Karmakar, and Shyamanta M Hazarika, “Human fall detection during activities of daily living using extended core9”, in *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*. IEEE, 2019, pp. 1–6. 94