# Fuzzy Finite State Machine for Human Activity Modelling and Recognition


**Gadelhag M Omar Mohmed**

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*

August 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 January 2017 and August 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 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. Many thanks for making your office accessible despite your busy schedules. 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.

I would like to pay tribute 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 Selwan and our lovely kids, Fozia, Mohamed and Ali. 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 father's soul who always encouraged me to carry on in my study, also I thank my mother, sisters, brothers.

Their prayers and faith have inspired and encouraged me to continue with this study. I owe a huge debt of gratitude to my brothers Omar and Dia, and my father in law Abdussalam for being looking after my business back in Libya.

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 research group. To everyone who in one way or the other contributed to the success of my research, I am most grateful.

**Gadelhag Mohmed**
**August 2020**

# Abstract

Independent living is a housing arrangement designed exclusively for older adults to support them with their Activity of Daily Living (ADL) in a safe and secure environment. The provision of independent living would reduce the cost of social care while elderly residents are kept in their own homes. Therefore, there is a need for an automated system to monitor the residents to be able to understand their activities and only when abnormal activities are identified, provide human support to resolve the issue.

Three main approaches are used for gathering data representing the human's activities; ambient sensory device-based, wearable sensory device-based and camera vision device-based. Ambient sensory devices-based systems use sensors such as Passive Infra-Red (PIR) and door entry sensors to capture a user's presence or absence within a specific area and record them as binary information. Gathering data using these sensory devices are widely accepted, as they are unobtrusive and it does not affect the ADLs. However, wearable sensory devices-based and camera vision device-based approaches are undesirable to many users especially for the older adults users as they more often forget to wear them and due to some privacy concerns.

Recognising and modelling human activities from unobtrusive sensors is a topic addressed in Ambient Intelligence (AmI) research. The research proposed in this thesis aims to recognise and model human activities in an indoor environment based on ambient sensory device-based data. Different methods including statistical, machine learning and deep learning techniques are already researched to

address the challenges of recognising and modelling human activities. The research in this thesis is mainly focusing on the application of Fuzzy Finite State Machine (FFSM) for human activities modelling and proposes ways for enhancing the FFSM performance to improve the accuracy of human activity modelling.

In this thesis, three novel contributions are made which are outlined as follows; Firstly, a framework is proposed for combining the learning abilities of Neural Networks (NNs), Long Short-Term Memory (LSTM) neural network and Convolutional Neural Networks (CNNs) with the existing FFSM for human activity modelling and recognition. These models are referred to as NN-FFSM, LSTM-FFSM and CNN-FFSM. Secondly, to obtain the optimal feature representation from the acquired sensory information, relevant features are extracted and fuzzified with the selected membership degrees, these features are then applied to the different enhanced FFSM models. Thirdly, binary data gathered from the ambient sensors including PIR and door entry sensors are represented as greyscale images. A pre-trained Deep Convolutional Neural Network (DCNN) such as AlexNet is used to select and extract features from the generated greyscale image for each activity. The selected features are then used as inputs to Adaptive Boosting (AdaBoost) and Fuzzy C-means (FCM) classifiers for modelling and recognising the ADL for a single user.

The proposed enhanced FFSM models were tested and evaluated using two different datasets representing the ADL for a single user. The first dataset was collected at the Smart Home facilities at NTU and the second dataset is a public dataset collected from CASAS smart home project.

# Publications

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

**Refereed Journal Papers:**

Mohmed, Gadelhag, Ahmad Lotfi, and Amir Pourabdollah. "Enhanced fuzzy finite state machine for human activity modelling and recognition." **Journal of Ambient Intelligence and Humanized Computing** (2020). https://doi.org/10.1007/s12652-020-01917-z

Mohmed, Gadelhag, Ahmad Lotfi, and Amir Pourabdollah. "Human activities recognition based on neuro-fuzzy finite state machine." **Technologies** 6.4 (2018): 110. https://doi.org/10.3390/technologies6040110

**Refereed Conference Papers:**

Mohmed, Gadelhag, Ahmad Lotfi, and Amir Pourabdollah. "Convolutional Neural Network Classifier with Fuzzy Feature Representation for Human Activity." 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE).

Mohmed, Gadelhag, Ahmad Lotfi, and Amir Pourabdollah. "Employing a Deep Convolutional Neural Network for Human Activity Recognition Based on Binary Ambient Sensor Data." Proceedings of the 13th PErvasive Technologies Related to Assistive Environments Conference. 2020.

https://doi.org/10.1145/3389189.3397991

Mohmed, Gadelhag, et al. "Unsupervised learning fuzzy finite state machine for human activities recognition." Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference. 2018. https://doi.org/10.1145/3197768.3201540

Mohmed, Gadelhag, et al. "Clustering-Based Fuzzy Finite State Machine for Human Activity Recognition." UK Workshop on Computational Intelligence. Springer, Cham, 2018. https://doi.org/10.1007/978-3-319-97982-3_22

Mohmed, Gadelhag, Ahmad Lotfi, and Amir Pourabdollah. "Long short-term memory fuzzy finite state machine for human activity modelling." Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments. 2019. https://doi.org/10.1145/3316782.3322781

Mohmed, Gadelhag, David Ada Adama, and Ahmad Lotfi. "Fuzzy Feature Representation with Bidirectional Long Short-Term Memory for Human Activity Modelling and Recognition." UK Workshop on Computational Intelligence. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-29933-0_2

# Contents

# Nomenclature

**Acronyms**

| | |
|---|---|
| ADL | Activities of Daily Living |
| ADW | Activity of Daily Working |
| AmI | Ambient Intelligence |
| ANN | Artificial Neural Network |
| Bi-LSTM | Bidirectional Long Short-Term Memory Neural Networks |
| CI | Computational Intelligence |
| CHMM | Coupled Hidden Markov Models |
| CL-HMM | Combined Label Hidden Markov Models |
| CNN | Convolutional Neural Network |
| CNNC | Convolutional Neural Network Classifier |
| CNN-FFSM | Convolutional Neural Network-Fuzzy Finite State Machine |
| DBN | Deep Belief Network |
| DCNN | Deep Convolutional Neural Network. |
| DMVP | Discovering Methods for Varying Pattern |
| DL | Deep Learning |
| DTEN | Deep Triplet Embeddings Network |
| E-NN | Ensemble Neural Network |
| FCM | Fuzzy C-means |
| FIS | Fuzzy Inference System. |
| FSM | Finite State Machine |
| FFSM | Fuzzy Finite State Machine |
| FLS | Fuzzy Logic System |

| | |
|---|---|
| HAR | Human Activity Recognition |
| HMM | Hidden Markov Model |
| LHMM | Linked Hidden Markov Models |
| LSTM | Long Short-term Memory |
| LSTM-FFSM | Long Short-term Memory-Fuzzy Finite State Machine |
| MF | Membership Function |
| MOPPNN | Multi-object Pattern Producing Neural Network |
| ML | Machine Learning |
| NN | Neural Network |
| N-FFSM | Neuro-Fuzzy Finite State Machine |
| PHMM | Paralle Hidden Markov Models |
| ReLU | Rectified Linear Unit |
| RNN | Recurrent Neural Network |
| PIR | Passive Infrared |
| RPTWSVM | Robust Parametric Twin Support Vector Machine |
| SVM | Support Vector Machine |

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Due to the rising cost of social care, the number of older adults who prefer to live independently in their own home has increased. This independent lifestyle cannot be achieved if the older adult suffers from mild cognitive impairment unless a suitable assistive environment is provided to monitor and recognise their daily activities. Hence, monitoring and recognising human activities in an indoor home environment are studied within the general topic of Ambient Intelligence (AmI) [1, 2]. eef

The social care system in the United Kingdom and worldwide is facing a staggering increase in the cost of looking after elderly people. The preferred option for many older adults is to stay in their homes while they receive assistive care [3, 4]. Research conducted in the area of Human Activity Recognition (HAR) for elderly people focuses on the importance of understanding the activity patterns and helping them to maintain their independent lifestyle. It could also be used for recommending a healthy lifestyle based on the information from monitored activities, and detecting early symptoms of any abnormalities [5]. Several techniques could be used for gathering information representing human activities from a real home environment; employing camera-based systems [6, 7], wearable devices [8, 9, 10] or binary ambient sensors [11, 12] to monitor human activities. Attention has predominantly been focused on data collected by unobtrusive sensors, which are more acceptable by users for privacy concerns [13]. An environment equipped with such monitoring devices is often referred to as an Intelligent or Smart

Environment.

An important factor in designing a smart environment for recognising and modelling human activities is that technology should not interfere with normal daily activities. Thus, all the employed devices should operate autonomously. To achieve this, objects and appliances within the monitoring environment can be supplied with sensors to collect information. The system of collecting and representing human activities based on sensors integrated into the environment is referred to as *ambient sensory device-based* system. Examples of sensors used in such system are; Passive Infrared (PIR) sensors, to capture the movements; electrical sensors, to measure the usage of the electrical devices; temperature sensors to measure the ambient temperature; mat pressure sensors, to measure bed occupancy; light density sensors, to monitor ambient light; and on/off magnetic switch sensors, to monitor doors' open/close positions.

The analysis of the data captured from the environmental sensors can be used to enhance the quality of users' lifestyle in the aspects listed below:

- Address health and safety concerns, e.g. by modelling users' activities, predicting the next activities and then giving help whenever it is needed.
- Increasing users' comfort in their residences, e.g. by adjusting the environmental conditions such as temperature and light based on the users' preferences.
- Optimise energy consumption, e.g. by controlling the usage of light and other electrical devices.

The ambient sensory device-based system records data as binary string information. Therefore different techniques are investigated for representing the sensor readings in an adequate format. To represent features of human activities, statistical and dynamic features in the time domain are the most commonly used features [9, 14]. They are suited for datasets with little noise.

There are many analytical tools to help with analysing the data gathered from a smart environment. Specifically, fuzzy sets and systems are established tools to handle the unpredictable uncertainties in using raw data in many practical applications [15]. For example, the fuzzy feature representation approach is a widely used approach for data representation and visualisation. It uses a fuzzy computational algorithm to represent features from one-dimensional input vectors. Also, it can be used to learn the degree of membership functions through the fuzzification of each input variable in the training data to represent human activities. The membership degrees represent the features used to model and recognise activities. When compared with other feature representation approaches such as deep learning Neural Networks (NNs) which are able to learn features from data, the features derived from the fuzzy feature representation can be flexibly constructed to reduce underlying uncertainties in data [15].

In order to model and recognise human activities using information obtained from ambient sensory devices, one of the promising techniques applied is based on the Finite State Machine (FSM) [16]. The classic FSM is employed to represent the states and the functionality of transitions between different states (in this context, the activities). However, considering the uncertainties incorporated in human activities, a fusion of Fuzzy Logic with FSM allows a more powerful tool to model the dynamic processes that may change over time [16, 17]. Considering the uncertainties incorporated in human activities represented in data obtained from sensors, it is argued that Fuzzy Finite State Machine (FFSM) is a suitable technique to deal with a large uncertain data collected from real-world environments. Furthermore, the system can assign a degree of truth to the occurrence of each activity.

In an FFSM model, the transitions between states are triggered by fuzzy variables instead of crisp values. This provides an accurate model supported by a reasoning mechanism that is represented with a degree of truth related to each state transition. Thus, the system can be in more than one state at any time based on the membership values or degree of belonging to each state [16, 18]. Readers are referred to [19] for the theoretical definition of the FFSM with some recent developments reported in [20, 21].

To model a dynamic process, when data changes over time, the FSM technique can also be utilised [22]. The FSM contains several states representing different actions and the mechanism of transitions between them. Many researchers have considered using the FSM to model and represent human activities. Since human behaviour is not restricted to a single state at any time and there are uncertainties associated with each state, it is reasonable to consider some degree of fuzziness within the FSM, thereby creating a more powerful tool to model dynamic processes that may change over time [18, 23]. In FFSM, both inputs and outputs are treated as fuzzy sets instead of being treated as crisp values. This allows the system to handle and process the input information with a degree of belonging, which often provides more flexibility and human comprehensibility [24]. The FFSM is one of the most suitable technique to deal with a large amount of uncertain data obtained from low-level sensory devices in AmI environments. In this case, the system can assign a degree of truth to the occurrence of each activity. The transitions between the system's states in the FFSM are triggered by fuzzy values rather than crisp values used in the classical FSM. This provides a realistic model supported by a fuzzy reasoning mechanism, represented by a degree of truth related to each state transition. Therefore, more than one state can be in an active mode at any time based on the membership values of each state [16, 18, 25].

## 1.1 Overview of the Research

In an AmI environment, the sensor readings representing human activities are used as input to the system developed to model and recognise human activities. Most of the work in recognising and modelling the activities of a single user is focused on using statistical methods such as Bayesian networks [26] and Hidden Markov Models (HMMs) [27, 28]. Other methods involving the application of computational intelligence methods such as NNs [29] and Fuzzy Inference System (FIS) [30]. Deep learning techniques such as Convolutional Neural Networks (CNNs) are also used in more recent research in this area [31, 32, 33]. The statistical methods are employed to find the relationship between features extracted from the collected sensor readings data, and eventually, identify user's activity pattern. However, these methods encounter difficulties when dealing with a large scale of sensor readings obtained from a real AmI environment [18]. By nature, several activities can be undertaken by a single user at the same time [34]. For example, people can watch TV while they are having their meal.

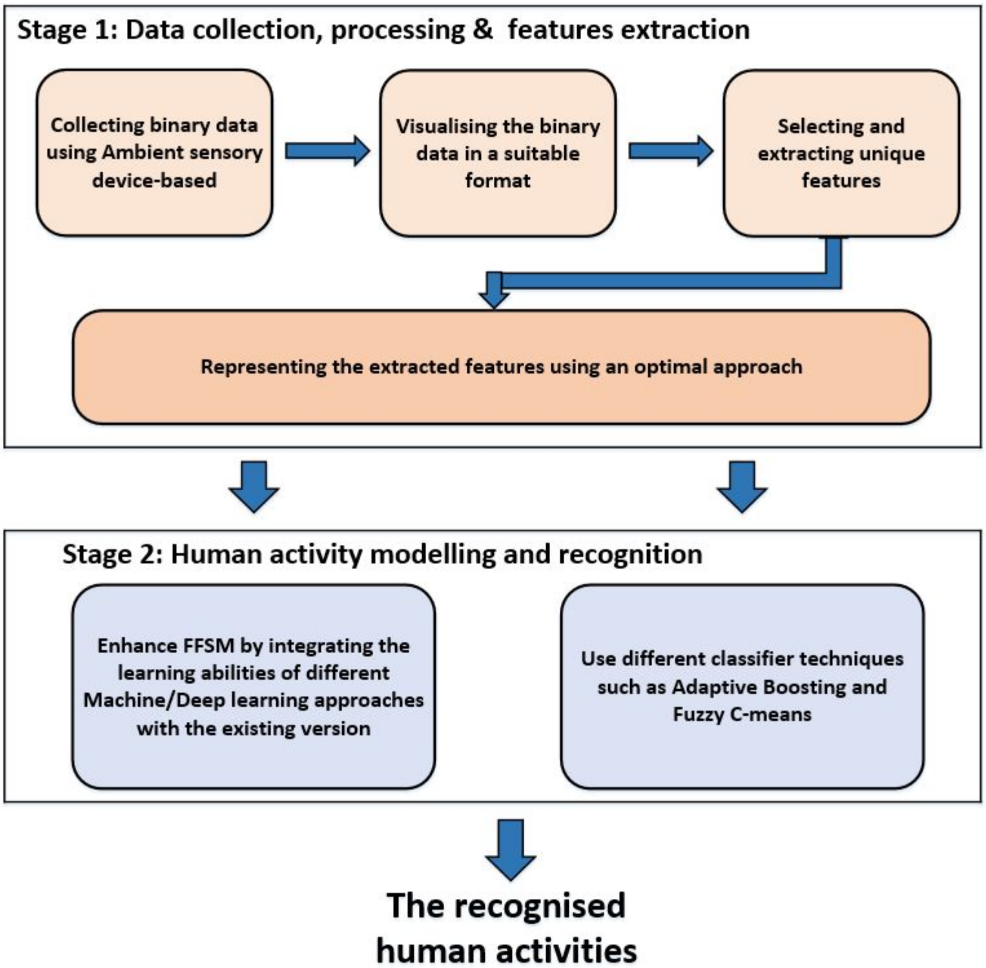



Figure 1.1: Schematic representation of the proposed human activity modelling and recognition framework.

In this particular scenario, it is not necessary to know which activity started first. However, it is essential to know the degree of involvement in each activity at that time. That means the existence of simultaneous activities; when an activity (e.g., eating) starts while the other activity is already started (e.g., watching TV). A specialised approach is required to recognise these non-sequential behaviours.

Figure 1.1 illustrates the schematic framework for human activity modelling and recognition proposed in this thesis. In the proposed model, during the first stage of the model, the sensor readings representing human activities are collected using ambient sensory devices-based. The collected sensor readings are then interpreted and converted into an understandable format, such as occupancy signal in time-series or greyscale images. Several numerical features representing each activity are selected and extracted from this occupancy signal. These numerical features can be extracted by employing practical knowledge-based techniques such as an ontology [35, 36], or using a computational intelligence technique integrated with the collected sensor readings. The extracted features are then represented using a suitable approach to obtain optimal features representation. One of the recommended methods for representing human activity data is the fuzzy feature representation approach. This technique fuzzifies the extracted numerical features to add more flexibility and make the information more human-like. During the second stage of the model, the generated fuzzy sets can be used as inputs to the different enhanced versions of FFSM for modelling and recognising the human activity. The proposed framework of enhanced versions of the FFSM combines the learning abilities of NNs, Long Short-Term Memory (LSTM) neural network and Convolutional Neural Networks (CNNs) for human activity modelling and recognition. The model's outputs are the modelled and recognised human activities based on the collected data.

In addition to the contribution of developing an FFSM model presented in this thesis, another novel approach of visualising binary sensor readings as greyscale images are investigated. This makes use of the information obtained using ambient sensory devices, including PIR and door entry sensors. A pre-trained Deep Convolutional Neural Network (DCNN) such as AlexNet is used to select and extract features from the generated greyscale image for each activity. The selected features are then used as inputs to an Adaptive Boosting (AdaBoost) and Fuzzy C-means (FCM) classifier for modelling and recognising the ADL of a single user.

The research reported in this thesis attempts to answer the following questions:

1. How binary sensor readings representing human activities, which are obtained from ambient sensory devices can be converted into an

occupancy signal in a time-series format for representing ADL for a single user?

2. How to extract numerical features representing each activity?

3. How to represent the extracted features using an optimal approach to overcome the uncertainty associated with the human activity data?

4. How the performance of the existing FFSM can be enhanced by integrating different computational techniques for modelling and recognising human activities?

In order to answer the above questions, the research aims and objectives are explained in the next section.

## 1.2 Research Aims and Objectives

The aim of this research is to investigate human activity modelling and recognition models. The research presented in this thesis addresses only the challenges in modelling and recognising in a single-occupancy real-home environment based on a dataset collected using ambient sensory devices. This involves investigating the ways that could be used to convert and interpret the binary sensory readings in a suitable format, feature extraction and representation, and enhancing an FFSM model by integrating it with different computational intelligence techniques. Once the human activities are recognised, the results obtained could be used in many different applications, including anomaly detection in daily human activities, energy consumption optimisation, addressing health and safety concerns, leading to an improved level of comfort and quality of life. To achieve this aim, the following research objectives are identified:

To determine the answer to question number one of the research questions mentioned earlier, the following objectives are identified:

- Use the ambient sensory device-based system to collect dataset representing ADL from a real smart home environment for a single user.

- Investigate suitable approaches for converting and visualising binary sensor readings collected into an understandable format. This is to achieve a better understanding of constituents of human activities from such data.

In related to question 2 of the research questions, the following objective is identified:

- Select and extract a set of numerical features from the generated occupancy signal and a greyscale image representing each activity.

To answer question 3 of the research questions, the next objective is determined:

- Use a feature representation approach for representing the extracted features. This will reduce the uncertainties associated with the data representing human activities.

The next aims and objectives are identified to answer question 4 of the research questions as follows:

- Investigate existing classification methods and models for human activity recognition and propose a model which is able to learn activities from the extracted features.

- Incorporate human experts' knowledge approaches with machine/deep learning techniques for creating a robust human activity modelling and recognition model. This will reduce the complexities associated with experts-based approaches and the most common machine/deep learning techniques which rely on a large amount of numerical data. This might be obtained by integrating the proposed FFSM with different computational techniques to enhance the learning ability of the system.

- Compare the performance of different approaches that are integrated with the FFSM to assess the most appropriate approach for recognising and modelling human activities.

- Test and evaluate the enhanced FFSM models and the investigated classification approaches using different data sets representing ADL for a single user.

## 1.3 Major Contributions of the Thesis

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

- An extensive literature review is presented discussing the state-of-the-art on human activity modelling and recognition. It encompasses various methods proposed and validated results from other researches related to this thesis.
- A novel framework for visualising binary string data as an occupancy signal and a greyscale image representing human activities is presented. Also, different feature extraction and representation approaches are employed.
- Integrating the FFSM with NN to incorporate the experts' knowledge used in FFSM with the learning abilities in the NN for governing the transition between the system states (Activities). The new model is referred to as a Neuro-Fuzzy Finite State Machine (N-FFSM).
- Integrating the FFSM with Long Short-Term Memory (LSTM) NNs to enhance the learning capability of the FFSM model for accurately generating the fuzzy rules that govern the transition between the system states. The new model is referred to as a Long Short-Term Memory-Fuzzy Finite State Machine (LSTM-FFSM).
- Integrating the FFSM with Convolutional Neural Network (CNN) to add the learning ability to model daily human activities based on the numerical and temporal information gathered from the sensory data. The new model is referred to as Convolutional-Fuzzy Finite State Machine (CNN-FFSM).
- Investigate the usage of a pre-trained CNN for selecting and extracting feature vector(s) representing human activity from generated greyscale image. The extracted feature vector is then used with different classifier for modelling and recognition of human activity.
- Testing and evaluating the proposed models using two different datasets gathered from real home environments representing ADL for a single user.

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.

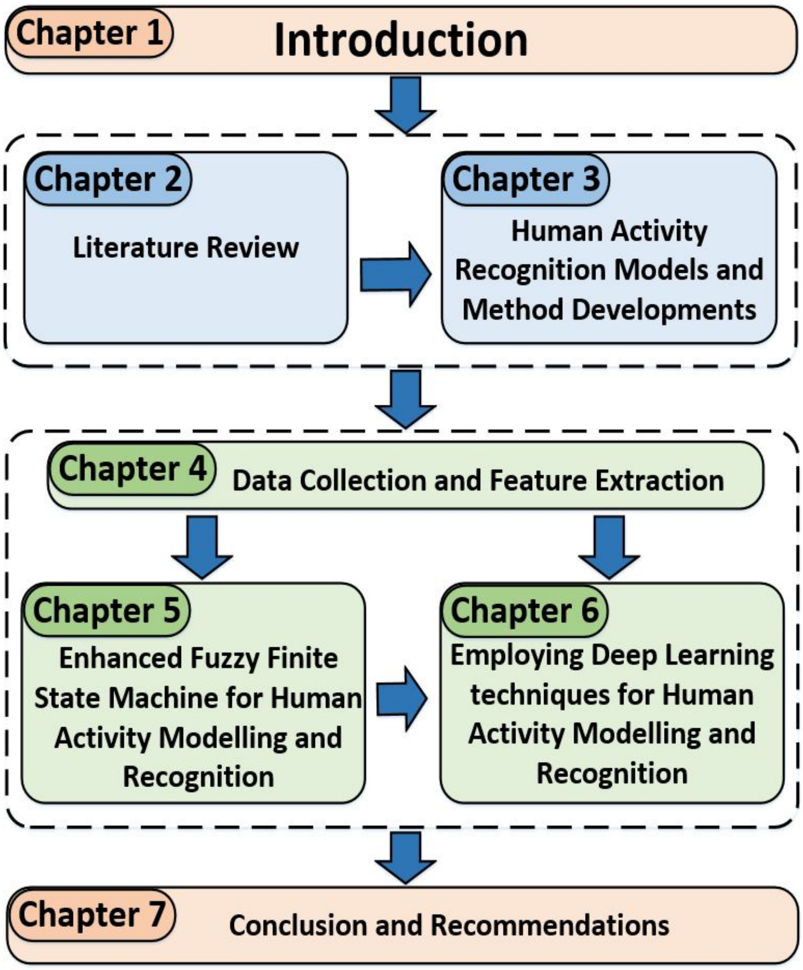



Figure 1.2: Thesis structure showing the organisation of the chapters and their respective dependencies.

## 1.4 Thesis Outline

This thesis consists of seven chapters. Figure 1.2, illustrates the structure of the thesis with an indication of how the chapters are linked. The idea of this 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 reviews the relevant literature works in the field of human activity modelling and recognition. In particular, the literature focuses on the statistical and computational intelligence techniques used for recognising and modelling human activity. The literature gives an overview of the available technologies that are used for modelling and recognising human activities. Also, the review of the previous researches is summarised to identify the research gaps and highlights how this work differs from previous research works. The methodology proposed for human activity modelling and recognition in this thesis is also introduced in this chapter.

**Chapter 3**: Human Activity Recognition Models and Method Developments - This chapter explains some of the existing techniques that are used for human activity modelling and recognition. Some of the presented techniques are used later on in Chapters 5 and 6 to develop models for modelling and recognising human activity for a single user based on binary sensory information. This chapter mainly reviews human activity modelling and recognising models which will be used to propose the enhanced Fuzzy Finite State Machine (FFSM).

**Chapter 4**: Data Collection and Feature Extraction - This chapter gives an overview of sensor intelligent environment for collecting information representing human activity using ambient sensory devices-based. A description of the data collection process from a real-word environment using low-level sensory devices to monitor the Activity of Daily Living (ADL) for a single user is presented. The away of visualising and representing the collected binary sensor information before the features extraction process is explained. The way of representing the extracted features from each activity is discussed.

**Chapter 5**: Enhanced Fuzzy Finite State Machine for Human Activity Modelling and Recognition - This chapter is an extension of the explanation provided in Chapter 3, for developing a Fuzzy Finite State Machine (FFSM) used for modelling and recognising human activity. In this chapter, the FFSM is introduced as a means of defining daily human activities and the transition between the states (here, the activities). Different machine/deep learning techniques are integrated with the existing FFSM. Three different models of FFSM are developed in this work, using NN, Long Short-term Memory (LSTM) recurrent neural network, and CNN. The developed models are referred to as N-FFSM, LSTM-FFSM and CNN-FFSM. Also, experiments are conducted in this chapter using the data sets explained in Chapter 4 to test and evaluate the enhanced FFSM models.

**Chapter 6**: Employing Deep Learning techniques for Human Activity Modelling and Recognition - In this chapter, a DCNN is employed with a fully annotated real data representing ADL for a single user living alone is presented. Data set A, presented in Chapter 4 is used in this chapter with the proposed model. First, the data collected is divided into training and testing sets on day by day basis. Then the activities in the training set are segmented and converted into greyscale activity images. The greyscale image is considered as a 2D image used for representing the numerical measurements such as sensor data. A pre-trained DCNN is employed to extract a unique feature vector from the generated greyscale activity image. The extracted feature vector is used as an input for two different classifiers to classify and recognising the ADL for that user, and then the obtained results are compared. Also, Another deep learning techniques such as Bidirectional Long Short-term Memory (Bi-LSTM) and CNN are employed for modelling and recognising human activities.

**Chapter 7**: Conclusion and Future Work - This chapter presents a summary of the findings of the research conducted in this thesis. The major achievements obtained in this thesis are discussed with a reflection on the research questions identified in Chapter 1. Following the summary of the achievements, the chapter also presents future work recommendations for applications of the work in this thesis. Some possible areas for enhancing the work is suggested for future work.

# Chapter 2

# Literature Review

## 2.1 Introduction

Human Activity Recognition (HAR) is a broad area widely studied in Computational Intelligence (CI). Many efforts are put into developing more suited solutions to enhance the current performance of existing approaches [33, 37]. Therefore, it is essential to review the current literature related to this research topic to justify the intent of the work in this thesis. This chapter provides a review of the relevant literature related to recognising and modelling human activities with a focus on the most common approaches used.

This chapter is structured as follows: Section 2.2 provides an overview of HAR based on binary sensory information of human activity as applied in this work. The related work on HAR including its applications is presented by reviewing the existing literature works as two main classes, which are statistical techniques in Section 2.3, and computational intelligence techniques in Section 2.4. A summary of the presented literature works related to human activity modelling and recognition is discussed in Section 2.5. Section 2.6 follows from the review of previous researches to identify the research gaps and highlights how this work differs from previous research works. The proposed methodology for this work is presented in Section 2.7 followed by Section 2.8 which summarises the information presented in this chapter.

## 2.2 Human Activity Recognition and Modelling

Learning and classification of human activities using CI techniques are known as HAR [38, 39]. Over the last few decades, the existing studies on HAR have been carried out to model, recognise and/or classify human activities. The HAR plays a vital role in many areas including health care, assistive robotics, security, amongst many others [40, 41].

An essential component of any HAR system is that how the information representing human activities is gathered or observed. Based on the information gathered from published papers, the data set representing human activities can be obtained using three different ways; using ambient sensory device-based, using wearable sensory device-based and using camera vision device-based. Observing human activities data using ambient sensory devices [4, 5, 27, 34] makes it easier to obtain information representing human activities within an indoor environment. These approaches utilise sensors to gather information about the environmental conditions and the interaction between the occupier and the environment. This includes recording motion detection, entrance door status (open or close), ambient temperature and light intensity, and locations. A comprehensive review of using ambient sensory device-based approaches for HAR can be found in [39, 42, 43, 44] with more recent researches in [32] and [45]. Obtaining the information representing human activities using systems with wearable sensory devices are undesirable to many users, especially for long-term activity monitoring of older adults as they often forget to wear them. Camera vision device-based systems are mainly used to obtain 2D or 3D information from images or video streams captured using camera devices.

More details about the three methods of collecting the data representing human activities are provided in Table 2.1 below. These details explain the difference between these three data collection approaches in terms of the type of collected data, accuracy, recorded parameter and number of features could be extracted.

In some application, data representing human activity could be collected using a hybrid system which often combined ambient sensors and wearable sensors. This combination of wearable and ambient sensors is of great interest in several

Table 2.1: A summary of the three methods used for collecting dataset representing human activities from real home environments.

| Data collection approach | **Ambient sensory devices-based** | **Wearable sensory devices-based** | **camera vision sensors** |
|---|---|---|---|
| Type of collected data | time-stamped binary data | accelerometer and gyroscope data | images / video stream |
| Collected data accuracy | low, as the collected data only has 1 or 0 values | medium/high | medium/high |
| Measured parameters | only environmental parameters includes movements and doors open and close | environmental and body data includes movements, speed, angle and body temperature | environmental and some body data. |
| Battery limitation | battery has longer life | battery run out quickly | in the most cases, it does not require a battery as it installed by cables which requires more work. |
| long-term data collection process | it can be used for long-term data collection | it is not suitable for long-term data collection process | it is not suitable for long-term data collection process |
| number of features could be extracted | limited number of features | more features could be extracted | more features could be extracted |
| Privacy concern | Low | medium | very high |
| Cost | very cheap | more expensive | very expensive |

applications in human activity modelling and recognition. For instance, when monitoring older adults while deploying interventions to improve balance control and reduce falls, one would be interested in using wearable sensors to track motion and vital signs. Specifically-designed data analysis procedures would then be used to detect falls via movement and vital sign data processing. In this context, ambient sensors could be used in conjunction with wearable sensors to improve the accuracy of falls detection and, most importantly, to enable the detection of falls even when subjects do not wear the sensors.

Many research works are conducted to monitor and analyse the activities of people using many different machine learning and deep learning techniques including Genetic-Fuzzy Finite State Machine (GFFSM) [22], Dynamic Bayesian Network Modelling (DBNM) [46], conventional Neural Networks (NNs) [3, 27, 47], Convolutional Neural Networks (CNNs) [4, 5] and Regression Models (RM) [27]. Most of these studies are employed to enhance a human's lifestyle and well-being by supporting their independent living, especially for older adults and disabled people [10]. As mentioned earlier, information representing human activities within an indoor environment can be obtained based on data collected using ambient sensory devices, wearable sensory devices or camera vision devices. In recent studies, some researchers have collected data using a smartphone's built-in accelerometer [13, 48, 49]. Attention has predominantly been focused on data collected by ambient sensory devices for security and privacy purposes [13].

Modelling and recognising human activities can be achieved by collecting information representing the human activities from an ambient intelligent environment, as such, different statistical and computational intelligence techniques can be employed. The HAR system is integrated into many real-world applications. For example, modelling human activities can be used to monitor the user's behavioural pattern and then detect or predict any deterioration in their health, security and comfort. Recognising the human activities comprises some challenges that may face the researchers; these challenges are addressed below [27, 50]:

1. *Recognising Concurrent or Simultaneous Activities* - By nature, several activities can be undertaken by a single user at the same time [50]. For

example, people can read a book while they are watching TV or eating. In this case, it is not necessary to know which activity started first. That means the existence of concurrent activities when an activity (e.g., eating) starts while the other activity is already started (e.g., reading a book). A specialised approach is required to recognise these non-sequential behaviours.

2. *Recognising Interleaved Activities* - In real life, a certain activity can be interrupted by another activity. For example, while the current activity of an office worker is recorded as "computer activity", a visitor comes to the office. In this case, the first activity is paused for the period of the second activity duration before the previous activity is resumed.

3. *Recognising Multiple Residents* - In many situations, more than one user is present in the environment at the same time. It is even harder than the previous scenarios, to recognise parallel activities for multiple people living/working together in the same place. Different statistical measurements are provided in this research area, but it is still considered as a challenge [10, 27].

There is an increasing amount of literature works for modelling and recognising human activities based on datasets collected from AmI environment. They explore different approaches and algorithms for recognising and modelling human activities. In the next two sections, the existing approaches are divided into two main classes, which are statistical and computational intelligence techniques. Other techniques not included in these two main categories are also reviewed in Section 2.4.5.

## 2.3 Statistical Techniques

Most of the research conducted in the areas of recognising and modelling human activities and future activities prediction is carried out using statistical techniques. These techniques are used to find the relationship between the action and the temporal data that is generated from sensors, and ultimately identify the activity of the user. Several graphical probabilistic-based techniques are applied to recognise and model human activities. The most popular probabilistic-based technique is the Hidden Markov Model (HMM), which has proven to be a good technique to recognise and model human activities. In particular, HMM has the ability to represent random variables, actions and temporal variations within data [26, 28, 45, 51]. In the following sections, most common statistical techniques used for recognising and modelling the human activities are reviewed.

### 2.3.1 Hidden Markov Model

The HMM is one of the most widely used statistical models for recognising and modelling human activities based on sensory data. The model uses some hidden states with Markov process that deal with unknown parameters to model human activities. The hidden states represent the human activities, and the sensor data represent the observable output of the model. Researchers in [2, 28, 51] have investigated different ways in which human behaviour can be detected and modelled using Markov Models (MM) and Hidden Markov Models (HMM). Their experiments are based on the data collected from some wearable sensors and cameras, with a focus on using Hierarchical Context Hidden Markov Model (HC-HMM) from video streams. Evolutionary computing based techniques and other machine learning techniques integrated with HMM are also employed; thus, the accuracy of the system for human activity monitoring is enhanced [13, 49, 52]. Relatively new research by Aicha et. al. [13] presented a new model based on Markov Modulated Poisson Process (MMPP) that promises to come up with a model to represent multi-visitors recognition with more accuracy. Another new Markov Modulated Poisson Process (MMPP) is presented in [53], which proposed a model to represent multi-visitor recognition with more accuracy. The only issue with their approach is the difficulty in processing a large amount of low-level data such as the data gathered from ambient sensory devices.

The research presented in [54], introduces an approach for temporal detection for social interactions. The reported approach at this stage detects intervals of the user or any social activities based on the data extracted from a continuous video stream recorded from an RGB-D camera. The authors developed a computational model for the temporal segmentation of human interactions. The model is working based on the features extracted from the upper body joints of the skeleton including shoulders, head and torso joints. Then, the minimum and maximum values for each feature are used to normalise and pre-formed the extracted features. At the final stage, these features are feed into two standard models; HMM and SVM. The obtained accuracy performance based on this research is 85.56%.

The issue raised while using basic HMM's is the difficulties in dealing with and processing large low-level sensory data (i.e. temporal data from different time scales and different sensors events). Also, in the case where multi-sensors are used, the sequence for each individual activity cannot be separated using these models, so a suitable data mining technique is applied before the HMM.

### 2.3.2 Bayesian Belief Network

Bayesian belief network (known as Bayesian Network) is considered as a statistical method which provides a general framework to recognise and model human activities. This approach is a useful tool to deal with uncertainty and process incomplete data [31, 55]. In [26], the authors proposed a novel way of implementing the task of recognition by using probabilistic graphical models such as Bayesian Network (BN) and Dynamic Bayesian Network (DBN). These techniques are widely used in different domains, including speech recognition and bio-sequence analysis. Furthermore, they have used the proposed DBN to recognise the current pair of activity-object and predict the most probable task based on the features extracted from Red Green Blue-Depth (RGB-D) raw data. This information is then used to make human-robot cooperation more efficient. The major issue encountered with Bayesian belief networks is the inflexibility of exact probabilistic inference [26, 45].

### 2.3.3 Finite State Machine

Finite State Machine (FSM) is one of the statistical techniques that is used to build a state diagram representing different actions. Generally, the transitions between the system's states are triggered by crisp values [18]. In [56, 57, 58], the authors used a classical FSM to recognise and locate the position of a single user in an apartment. The data applied to the FSM in [57] was extracted from video steams.

Many applications have been effectively implemented by using the different types of FSM, such as pattern recognition, transfer learning for human activity, dynamic modelling, and sensitivity analysis[18]. Also, it can be combined with different computational techniques such as econometric models, fuzzy logic systems, self-learning models include a neural network as it goes by filtering information through multiple hidden layers, in a similar way to humans.

As detecting and tracking people within a crowded environment are essential for surveillance and safety, in [58], a new approach for Detecting or tracking people and objects in complex scenarios has been presented using FSM linked with a motion-based image features. The presented approach describes two levels in which activities should be considered as: local activities which are necessary to generalise the detection process; and global activities which are used to detect behaviour patterns that involve not only a single object/people, but also for groups of objects/people in the same environment. Authors highlighted some parameters that must be inferred from the captured video stream for the objects being tracked in the scene, such as speed or direction. The system then takes the initial segmentation to calculate these parameters. Next step is to identify a set of queries that are calculated to detect movement, orientation and location of the objects. The results obtained from this research are promising.

In [59], Authors propose a human activity recognition method based on the FSM model to recognise five different activity using a set of features extracted from a video stream. These five activities are; *Walk through* which means that a person walks through the monitored area without stopping, *Observation* which means that a person repeats the following action pattern, walking-standing-walking, with the large direction variation inside the interested area, *Rest* this activity has two sub-activities (Normal rest-activity and Abnormal rest-activity), *Browse* which means that a person walks to the special object/locution then stop near it, and *Being Idle* which means the current activity is not matched to any of the activities above. In this research, the basic actions for the house users' are extracted and calculated for each person in the interested area at their properties. Then, an activity stream with related features such as movement and the referenced location is recognised using the predefined FSM model. The achieved experimental results show a good recognition accuracy up to 86.96% on average for the five recognised activities.

In [37], a global method based on probabilistic finite-state automata and the definition of the normalised likelihood and perplexity is proposed to manage and identify the ADL for a single user based on a data set collected using ambient sensory devices-based. In order to reduce the computational complexity and to be able to complete ADL recognition, such as eating, dressing, cooking, drinking, and taking medicine some results about a simplified normalised likelihood computation are proved. The research presented in [37], treats two main problems to recognise the ADL as: Activity Discovery (AD) and Activity Recognition (AR). In particular, AD step is a method used to reduce the need for expert knowledge by using learning algorithms to discover activities from the sensor raw data to identify activities. The main need for the AR is to detect the activity actually performed by the inhabitant. When this approach tested with a real data set repressing the ADL for a signal user; it shows the efficiency for identifying and recognising the daily activities.

### 2.3.4 Support Vector Machine

Support Vector Machine (SVM) is commonly used in recognising and modelling human activities to detect and predict users' abnormal activities [60]. In [61], the proposed algorithm enables the recognition of users' activities based on a dataset that is collected from a smartphone. The activity is identified for assessment initially. Then, an SVM classifier is trained to classify the activities. In [7], a novel Robust Parametric Twin Support Vector Machine (RPTWSVM) classifier is proposed in a human activity recognition framework for binary classification problems. The efficacy of the proposed algorithm is tested and evaluated on standard UCI datasets.

### 2.3.5 Other Statistical Models

There exist some other probabilistic models used to recognise and model human activities. For example, the Expectation Maximisation (EM) algorithm reported in [40] is utilised to minimise the uncertainty in the data that represents human activities. In particular, the EM algorithm is applied to the data collected for finding the highest probability of the pixel values in a Gaussian Mixture Model (GMM) and then matching them with the background in the image sequence. In [62], the research proposed came up with a novel approach to build a probabilistic sequential frame to recognise human activities in a video stream. They create a statistical model to extract foreground features from the image background. Then the extracted foreground blobs are used to create the previous and current frames using specific rules. Once the foreground blobs are combined with non-zero pixels, each corresponding blob represents one activity.

In the study introduced in [63], the research group proposed approach consists of a probabilistic ensemble of classifiers as a dynamic mixture model considering the Bayesian probability. Each base classifier contributes to the inference in proportion to its posterior belief. The data used with the proposed approach is gathered from 3D skeleton-based features extracted from RGB-D sensor data are used to recognise daily activities for a single user and the risk situations. The proposed system was tested and evaluated using two different public datasets MSR-Action3D and MSR-Activity3D, and the achieved results were compared with other recent methods. Moreover, the proposed approach was implemented using Robot Operating System (ROS) environment to validate the Dynamic Bayesian Mixture Model (DBMM) running on-the-fly in a mobile robot with an RGB-D sensor on-board to identify the daily activities for a robot-assisted living application.

## 2.4 Computational Intelligence Techniques

Computational intelligence techniques are widely used to recognise and model human activities as an alternative to statistical methods. The following sections summarise some of the computational intelligence techniques that are used for modelling human activities.

### 2.4.1 Neural Networks

Neural Networks (NNs), also known as Artificial Neural Networks (ANNs), are widely used for pattern recognition to process numerical information gathered from sensors in AmI environments [64, 65]. Many different ways are proposed that employ NNs to recognise the ADL of a single user who lives within an intelligent environment. For example, the research presented in [66] used Recurrent NNs (RNN) for detecting the abnormal activities for a single user based on a data set collected using ambient sensory devices-based representing the ADL for a single user. The researchers in this work assumed that by adding an extra layer to the existing design of the network, the ability of the system to detect abnormality is increased. The extra layer is added at the end of the network, where the outputs of the neural network are inputs to this layer.

In [67], a Multi-Layer Perceptron (MLP) neural network is used to identify and monitor movements based on a dataset collected from a Wireless Sensor Network (WSN). For instance, in [68], a complex assistive system with adaptive learning capabilities for human activities is developed using NNs. The researchers use the Internet of Things (IoT) technology with NNs to monitor in real-time a user's current activity in order to be subsequently analysed for a medical diagnosis. In [69], the authors developed a prototype control system called Adaptive Control of Home Environment (ACHE). The developed system is used to monitor an environment (home or office) to understand the user's behavioural patterns and then predict future activities that would be taken by the user. NNs is used for modelling and predicting human activities in [29], where a NN model is combined with an adjustable fuzzy clustering algorithm. This method is experimented using wearable sensor data to test the proposed algorithm's ability to improve the recognition accuracy of user activities.

In general, NNs are useful tools to model and predict time series data. Most of the researches conducted are carried out using feed-forward neural networks which connect input information to the outputs without considering any feedback connections. Therefore, temporal values are not taken into consideration. On the other hand, Recurrent Neural Networks (RNNs) have proven to be a powerful tool to solve the difficulties of the temporal relationships of inputs and outputs at different time steps. In building models for activity recognition, RNN is one of the widely used algorithms due to their ability to learn sequences in data. Long Short-term Memory (LSTM) is a type of RNN applied in most cases when temporal information is considered. In a recent work, LSTM was proposed as a method for modelling human activities with the fusion of LSTM and fuzzy finite state machines. The LSTM algorithm is also used in [70] for learning activity sequences. The algorithm models activities with the advantage of a memory which recalls instances that have occurred previously. Furthermore, an improvement to the LSTM algorithm known as Bi-LSTM was introduced in [71]. This learns data sequences by considering both past and future instances [8, 72, 73].

The work in [65] proposed a sequential Meta-Cognitive learning algorithm for a Neuro-Fuzzy Inference System (McFIS) to develop a classifier for human actions recognition based on video sequences. They used a four-layered NN to determine the number of rules and their corresponding parameters. The motion features are used for each action by extracting the accumulated motion information over a small time window. The results obtained from this work indicate superior performance of the McFIS classifier compared with a standard Support Vector Machine (SVM).

### 2.4.2 Data Mining Techniques

Several data mining techniques are investigated in [74] including some K-Nearest Neighbour (KKN), Lazy Locally-Weighted Learning (LWL) and Multi-Layer Perceptron (MLP). These techniques are used to identify and model the user's activities and then predict future activities based on that. In [75], the authors used Spatio-Temporal Relations to recognise and model everyday activities pattern for a single user. The same Allen's temporal relations are employed with a probabilistic framework to represent local temporal activities in [76]. In general, for particular activities, Allen's temporal relations method encountered a low rate of efficiency.

### 2.4.3 Fuzzy System

Fuzzy systems have proven their efficiency to model and recognise human activities based on vague or uncertain data that are collected from sensor networks in AmI environments [18, 22, 29]. A fuzzy logic system for human gait modelling and recognition is proposed in [22]. The system was enhanced by adding Genetic Algorithms (GAs) to add learning capabilities to it. In [77], a fuzzy logic system is used to recognise user's activities in a home environment based on the data collected from physiological monitoring devices such as cardiac frequency, activity or agitation, posture and fall detection sensor. The target of the research was to monitor users' ADLs to offer them a safe, comfortable and appropriate environment. In [78], a fuzzy predictor model is proposed to create a prediction model to predict a user's future activities within a smart environment. The results presented from work are compared with standard time series prediction models.

Fuzzy systems could be used to increase the efficiency of the classical Finite State Machine (FSM) by proposing Fuzzy Finite State Machines (FFSM) where transitions between states are triggered by the sense of fuzzy inferences instead of using crisp values [22]. This has the advantages of smooth modelling and reasoning with a degree of truth which proves more accurate. Thus, the system can be in more than one state at the time based on the membership value for each state [16][18]. The main advantage of using the fuzzy sense is that it can deal with uncertain data and represent it in more than one state at the same time as membership degrees. In [18], it is shown that human activities can be modelled using an FFSM technique using sequential events based on a dataset collected from a real smart office environment. Although the authors represented the human activities by fuzzy states, they face difficulties in generating fuzzy rules solely based on expert knowledge only.

### 2.4.4 Deep Learning Techniques

Recently, deep learning techniques are widely used in modelling human activities [68, 79, 80]. The most common technique is the Convolutional Neural Network (CNN). For example, in [81], a 3D CNN is employed directly to 3D raw data to extract the input features instead of using handcrafted approaches. The model generates multiple channels of information from the input frames, and the final feature representation combines information from all channels. In [82], the authors proposed a 1D CNN to recognise user's activities such as walking, running, and staying still based on accelerometer data that is collected using a smartphone. The x,y and z accelerometer data are used to learn 1D CNN. Hybrid computational techniques, such as data mining [83], pattern recognition and human activity profiling using Convolutional Neural Network (CNN) [84] are also used in the context of ADL and ADW in order to divide monitored human behaviours into activities and preferences [17]. The CNN is enhanced in [85], by proposing an end-to-end deep network called Tube Convolutional Neural Network (T-CNN). The architecture of the proposed T-CNN is a unified deep network that is used to recognise actions (activity) based on 3D features extracted from a video stream. Many published literature's addresses the issue of modelling human behaviour using wearable sensors [6, 86]. Developing activity recognition systems using a smartphones' built-in accelerometer together with employing CNNs to model the activities are addressed in [87].

In [11], authors investigated methods that could be used for feature representation by creating a binary image based on data collected from Passive Infrared (PIR) sensors. To this end, they proposed a novel technique called Episode Image to generate a binary image based on binary signals from the PIR sensors. Once the binary images had been generated, they were fed into a Deep Convolutional Neural Network (DCNN) classifier where eight features were extracted from each episode to classify the activity travel pattern. The same authors, in [4], report their use of the same feature representation approach with the DCNN classifier for human posture recognition. CNN is considered as one of the recently-used algorithms for human activity recognition and modelling selected due to its ability to learn fruitful features and capture local dependency and spatial information from the given data. In some recent works on human activity recognition [5, 88], researchers have focused on employing the CNN with binary datasets in order to recognise human activities and to detect any abnormal activities in the users' behavioural patterns, based on trained CNN. In [88], the Aruba test-bed data, produced by the Centre for Advanced Studies in Adaptive Systems (CASAS), representing Activities of Daily Living (ADL), were used to recognise human activities and to detect abnormalities in behavioural patterns.

The research study presented in [89], proposed a framework for human daily activity recognition based on depth data. The first step in the proposed framework is to divide each activity into several actions of variable size identified by key poses. Then, different features are extracted from each action window including static and dynamic or temporal features. The features representing the positions of the skeleton are defined as key poses. The static features have projected the distances between two joints and projected angles based on three joints. Dynamic features are used to present velocities of joints coordinate and projected angular velocities. These sets of the extracted features are used to train a Random Forest (RF) classifier. Then, an extension of the RF classifier named the Differential Evolution Random Forest (DERF) algorithm is proposed. The obtained accuracy performance of the proposed human activity recognition framework is a precision of 81.83% and recall of 80.02%.

### 2.4.5 Hybrid Techniques

There exist hybrid modelling and recognising human activities techniques combining more than one of the techniques mention earlier. For example, in [52], the authors presented a novel approach for monitoring people's activities using an indoor localisation system based on Stigmergy technique. They suggested that further work is required to implement the same concepts for enhancing the system's ability to monitor human behaviour. This enhancement can be processed after training the system using a dataset collected from a real environment. In [90], a framework is proposed to integrate temporal and spatial contextual information for determining the wellness of an older adult living alone in a home environment. Swarm intelligence method is used in [52] to monitor an older person's activities via indoor position-based stigmergy. Other evolutionary computing and machine learning techniques based on MMPP are similarly employed to enhance human activity monitoring accuracy. They used a dataset collected using a smartphone's accelerometer [13, 49].

## 2.5 Summary

A summary of the related literature works in the area of human activity modelling and recognition is presented in Table 2.2. The table presented summarises the existing publications in the context of the year of publication, HAR model approaches used and application, dataset name, sensor modality as well the number of activities being recognised and the achieved results performance. Although some of the approaches used for human activity modelling and recognition can be used with all sensors modalities, most of them are only specific to certain types of data that are collected using a certain method of data collection. A description of used models are presented in Table 2.3.

Different types of sensor modalities for collecting data representing human activities are grouped into three aspects: Body-worn sensors, Object sensors, and Ambient sensors. Table 2.4 briefly outlines the sensors modalities mentioned in the literature work summary in Table 2.2.

Table 2.2: A summary of the related research studies for human activity modelling and recognition in the context of year of publication, HAR model approaches used and application, dataset name, sensor modality as well the number of activities being recognised and the achieved results performance.

| Reference | Year of publication | HAR Model | Application | Dataset name | Sensor Modality | Number of Activities | Achieved accuracy performance |
|---|---|---|---|---|---|---|---|
| [91] | 2020 | E-NN | ADL | CASAS | Ambient | 12 | 80.39 % |
| [92] | 2020 | DTEN | Exercise, ADL Physiotherapy | MHEALTH, WISDM, SPAR | Ambient\ Body-worn | 12, 18, 7 | 99.01 % 91 % 99 % |
| [25] | 2019 | MOPPNN | ADL | self-collected | Camera vision, EEg media | - | 99.62 % |
| [93] | 2019 | DMVP | ADL | CASAS | Ambient | 13 | 93.46 % |
| [23] | 2018 | FTW-LSTM | ADL | CASAS | Ambient | 8 | 99.47 % |
| [11] | 2018 | DCNN | ADL | Aruba | Ambient/ object | 8, 10 | 95.10 % 79.00 % |
| [94] | 2018 | DBN | ADL | self-collected | Body-worn/ smartphone | 12 | 95.85 % |
| [7] | 2017 | RPTWSVM | ADL | self-collected | Ambient | 8 | - |
| [64] | 2017 | PHMM, CHMM, CL-HMM, LHMM | multi residant /ADL | CASAS | Ambient | 11 | 74.97 % 82.81 % 85.57 % 89.46 % |
| [95] | 2016 | b-LSTM-S | ADL/ Human Gait | Am-Gia | Ambient/ Body-worn | 2 | 92.70 % |
| [96] | 2016 | CNN, DC-LSTM | ADL | self-collected | Wearable | 16 | 88.30 % 91.70 % |

| Reference | Year of publication | HAR Model | Application | Dataset name | Sensor Modality | Number of Activities | Achieved accuracy performance |
|---|---|---|---|---|---|---|---|
| [97] | 2016 | CNN | ADL/ factory | WISDM | Body-worn | 6 | 89.30 % |
| [98] | 2016 | CNN | ADL | ActiveMiles | Body-worn | 7 | 94.61 % |
| [79] | 2015 | CNN | ADL | ActRecTut | Ambient/ Body-worn | 12 | 85.10 % |
| [99] | 2015 | DBN | ADL | DSADS | Body-worn | - | 83.30 % |
| [100] | 2015 | CNN | ADL | USC-HAD | Body-worn | 12 | 95.18 % |
| [101] | 2015 | CNN | ADL | ActiveMiles | Body-worn/ object | 7 | 94.79 % |
| [102] | 2015 | CNN | ADL | ActiveMiles | Body-worn/ object | 7 | 90.00 % |
| [99] | 2014 | CNN | ADL | ActiveMiles | Body-worn | 6 | 76.80 % |
| [18] | 2014 | FFSM | ADW | self-collected | Ambient | 6 | - |
| [65] | 2012 | McFIS | ADL | self-collected | Camera-vision | 6 | 96.16 % |
| [103] | 2011 | DBN | ADL | Darmstadt Daily Routines | Body-worn | 35 | 73.20 % |

Table 2.3: Different model approaches used for HAR applications.

| HAR Model | Description |
|---|---|
| DNN | Deep fully-connected Neural Network and artificial neural network with deep layers. |
| CNN | Convolutional Neural Network with multiple convolution operations for feature extraction. |
| DC-LSTM | Deep Convolutional Long Short-term Memory. |
| RNN | Recurrent neural network which is a network with time correlations and LSTM. |
| DBN | Deep belief network that is restricted to Boltzmann machine. |
| E-NN | Ensemble Neural Network contains two non-parametric benchmark classifiers. |
| DTEN | Deep Triplet Embeddings includes extracting features from a neural network classifier optimising the embedding learned features. |
| MOPPNN | Multi-object Pattern Producing Neural Network. |
| DMVP | Discovering Methods for Varying Pattern. |
| PHMM | Paralle Hidden Markov Models. |
| CHMM | Coupled Hidden Markov Models. |
| CL-HMM | Combined Label Hidden Markov Models, it meant to deal with the parallel activities and cooperative activities. |
| LHMM | Linked Hidden Markov Models in which activities of all residents are linked at each time step. |
| RPTWSVM | Robust Parametric Twin Support Vector Machine. |

From the review presented in this chapter, modelling and recognising human activity is evidently well-studied area with applications seen in many disciplines. Therefore, the need to further research into solutions to improve current existing systems. Although there have been many successes recorded in modelling and recognising human activity using ambient sensory device-based, the complexities associated with occlusions, varying illuminations, activity data representation, scale variance and activity similarity, remain challenging in many applications. These difficulties have some effects on the computational requirements of many human activity modelling and recognition systems. The conclusions from the

Table 2.4: Description of sensor modalities for collecting data representing human activity in HAR applications.

| Modality | Description | Examples |
|---|---|---|
| Ambient | Applied in an ambient environment to detect users' interaction. | PIR, off/on switch sensor etc. |
| Body-worn | it is a kind of wearable sensors that Worn by the user to describe the body movements. | Wristband sensors, watch, accelerometer, gyroscope and smartphones etc. |
| Object | Attached to objects for capturing objects' movements. | RFID, accelerometer on fridge or cup etc. |

review on the presented literature work are outlined as follows:

- As the information representing human activity is an integral component of human activity modelling and recognition systems, the way of how this information is gathered or observed has an important effect on the system. Therefore, suitable data for human modelling and recognition systems must be obtained as this has a defining impact on the system. Besides, the approach used for the modelling and recognition process should be investigated and selected based on the performance obtained with the information modality and other relevant factors.

- Based on the provided literature works, most of the researches focus on activity classification for a single user. However, activity pattern discovery requires more investigation to provide a better understanding of the nature of activities, including recognising the activities occurring in the same place (e.g. dishwashing and meal preparation).

## 2.6 Research Gap

To gather the activity information from an environment, a camera vision system could be used. However, it is not widely accepted by many users due to some privacy concerns. Also, using wearable sensor-based systems for gathering data representing human activities can be intrusive, and as such may not be the best way to gather data for a long period. Many users often find putting on wearable sensors uncomfortable and may forget to wear them while carrying out normal activities. Furthermore, as human activities differ in nature and sequence of occurrences, it is often difficult to understand the nature of human actions such as the position and location as it requires huge processing using special equipment. This results in limitations in effectively creating models for human activity modelling and recognition. On the other hand, ambient sensory device-based systems offer enough information for representing human activities and they are widely accepted by users as they are not affecting their daily routine. For example, PIR, door, temperature, humidity and mat-pressure sensors can provide a range of temporal information that allows extracting features for high-performance activity modelling and recognition algorithms. An important factor in designing a smart environment for recognising and modelling human activities is that the technology used to gather the human activity data should not interfere with the normal daily activities of the users. Thus, all the devices employed should operate autonomously.

The existing approaches that use data obtained using ambient sensory devices for modelling and recognising human activity mostly focus on the technical aspects (systems ability to recognise activities accurately) [64, 104, 105]. These researches have been directed towards evaluating an algorithm's ability to attain good performances for modelling and recognising human activity. However, most of the approaches used for modelling and recognising human activity are either using computational intelligence techniques or linguistic information assigned by human experts which are not enough for a successful and robust human activity modelling and recognition model. Therefore, the need for integrating human experts' knowledge with the learning capabilities of computational intelligence techniques together for generating a successful and robust model that can be used for modelling and recognising human activities.

To address the identified research gap, this research uses an ambient sensory device-based approach to obtain information for developing a framework capable of human activities modelling and recognition based on the integration of experts' knowledge with available computational intelligence techniques. In the following chapter, the techniques used to develop this framework are described. Also, the methodology used is explained. The proposed methodology for human activity modelling and recognition is presented in the next section.

## 2.7 Proposed Methodology

This research presents a novel human activity modelling and recognition approach to support the independent living lifestyle, especially for older adults based on the binary data gathered using ambient sensory device-based. The proposed approach includes four steps; data collection, data representation and visualisation, feature extraction and representation, and Activity modelling and recognition model.

The first step is to collect a data set representing ADL for a single user by detecting movements, open/close doors, bed/sofa occupied in the monitored environment using ambient sensory device-based. The recorded sensory information is stored as a binary sensor reading containing time-stamp when the sensor is triggered, sensor ID, and sensor status. The sensor ID identifies the location where this sensor is installed (e.g., kitchen, toilet and living room) and the sensor status value is either 1 when detecting an event (e.g., movement, entry door opens/close, or bed/sofa occupied) or 0 otherwise. The process of data collection stage is explained in Chapter 4. The second step is to represent and visualise the gathered binary sensor readings in a proper format that allows for segmenting each activity and extracting features in the next step. Different approaches are currently used for converting the binary sensory readings. However, two novel approaches to convert such binary data into an occupancy signal and 2D greyscale image are employed in this research. More details about this process are explained in Chapter 4. In step three, a process of extracting and representing several features which describe each activity is employed. The

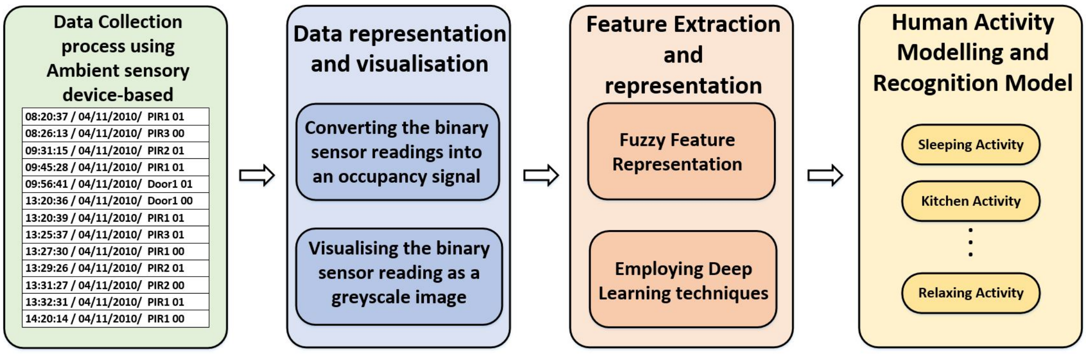



Figure 2.1: Methodology framework of the research presented in this thesis.

features could be extracted and represented using several approaches, includes fuzzy feature representation approach and other deep learning techniques. The extracted features are then used as inputs to the proposed activity modelling and recognition model in step four. An enhanced version of FFSM is proposed in Chapter 5 to be used with the fuzzy represented features. Also, other deep learning techniques are employed for human activity modelling and recognition are conducted in Chapter 6.

To test and evaluate the proposed human activity modelling and recognition approaches, binary sensor readings were recorded in a realistic home environment using an ambient sensory device-based. The data recording process was conducted in a controlled environment, and the participants performed normal day to day activities such as sleeping, toilet, meal preparation, dish washes, relaxing, dining and leaving home either from the front or back doors. The experiments were carried out using two different data sets. The first data set was collected by our research group. The other data set is publicly available datasets. Figure 2.1 shows a diagrammatic representation of the stages of the methodology framework in this research. Different stages include; 1) data collection, 2) representation and visualisation, 3) feature extraction and representation and 4) human activity modelling and recognition model.

## 2.8 Discussion and Research Opportunity

This chapter presented the state-of-the-art of the existing research related to human activity recognition and modelling technologies. The review presented human activity modelling and recognition research works based on ambient sensory device-based information as related to the work in this thesis. Different techniques to model and recognise human activities have been investigated. Although there are still gaps in practical implementations of such systems, its importance cannot be overemphasised. The solutions for developing a robust human activity modelling and recognition system is discussed to meet the needs of enhancing independent lifestyle. A methodology of the work presented in this thesis is presented.

By incorporating this concept, the performance of human activity modelling and recognition systems can be improved using data obtained using ambient sensory devices. From the literature review provided in this chapter, it is obviously seen that the use of simple, low-cost sensors can be used to obtain rich information (which is relevant to any computational system) for activities modelling and recognition. This is investigated in this research. To reiterate the focus of this research, PIR, door entry switch and mat-pressure sensors are used to gather information representing human activities for modelling and recognition of human activities.

# Chapter 3

# Background on Machine Learning Techniques for Human Activity Recognition

## 3.1 Introduction

In this chapter, the existing tools and techniques for human activity modelling and recognition are presented. These tools and techniques are used later on in this thesis to develop the modelling and recognising human activity for a single user within an AmI environment based on binary sensory information. The information about an AmI environment is collected using ambient sensory devices, and they are presented as time-series. The main focus of this chapter is on tools and techniques suitable for integration with Fuzzy Finite State Machine. Hence, more information about Finite State Machine; Fuzzy Logic Systems; Fuzzy Finite State Machine; Neural Networks; Recurrent Neural Networks including Long Short-term Memory and Bidirectional Long Short-term Memory; Convolutional Neural Network; Adaptive Boosting and Fuzzy c-means clustering techniques are provided in the subsequent sections.

To represent human activities in a numerical or analytic model, it is essential to use the correct tools to describe the activities and the link between them. Activities of an occupant in a home environment is a collection of tasks and events which are represented sequentially. Fuzzy Finite State Machine (FFSM) is introduced as a means of defining activities and the transition between activities or states. In this section, a depth explanation about the existing human activity modelling and recognition technologies are provided.

## 3.2 Finite State Machine

Finite State Machine (FSM) is a model to represent sequential processes [18], using for events modelling based on states and state transitions [106]. The FSM is considered as one of a widely used model in the area of the software industry. In particular, it is popular for designing the modelling systems such as embedded systems and control systems includes traffic lights, text parsing and activity recognition [107]. Mainly, there are two different types of the FSM: Mealy state machines and Moore state machines [106, 108]. Typically, FSMs are modelled and implemented as Mealy machines, which are deterministic machines that perform the outputs on their state transitions after the inputs are received [22, 109].

The FSMs are significantly used for understanding the decision making logic as well as for modelling the dynamic data (changing over time) [108]. In the FSM, the outputs which indicate the next state, are a present state and the input function. This means that the selection of the next state mainly depends on the input value (input function) as well as the current state. As in the sequential logic system, the history of the past input is required for deciding the output. Therefore FSM proves very cooperative in understanding sequential logic roles. There are two methods for arranging a sequential logic design, namely Mealy machine as well as Moore machine [110]. In this section, the theory and implementation of FSM are discussed. The definition of a finite state machine, which is also known as finite state automation, is a calculation model that can be executed with the help of hardware and/or software [111]. This is used for creating sequential logic as well as a few computer programs. FSM could be used to solve the problems in fields like mathematics, games, linguistics, artificial intelligence and dynamic data modelling [112].

Mealy state machine consists of two parts, namely combinational logic and

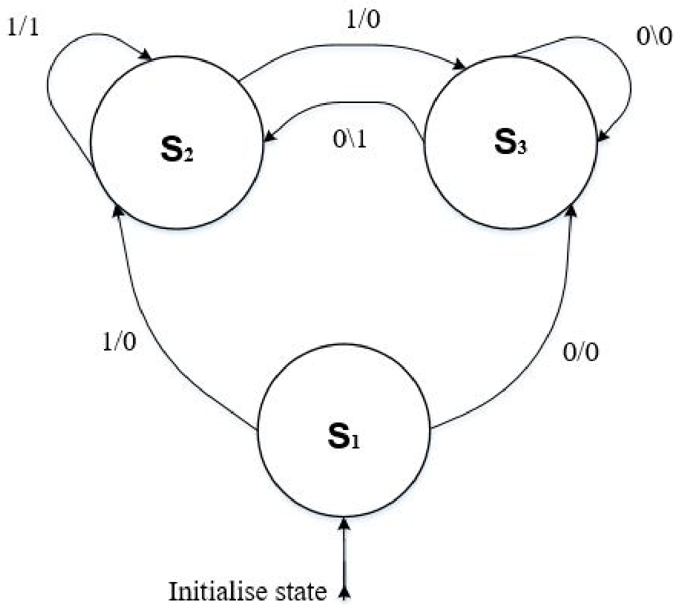



Figure 3.1: An example of a three states finite state machine.

memory [110]. The memory in the system can be used to provide some of the previous outputs as combinational logic inputs. Therefore, based on the current inputs and states, this machine can produce outputs for the next state.

Moore state machine consists of two parts, namely combinational logic and memory as well. But in this case, outputs are a function of current state only, and Output change occurs synchronously with state change [111].

Generally, FSM is represented as a quintuple of $M = (I, O, S, \delta, \lambda)$, where $I, O$, and $S$ represent the set of inputs, outputs and the system states, respectively. $\delta$ is the state transition function $\delta : S \times I \rightarrow S$. $\lambda$ is the output function $\lambda : S \times I \rightarrow O$. The transitions between the model's states are based on crisp values that are controlled depending on the values of the inputs and the current state of the machine [94]. There are two different possibilities for the transition coming from each state; to remain in the same state or transmit to the next state [113].

To implement an FSM, it is essential to define the states and transitions. Figure 3.1, shows a simple design of a FSM with 3 states representing as $S_1$, $S_2$ and $S_3$. The initialise starting state is $S_1$, which will remain active until the condition of transmission is true. When that happens, the current state is transitioned to $S_2$ or $S_3$, which remains active until the condition parameters change. Since there are transitions connecting between the three created states, always the model can transit from the active state (current state) to any other state (next state) based on the transition conditions (rules).

The FSM is designed in such a way that it can model the sequence of activities. If there are uncertainties associated with states or the transition between states, the FSM needs to enhanced with new capabilities. One possible tool which can handle uncertainties is Fuzzy Logic, and in the next two sections, more details are provided in this regard.

## 3.3 Fuzzy Logic Systems

Fuzzy Logic Systems (FLS) can efficiently model uncertain numerical data. It can be used to represent the user's activities as a linguistic variable to make it more understandable for users. The FLS is widely used in many applications including; control systems, image processing, pattern recognition and video games [25, 72, 114, 115]. The FLS is a computational structure that is based on the theory of fuzzy sets, **if-then** rules and fuzzy logic. A FLS consists of three conceptual components, as follows:

- A rule base including all fuzzy rules for the decision making.
- A database where all the Membership Functions (MF) associated with the inputs and the outputs are defined. This also includes all terms used in fuzzy rules and linguistic variables of the fuzzy system.
- An inference engine, where deriving conclusions using the given rules and parameters to infer a relevant output.

The rule base and the database are used to form the knowledge base for the entire system. Figure 3.2, shows the general structure of a FLS consists of a knowledge base (rule base and database) and an inference machine. It operates as described in the following sequence:

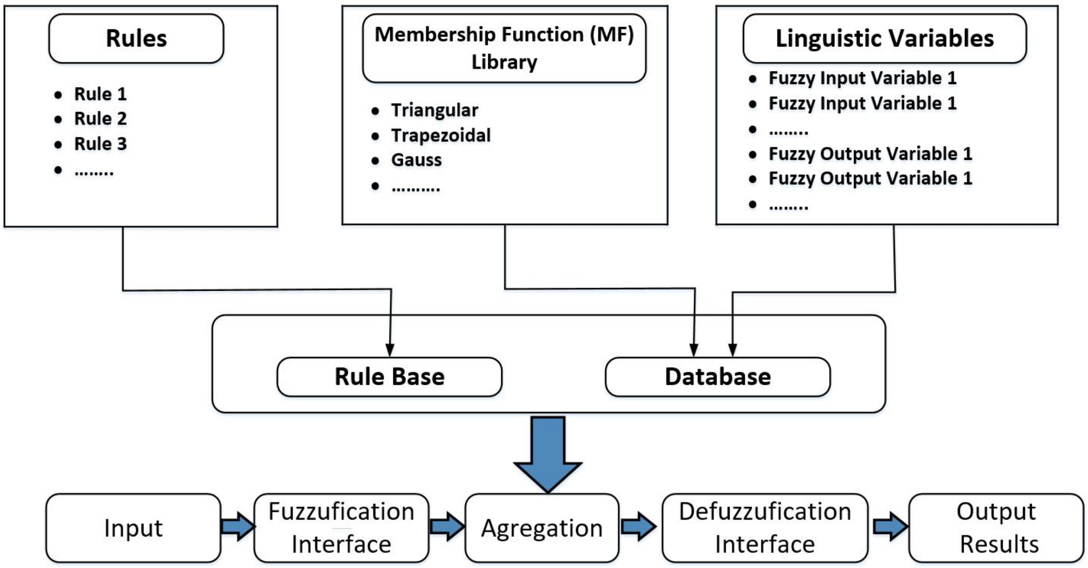



Figure 3.2: An schematic diagram representing the structure of a fuzzy logic system.

- The input data are fuzzified first, in order to obtain membership degrees for each of the parameters of the input variables.
- The knowledge base using the inference machine applies the aggregation rules. Thus, membership degrees to the terms of output variables are calculated.
- The defuzzification process is applied in order to obtain the output result.

Considering the explanations provided in the earlier section about FSM and the information provided above about FLS, a combination of these two techniques could be used to come up with an enhanced version of the FSM. Enhanced FSM is proposed due to the shortcomings in the classical FSM.

## 3.4 Fuzzy Finite State Machine

Fuzzy Finite State Machine (FFSM) is an extended version of the classical FSM. The FSM can be represented as a model made of two or more states; each state represents one event from a sequence of events in a dynamic process. Only one single state of this model can be active at a time. The model is moved from one state to another by triggering crisp values. In human activity recognition and modelling, a user may be associated with multiple states. This would require to be quantified with a degree of belonging (degree of fuzziness). Once the fuzziness aspect is added to the state transitions in the classical FSM, the transitions are not triggered based on crisp values, but using fuzzy variables [18, 22, 24]. This implies that the current *activated* state of the model is not necessarily one state, but it could be more than one state at any given time with belonging degrees [24].

In an FFSM, the system's states are represented as a set of linguistic variables $S(t) = [s_1(t), s_2(t), ..., s_i(t), ..., s_N(t)]$ where $N$ is the number of states. For a non-sequential system at a time $t$, the system's states are represented as a state vector $S(t)$. When the system evolves in time, the next state is represented as a vector $S(t+1)$.

In general, as in [18, 22], the FFSM is defined as a tuple of parameters $\big(S(t), U(t), f, Y(t), g\big)$. where;

- *Fuzzy State*, $S(t) = [s_1(t), s_2(t), ..., s_i(t), ..., s_N(t)]$ is presenting a vector identifying the system's states at time $t$ and $N$ is the number of states. Each individual state at time $t$ is $s_i(t); i = 1...N$ is a numerical value that is in fact the membership grade (between 0 and 1) given to each linguistic variable $s_i(t)$ within the set of FFSM's states $S(t)$.

- *Input Vector*, $U(t) = [u_1(t), u_2(t), ..., u_j(t), ...., u_P(t)]$ is the input vector at time $t$ representing the associated value to the linguistic variables that are generally obtained after a fuzzification process for the input data. $P$ are the number of input variables. This input data could be a sensors' data, a combination of different signals, or any other calculation to numerical data. The fuzzification process that is designed based on experts' view to translating the numerical input values to a set of membership grades given to each linguistic label that defines all the acceptable values in the input vector. The labels that are associated with the input $u_j(t)$ is represented as $A_{u_j} = \{A^1_{u_j}, A^2_{u_j}, ..., A^M_{u_j}\}$, where $M$ is the number of the associated linguistic labels [16].

- *Transition Function*, $f$ is the state transition function that is mainly used to calculate the next state vector $S(t+1)$, at each time instant $t$. The transition function $f$ controls the allowed transitions between the defined system's states. Also, it is implemented as a set of fuzzy rules. There are different ways to define the rules, e.g., using the human expert knowledge or learning from the numerical input-output data by applying machine learning algorithms such as Artificial Neural Network (ANN) and Genetic Algorithm (GA) [29, 30, 116]. A combination of these approaches can also be implemented to have one framework containing the rules that are generated by learning from the numerical data and those assigned by the human experts' knowledge [116].

- *Output Vector*, $Y(t) = [y_1(t), y_2(t), ..., y_k(t), ..., y_Q(t)]$ is the output vector consisting of crisp values associated to each output at the time $t$ and $Q$ is the number of output variables. Values in the output vector $Y(t)$ are calculated based on the current state of the system $S(t)$ and the input vector $U(t)$.

- *Output Function*, $g$ is the output function that is used to calculate the value of output vector $Y(t)$, at each time instant $t$.

The states and outputs of the time-invariant FFSM [16, 22] are expressed as:

$$S(t+1) = f\big(S(t), U(t)\big) \tag{3.1}$$

$$Y(t) = g\big(S(t), U(t)\big) \tag{3.2}$$

The states' transition mechanism between two exemplary states $s_m$ and $s_n$ in the FFSM is illustrated in Figure 3.3. Considering the complexity of modelling a large scale dataset, it may be impossible to analytically identify the functions

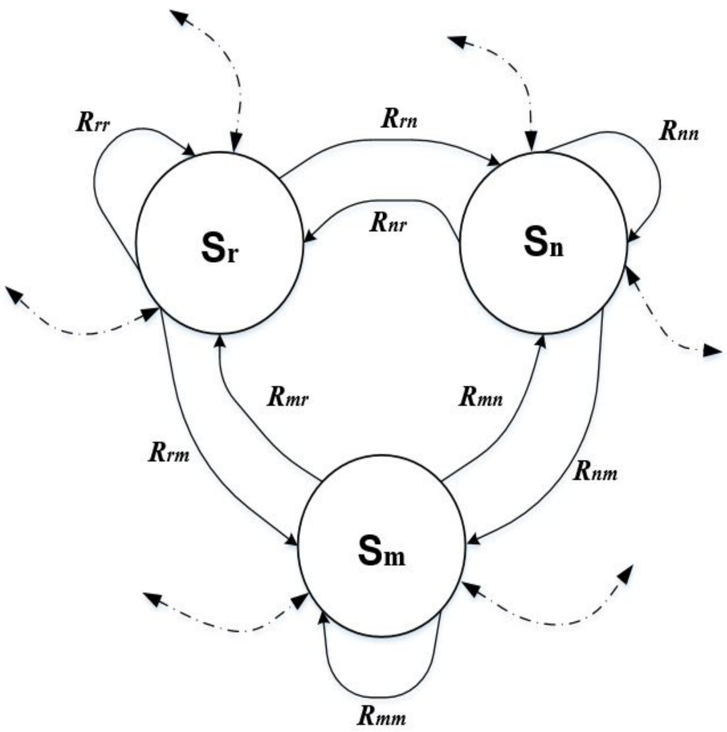



Figure 3.3: An illustration of states' transition diagram of a fuzzy finite state machine.

$f$ and $g$. This complexity will be even harder when it is used for time-invariant models [18, 22]. Therefore, a rule $R_{mn}$ is used to establish the relationship between states $s_m$ and $s_n$. These states' transitions can be expressed as a general fuzzy rule format [24], as follows:

$$R^{\lambda}_{mn}\textbf{: IF } (S(t) \textbf{ is } s_m) \textbf{ AND } H_{mn} \textbf{ THEN}$$
$$S(t+1) \textbf{ is } s_n \qquad \lambda\mathbf{=1,\ldots,}\Lambda$$

where the fuzzy rule has the following parts:

- *The Antecedent Part*: is a combination of two terms; the first term, ($S(t)$ is $s_m$) is used to determine if the state $s_m$ is an *activated* state in time instant $t$. Therefore, the system can change from state $s_m$ to state $s_n$ or remains in state $s_m$, only if $m = n$, the second term of the antecedent part is $H_{mn}$ which represents all constraints imposed on the input variables that are required to either remain in state $s_m$ (when, $m = n$) or change to state $s_n$, e.g., $H_{mn} = (u_1(t)$ is $A^3_{u1})$ **AND** $(u_2(t)$ is $A^4_{u2}$ **OR** $A^2_{u2})$.

- *The Consequent Part*: $(S(t+1)$ is $s_n)$ is the **THEN** part of the fuzzy rule, which determines the next value of the state vector $S(t+1)$ for being in state $s_n$. The linguistic variables of the consequent are considered as being singletons, i.e. all elements of the $S(t)$ vector are zero, except for the $m^{th}$ element which is 1 [22].

For a rule-base consisting of $\Lambda$ rules, the next value of the state vector $S(t+1)$ is the weighted average utilising the firing degree of each rule, defined as:

$$S(t+1) = \frac{\sum_{\lambda=1}^{\Lambda} w_\lambda . S(t)}{\sum_{\lambda=1}^{\Lambda} w_\lambda} \quad if \sum_{\lambda=1}^{\Lambda} w_\lambda \neq 0 \tag{3.3}$$

$$S(t+1) = S(t) \qquad if \sum_{\lambda=1}^{\Lambda} w_\lambda = 0 \tag{3.4}$$

More details about the transition function and how they can be enhanced are explained in Chapter 5.

## 3.5 Neural Networks

Neural Networks (NNs) have been employed widely for pattern recognition and particularly for human activity modelling and recognition due to its capability of learning from the input data and create a network model [117]. The network model is trained to learn the relationships between inputs and outputs date based on some samples of the input data and the target required outputs (classes) during the training mode. The trained network model can be applied on new data which was not previously exposed to the network for classification and recognition purposes [29, 68]. The NN's models consist of a collection of inputs, output, and processing units called nodes or neurons [118]. In a general architecture of the NN models, the nodes are organised into three layers: an input layer, where the input data will be applied to the networking model; hidden layer, where all of the computational operations are performed; and output layer, where the final

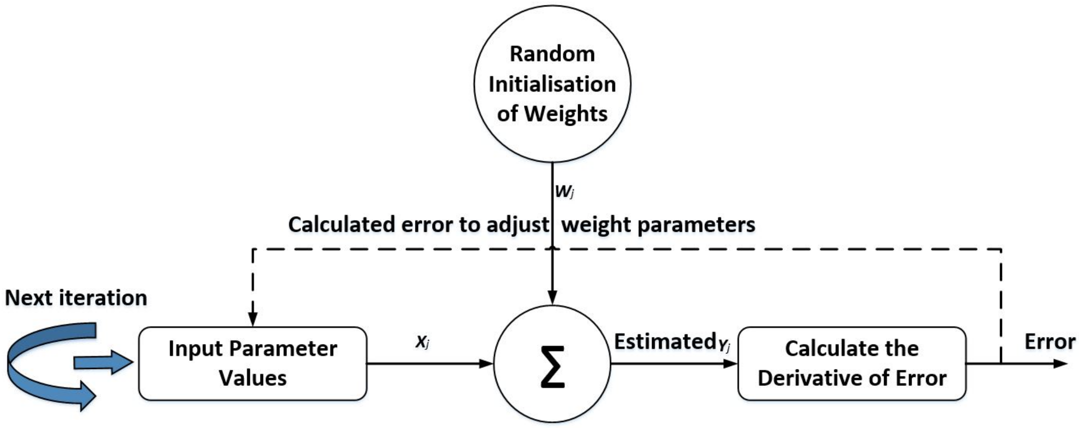



Figure 3.4: The learning process of weight adjustment for a single node or neuron of the neural network.

outputs are obtained. Each node or neuron performs the simple operation to process these inputs and produce the output, and then the output of this neuron is forwarded to the next neuron in the next layer [25].

Each node contains an activation function $\phi(W - B)$, where $W$ are the weighted summation of the given inputs and $B$ is the bias value. In most cases, weights are initialised to a small random value that will be updated during the training mode. The weighted summation $W$ is given from inputs $1, 2, 3, .., n$, where $n$ is the last value in the input vectors, and it will be calculated as shown in Equation 3.5. A number of input vectors are provided to the algorithm through the input layer in order to determine the corresponding desired output (classes) in the output layer. In each iteration and when the input data is presented to the system, there is an error created at the output layer. This error represents the difference in values between the real system output (during the training mode) and the desired response value (classes). This error will be used to feed into the network model to adjust the weights in the next iteration during the training mode.

$$W = \sum_{i=1}^{n} weight_i \times input_i \tag{3.5}$$

Before the training mode to learn the relations between inputs and outputs, the given data set is divided into two sets; training data set for the learning phase, and testing data set for testing and evaluation the model. During the learning phase, a set of training input vectors are presented at the input layers, including feature vectors (input samples) and their corresponding desired output (output classes) vectors. Initially, small weight values are assigned randomly to the hidden layer's nodes. Then the NN model adjusts the assigned weights according to the difference between the network's actual outputs and the desired output corresponding to this input vector. A sample scenario of the learning process of a single node or neuron is illustrated in Figure 3.4. Every single node in the NN model is represented as illustrated in Equation 3.6:

$$Y_j = f_j \sum W_{ij} \times x_i \tag{3.6}$$

where, $x_i$ is the input data applied to the NN model through the input layer, $W_{ij}$ is the difference of weights between $i^{th}$ node of the previous layer and $j^{th}$ node of the current layer, and $f_j$ represents the activation function in that node. Several types of activation functions are around within the neural network domains including; linear, sigmoid and hyperbolic tangent [29].

The NN used in this thesis is based on the principles of the Multi-Layer Perceptron (MLP) network with Backward Propagation (BP) learning algorithm because it is easy to train and for its high-performance accuracy. The MLP-BP is mainly used in this work to enhance the capabilities of the proposed FFSM for modelling and recognising human activity due to its abilities to learn the relations between the input and output data. The process of the BP can be divided into two main operations, feed-forward and back-propagation operations. During the feed-forward operation, the input features are fed to the input nodes through the input layer. Once this operation ends, the back-propagation operation starts to organise the output by adjusting the weights of the NN model.

The learning capability of the NNs is used to update the state transition parameters of the FFSM, hence enhance the performance of the model. The details of the enhance system are provided in Chapter 5. Moreover, the learning abilities of different techniques, such as RNN could be used to enhance the performance of the model.

## 3.6 Recurrent Neural Networks

A Recurrent Neural Network (RNN) is a class of ANN which includes built-in loops that allow previous information to be incorporated in the network. It has a feedback connection to the network itself, which allows activation functions to flow back in a loop to learn sequences and information to persist [119]. Hence, the network exhibit temporal dynamic behaviour. Figure 3.5 illustrates a general architecture of the RNN with two layers. The RNNs usually have an input layer, some hidden layers and an output layer, with some context units connected to the input layer [120]. In this model, the Multi-Layer Perceptron network is used to compute $y(t)$ and it could be formulated as:

$$y(t) = f(x(t-1), .., x(t-D_x), y(t-2), .., y(t-D_y)) \tag{3.7}$$

where $x(t)$ and $y(t)$ are the input and the output of the network at time step $t$, respectively. $D_x$ and $D_y$ are the input and the output memory information orders with $D_x \geq 1, D_y \geq 1 and D_y \geq D_x$. The used function for computing the input and output of the network is expressed as $f$. The output results $y(t)$ is connected to the input layer to be re-used as $x(t-1)$ with the input, and $y(t-1)$ with the output value.

In the domain of modelling sequential data, RNNs are considered as an extremely powerful tool [121]. RNNs could be employed for solving temporal recognition and classification problems with sequential and non-sequential applications such as video captioning, word prediction, word translation, image processing, speech recognition, speech processing [122]. It is well known that NN training algorithms are widely used for many temporal classification problems. However, to achieve high accuracy during the learning process, several computational temporal techniques are required [122]. Therefore, different types of RNN are proposed for learning the relations between the input temporal data and their corresponding classes. Different types of RNN architectures up to now are available such as Long Short-term Memory (LSTM), Bidirectional Long Short-term Memory (Bi-LSTM), Recursive Neural network (ReNN) and Echo State Network (ESN) [119].

In the next section, LSTM and Bi-LSTM networks are explained in details,

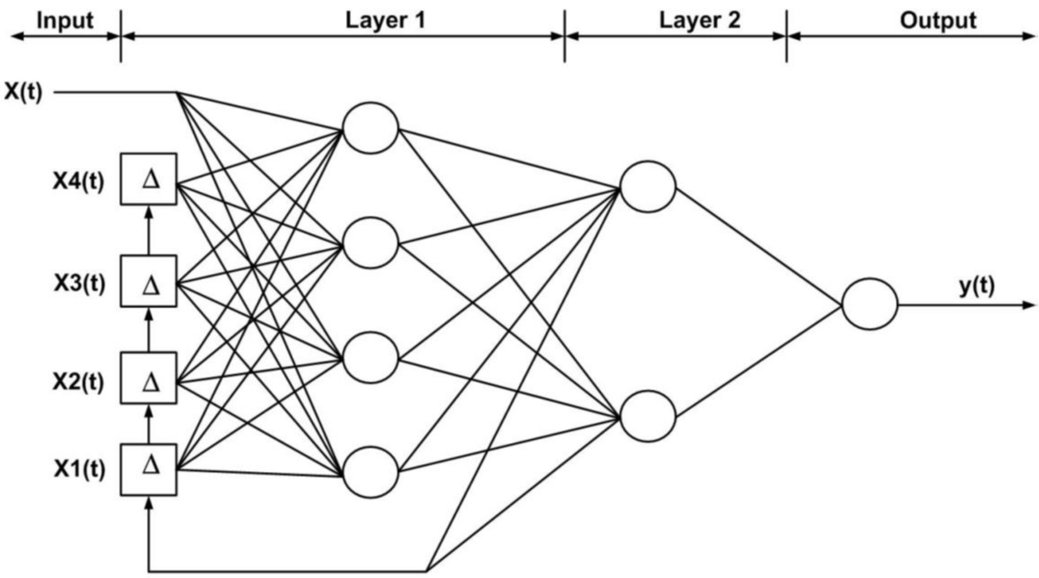



Figure 3.5: A general architecture of Recurrent Neural Network with two layers.

which considered as one of the most widely used RNN.

### 3.6.1 Long Short-Term Memory

A Long Short-Term Memory (LSTM) is a particular kind of RNNs designed to solve vanishing and gradients problems in the classical RNNs [72, 123]. LSTM is a powerful tool for sequential learning tasks that are represented as temporal data. It can also remember previous information for long periods. These characteristics make LSTM especially useful for temporary data classification problems. LSTM cell consists of three gating mechanisms to provide the ability to remove or add information to the memory cell. These three gates are used to regulate the impact of the input through the input gate, the previous cell state through the forget gate and the output through the output gate. The essential gate in the LSTM cell is the forget gate as it decides the information that is going to be remembered or forgotten from the previous states.

The state equations for each LSTM cell as they defined in [72, 123] are formulated as follows:

$$f_t = \sigma(w_{xf}.x_t + w_{hf}.h_{t-1} + b_f) \tag{3.8}$$

$$i_t = \sigma(w_{xi}.x_t + w_{hi}.h_{t-1} + b_i) \tag{3.9}$$

$$v_t = tanh(w_{xv}.x_t + w_{hv}.h_{t-1} + b_v) \tag{3.10}$$

$$o_t = \sigma(w_{xo}.x_t + w_{ho}.h_{t-1} + b_o) \tag{3.11}$$

$$C_t = f_t.C_{t-1} + i_t.v_t \tag{3.12}$$

$$y_t = o_t.tanh(C_t) \tag{3.13}$$

where, $f, i$ and $o$ denote the forget, input and output gates. $C$ and $y$ are the cell state and output respectively at time $t$. $b_j$, $j \in \{f, i, o, v\}$ are the bias units for the forget, input and output gates and the remembering inputs. $w_{ij}$ is the weight connection between $i$ and $j$. $\sigma$ is a logistic sigmoid function and $tanh$ is the tangent hyperbolic function.

LSTM learns the temporal relations in the given data by storing the information through multiple time steps. In Chapter 5, the proposed approach is introduced, which explains how the learning abilities of the LSTM in the temporal data is integrated with the sequential events modelling in the FFSM.

### 3.6.2 Bidirectional Long Short-Term Memory Neural Networks

Neural Networks are shown to be a good model to represent sequential forms of data such as the time series data [72]. The LSTM has the ability to store past information by looping inside its architecture, which keeps the information from the previously learned iteration. The LSTM's architecture contains; input gate to handle the input information in the input layer, output gate in the output layer which has a Softmax activation on it, and a forget gate that controls the internal cells to either save or forget the previous information. The main purpose of using the forget gate is to reduce the risk of vanishing gradients. As the LSTM can only get the information from the previous context, a further improvement was made by fusing another LSTM network to give the Bidirectional LSTM (Bi-LSTM) model. The Bi-LSTM model is a combination of two - forward and backwards - LSTMs, which can handle information from both future and past directions.

The Bi-LSTM is a combination of two directional LSTM's, the input data will be processed firstly in forward direction ($\rightarrow$) and then re-processed again in backward direction ($\leftarrow$). Bi-LSTM's consists of input gate ($i_t$), forget gate ($f_t$), output gate ($o_t$), memory cell ($c_t$) and the hidden state ($h_t$) for both forward ($\rightarrow$) and backward ($\leftarrow$) directions at each time step ($t$). Therefore the hidden state in the Bi-LSTM is calculated based on the combination of the forward hidden state ($h_t^{\rightarrow}$) and the backward hidden state ($h_t^{\leftarrow}$) as formulated in equation 3.14.

$$h_t = h_t^{\rightarrow} \oplus h_t^{\leftarrow} \tag{3.14}$$

where the elements of the forward state ($\rightarrow$) are formulated as follows:

$$h_t^{\rightarrow} = o_t^{\rightarrow} \otimes tanh(c_t^{\rightarrow}) \tag{3.15}$$

$$c_t^{\rightarrow} = f_t^{\rightarrow} \otimes c_{t-1}^{\rightarrow} + i_t^{\rightarrow} \otimes tanh(W_c^{\rightarrow} U_t^{\rightarrow} + V_c^{\rightarrow} h_{t-1}^{\rightarrow} + b_c^{\rightarrow}) \tag{3.16}$$

$$i_t^{\rightarrow} = \sigma(W_i^{\rightarrow} U_t^{\rightarrow} + V_i^{\rightarrow} h_{t-1}^{\rightarrow} + b_i^{\rightarrow}) \tag{3.17}$$

$$f_t^{\rightarrow} = \sigma(W_f^{\rightarrow} U_t^{\rightarrow} + V_f^{\rightarrow} h_{t-1}^{\rightarrow} + b_f^{\rightarrow}) \tag{3.18}$$

$$o_t^{\rightarrow} = \sigma(W_o^{\rightarrow} U_t^{\rightarrow} + V_o^{\rightarrow} h_{t-1}^{\rightarrow} + b_o^{\rightarrow}) \tag{3.19}$$

on the other hand, the elements of the backward state ($\leftarrow$) are formulated as follows:

$$h_t^{\leftarrow} = o_t^{\leftarrow} \otimes tanh(c_t^{\leftarrow}) \tag{3.20}$$

$$c_t^{\leftarrow} = f_t^{\leftarrow} \otimes c_{t-1}^{\leftarrow} + i_t^{\leftarrow} \otimes tanh(W_c^{\leftarrow} U_t^{\leftarrow} + V_c^{\leftarrow} h_{t-1}^{\leftarrow} + b_c^{\leftarrow}) \tag{3.21}$$

$$i_t^{\leftarrow} = \sigma(W_i^{\leftarrow} U_t^{\leftarrow} + V_i^{\leftarrow} h_{t-1}^{\leftarrow} + b_i^{\leftarrow}) \tag{3.22}$$

$$f_t^{\leftarrow} = \sigma(W_f^{\leftarrow} U_t^{\leftarrow} + V_f^{\leftarrow} h_{t-1}^{\leftarrow} + b_f^{\leftarrow}) \tag{3.23}$$

$$o_t^{\leftarrow} = \sigma(W_o^{\leftarrow} U_t^{\leftarrow} + V_o^{\leftarrow} h_{t-1}^{\leftarrow} + b_o^{\leftarrow}) \tag{3.24}$$

where, $U_t$ is the input data at time step $t$; $U \in \{u_1, u_2, .., u_k\}$. $i, f$ and $o$ denote the input, forget and output gates respectively. $c$ and $y$ are the cell state and output respectively at time step $t$. $b_j$, $j \in \{f, i, o, c\}$ are the bias units for the forget, input and output gates and the memory inputs respectively. $W_{ij}$ is the weight connection between $i$ and $j$. $\sigma$ is a logistic sigmoid function and $tanh$ is the tangent hyperbolic function. $V$ is the cell matrix.

The Bi-LSTM is used to learn the sequential relations in the given data by storing the information through multiple time steps using the combination of forward and backward states. The next section introduces experiments with the proposed approach, which integrates the learning abilities of the Bi-LSTM model in the temporal data with sequential human activities.

The learning abilities of different deep learning techniques such as CNN could be used to enhance the performance of the model due to its capabilities for pattern recognition based on a complex dataset. The following section explains the framework architecture of CNN.

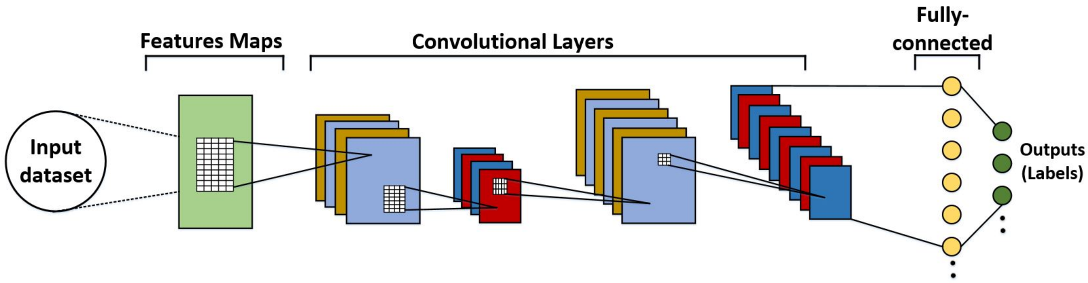



Figure 3.6: The framework architecture of the Convolutional Neural Network.

## 3.7 Convolutional Neural Network

Convolutional Neural Networks (CNNs) are a category of ANNs that have proven very effective in areas such as image recognition and classification. A general framework architecture of the CNN, containing two conventional layers or more, dependent on the complexity of the input data, and one fully-connected layer, is shown in Fig 3.6. Each one of these convolutional layers has multiple feature filters for accurately optimising values during the training phase. Additionally, each convolutional layer is followed by a max-pooling layer that has a window of a specific size to ensure that the outputs from each conventional layer are smaller than the inputs. The fully-connected layer in this architecture is a traditional MLP that contains a softmax activation function for the output layer. By operating the softmax activation function in the output layer, CNN can be used to recognise and classify the input features into various classes based on relations learned during the training stage [88]. In a scenario where complex input data is expected, the CNN architecture may be modified to contain more than two of the convolutional and max-pooling layers with different sizes of border filters to process such data [4]. Also, more than one fully-connected layer could be considered after the top convolutional layer.

The input data for a CNN is formulated as a matrix $c$ in dimensions $h \times w \times d$, where $h$ is the matrix height, which will be represented as the total number of activity. $w$ is the matrix width, which is the number of fuzzy set for each input variable. $d$ is the number of channels in the input matrix $q$ [5]. In cases where the input window has only one class, the number of channels $d$ is 1. In the training process of the CNN, the standard forward and backward propagation algorithms are employed to select the values of the CNN parameters. The selected features are mapped by the convolutional operator as follows [88]:

$$V_t = \frac{1}{1 + exp(d_\eta + \sum_\iota \kappa_{\iota\eta} \vartheta x_\iota)} \tag{3.25}$$

where $\vartheta$ is the convolutional operator, $\kappa_{\iota\eta}$ is the convolutional filter for the $\iota$-th input, $V_t$ is the generated $\eta$-th output feature map, which is achieved by selecting the most effective features over the non-overlapping pooling regions from the input data $x_\iota$, and $d_\eta$ denotes the bias.

Different CI techniques such as Adaptive Boosting could also be used for creating a robust modelling and recognition model for human activity due to its capabilities for pattern recognition based on a complex dataset. The following section explains the framework architecture of Adaptive Boosting.

## 3.8 Adaptive Boosting

The Adaptive Boosting (AdaBoost) classifier is based on a combination of different weak classifiers. One of the advantages of the AdaBoost is that it uses the features of each class separately when the use of the whole features may not be necessary at the time [124]. Also, it adjusts the errors of the weak classifiers, allowing the achievement of high accuracy with far less tuning of other powerful classifiers [125]. An extension of the original AdaBoost algorithm which is used to create a strong classifier from a number of weak classifiers. The improved algorithm is referred to as AdaBoost-M2 algorithm. The pseudo-code of AdaBoost-M2 algorithm is provided Algorithm 1. The weights of the training samples in the first iteration are equal. In the next iteration, the incorrect samples that were inaccurately classified in the previous round carry more weight than the correctly classified ones, to force the weak learner to focus on the hard samples in the training set.

The idea behind improving the performance of decision trees in the AdaBoost technique by converting several weak classifiers to one single strong learner has achieved the best accuracy performance. The decision tree is a type of supervised

**Algorithm 1** AdaBoost-M2: An extension of the original AdaBoost algorithm.

**Input:**

- A training set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ with samples $x_n \in X = \{1, 2, \ldots, p\}$ and labels $y_n \in Y = \{1, 2, \ldots, k\}$

**Initialisation:**

- Define the number of iterations **T**.
- Initialise all weights for the training samples $w_1(i) = \frac{1}{n}$.
- Weak learning **WeakLearn**.

**for** $t = 1, \ldots, T$ **do**

1. Select classifier $h_t : R^d \rightarrow \{-1, +1\}$ using distribution $w_t$.
2. Set error $\varepsilon = \sum_{t=1}^{N} w_t(i)[h_t \neq (x)]$.
3. Choose $\alpha_t = \frac{1}{2}(ln\frac{1-\varepsilon_i}{\varepsilon_i})$.
4. Call **WeakLearn** for providing it in the distribution.
5. Update weight distribution over the whole samples:

$$\omega_{t+1}(i) = \frac{\omega_t(i)exp(-\alpha_t h_t(x_t))}{Z_t}.$$

where $Z_t = \sum_{t=1}^{n} \omega_t(i)exp(-\alpha_t Y_t(x_t))$ is the normalisation factor.

**Output:**

- Combined classifiers: $H_T(x) = sign(\sum_{t=1}^{T} \alpha_t \times h_t(x))$

**end for**

machine learning that works by specifying the inputs and the desire corresponding outputs in the training data set. Then the training data set is continuously split according to a certain parameter to learn the relations between the inputs and their corresponding output by adjusting the weights during the training mode.

As human behaves in an unpredictable way, different CI techniques such as Fuzzy C-means could also be used for creating a robust for human activity modelling and recognition model by adding a fuzziness ability to overcome the uncertainties in the data representing the human activity.

## 3.9 Fuzzy C-Means Clustering

It is considered as a data-driven approach that is the fuzzy extension of the k-means algorithm for creating fuzzy partitions for each class in the input data [12, 126]. After the partition operation is calculated, each feature has a degree of membership in the associated fuzzy partitions, which results in a fuzzy version of the original numerical feature vector (input). This operation eliminates any loss of information that arises due to sharp partition boundaries and prepares the dataset for performing fuzzy associative classification [127]. $X$ is the number of input samples, $X = \{x_1, x_2, \ldots, x_p\}$, and $X_i$ is the number of features of the $i^{th}$ samples $X_i = \{x_{i1}, x_{i2}, \ldots, x_{iq}\}$. So, the objective function is given as:

$$J_m = \sum_{i=1}^{p} \sum_{k=1}^{c} \sum_{j=1}^{q} u_{ik}^{m} \|x_{ij} - v_{kj}\|^2, 1 \leq m \leq \infty \tag{3.26}$$

where $m > 1$ is the fuzzification parameter, $v_{kj}$ is the centre of the $j^{th}$ feature in the $k^{th}$ cluster and $u_{ik}$ is the grade of membership of the $i^{th}$ observation in the $k^{th}$ group.

The fuzzy clustering could be performed by optimising the objective function shown in Equation 3.26, with the update of membership degree $v_{kj}$ and $u_{ik}$ as given in Equations 3.2 and 3.3 respectively.

$$v_{kj} = \frac{\sum_{i=1}^{p} u_{ij}^{m}.x_i}{\sum_{i=1}^{p} u_{ij}^{m}} \tag{3.27}$$

$$u_{ij} = \frac{1/(x_i - v_{kj})^{\frac{2}{m-1}}}{1/\sum_{k=1}^{c}(x_i - v_{kj})^{\frac{2}{m-1}}} \tag{3.28}$$

Based on the provided explanation about the existing techniques for human activity modelling and recognition so far, they are utilised in the proposed enhancement version of Fuzzy Finite State Machine (FFSM) in Chapter 5 and other deep learning techniques in Chapter 6 for creating a different robust version of FFSM.

## 3.10 Discussion

This chapter provides an overview of different techniques to be used later on in this thesis. The reported techniques are used for different applications related to human activity modelling and recognition. The modelling and recognition techniques include statistical methods such as FSM have been shown their dynamic nature that helps in handling the temporal and sequential relationships of inputs between observations at different time steps. In particular, the FFSM, which is the extended version of the classical FSM and it was developed by integrated the (fuzziness degree) with the dynamic nature of the existing FSM. FFSM has been shown its capability of implementing the relations between the fuzziness and the dynamic nurture human behaves. The only challenge coming over at this stage is that the developed FFSM still considering only the experts' knowledge for creating the fuzzy rules that controlling the transitions between the defined states. The existing FFSM model will be enhanced by combining it with different computational techniques for enhancing the ability of the model to learn the related parameters that will be used for generating the fuzzy rules in Chapter 5. The next chapter explains the data collection process, visualising the binary sensory readings in a proper format, and feature extraction and representation process.

# Chapter 4

# Data Collection, Processing and Feature Extraction

## 4.1 Introduction

To recognising human activities in a smart environment and develop a modelling technique to represent their occupier's behaviour, it is essential to monitor the activities first. Once an activity is identified, it would be possible to offer assistance and support according to the needs. Moreover, once the activities of daily living are recognised, some future activities could be predicted. For example, spending a long time in the bedroom, using the toilet many times during the day, or unusual absence for long periods could be an indicator for some possible issues [33]. Therefore, it is important to collect data representing human activities from a real environment, in particular for the older adults living alone. It is also essential to respect the user's privacy with no disruption to the routine activities. When the monitored person is living at a distance, the collected data could be used to send reports or alerts in the form of e-mail or phone calls to family members or carers. Thus, their family/health-services are informed if any problems in their health are recognised, which could lead to other issues.

This chapter gives an overview of the data collection system from a real-world environment using ambient sensory devices. The collected data will be recorded as a binary string data. As an important step, the collected binary string should be visualised and represented in a suitable format before the activities are identified, and relevant features are extracted. Different techniques are suggested for visualising and representing the binary sensor readings.

This chapter is organised as follows: in Section 4.2, an overview of ambient intelligence systems is presented. The process of using ambient sensory device-based to collect data representing human activities is discussed in Section 4.3, where two different data sets are explained in detail. Data handling techniques for visualising and representing the collected data is discussed in Section 4.4. Feature extraction and representation processes are introduced in Section 4.5. Some conclusions are drawn in Section 4.6.

## 4.2 Ambient Intelligence Environment

Ambient Intelligence (AmI) is considered as a new discipline to model peoples' daily activities within an environment using sensors installed in the environment [60]. The quality of users' lifestyle can be improved by employing different applications of AmI through conducting relevant environmental, particularly for older adults who prefer to live independently in their own homes [128]. To be able to meet this target, human activity modelling and recognition systems are required to identify users' activities during their daily routine [82, 129].

Ambient sensory device-based monitoring systems are the most commonly used approach for data collection in AmI applications for human activity modelling and recognition [38, 130]. They form a pervasive infrastructure where the user is embedded by a large number of interconnected sensory devices such as Passive Infrared (PIR), on/off door entry switch sensors, bed pressure sensors and electricity plugs. The reasons behind the ubiquitous of the ambient-based systems for gathering data representing human activity because these systems do not interfere with regular daily activities. To collect activity data autonomously without human interference, items and objects within the monitoring environment are equipped with sensors to collect relevant information [131, 132].

Human activity modelling and recognition systems could be a combination

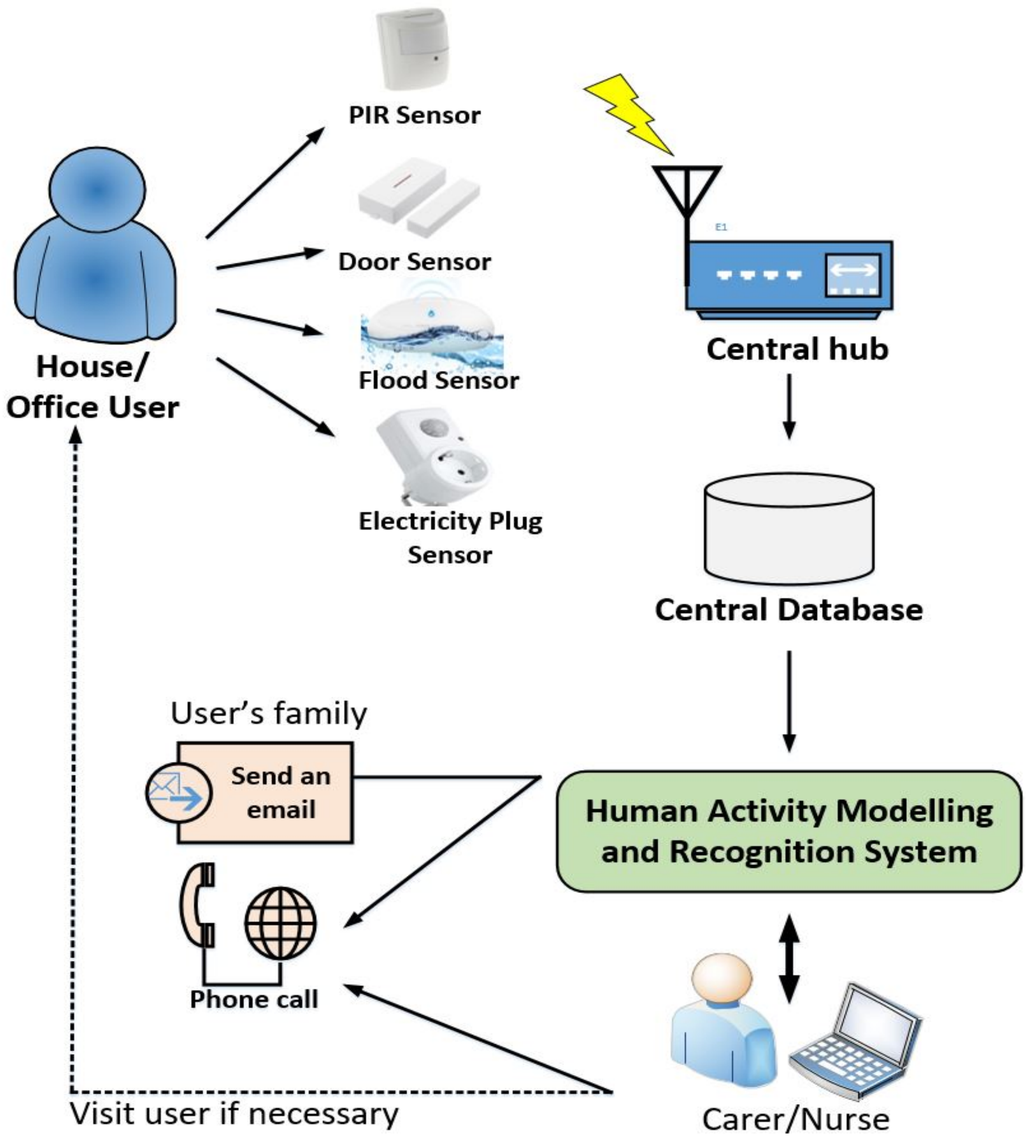



Figure 4.1: A schematic representation of human activity modelling and recognising.

of different techniques include learning algorithms [133], speech recognition [134], and human gesture classification [135, 136]. Human activity modelling and recognition systems are formed user's daily activities to help taking decisions at the right time related to the information gathered and historical data accumulated.

Figure 4.1 illustrates a schematic representation of human activity modelling and recognition system within an intelligent environment. The data gathered using ambient sensory device-based is communicated with a central hub and eventually stored in a central database. The communication channel between the used sensors and the Central hub could be either wired or wireless technology.

## 4.3 Data Collection

Data collection process for human activity modelling and recognition is the process of gathering information representing human activity within an intelligent environment. It consists of spatially distributed independent sensory devices that are used to monitor physical or environmental conditions in different locations [77, 137].

The Ambient sensory device-based systems could be used deploying wired or wireless technologies to connect the central hob with the installed sensors in the environment. However, wireless technology is widely used for communicating the used sensors with the central hub for gathering the data as it does not require an infrastructure be made in the monitoring area. Also, the wireless-based technology for the data collection systems is the preferred option, as it is easy to install them in existing homes. However, some other wired-based data collection systems are still in use such as X10 technology or other well established wired sensor networks in which the used sensory devices are communicating with the central hub via electrical power lines [138]. As all used sensors in the wireless-based data collection systems are powered by batteries, therefore, sensors' power consumption is the most important concern [130]. Figure 4.2, represents the main steps of the data collection system using ambient sensory device-based for collecting ADL for a single user.

The collection process is started by installing different types of sensors in the monitoring environment to monitor physical (e.g., detect objects movement, doors open/close status and bed/sofa occupancy) or ambient environmental conditions (e.g., ambient temperature, ambient light intensity and humidity) in different locations. Once this information is detected, it will be saved in a central database in a format of the binary string containing the sensors' readings as $0, 1$ as it is shown in Figure 4.3. Different data visualisation and representation techniques

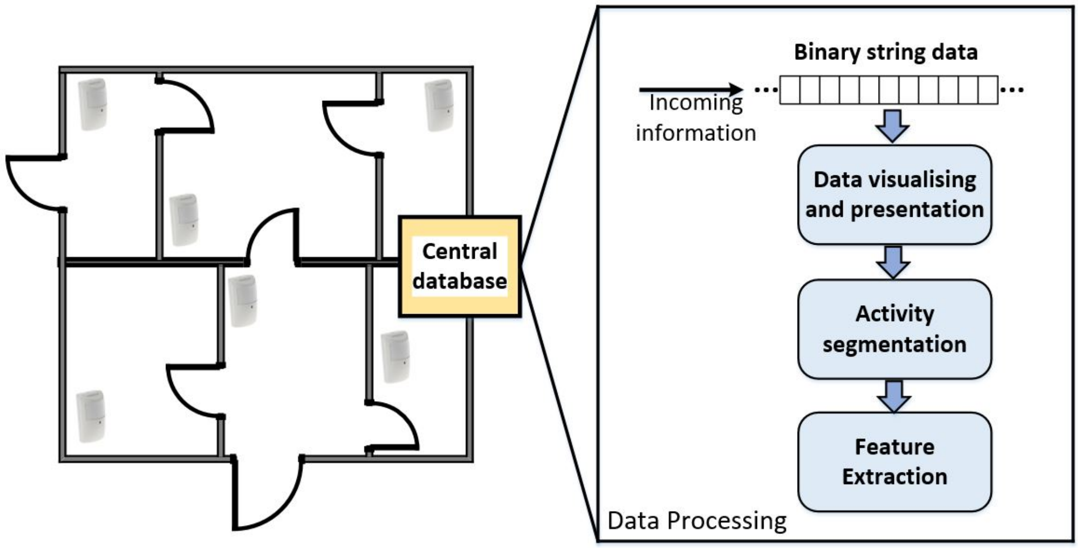



Figure 4.2: The data collection process.

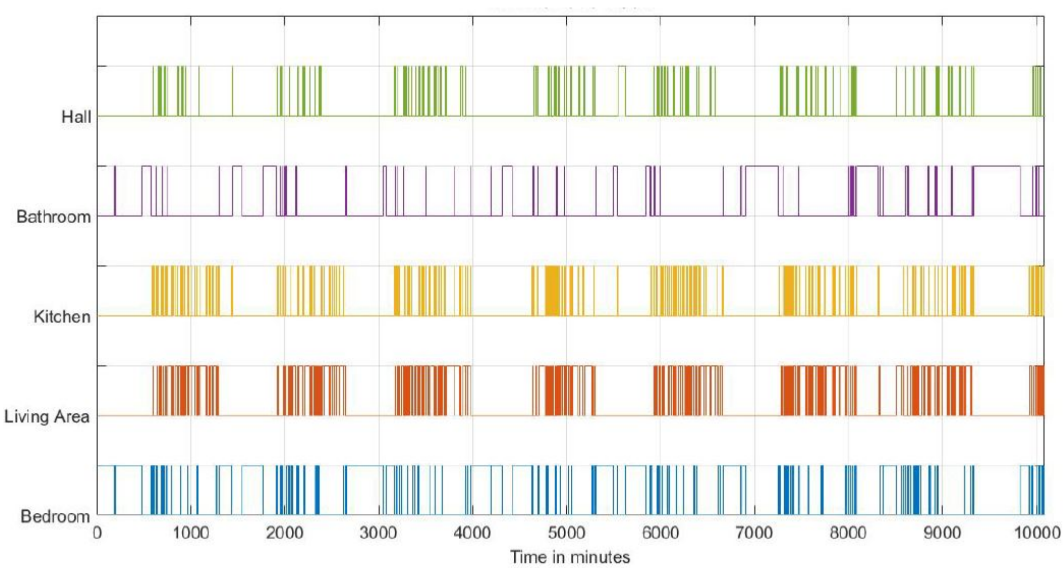



Figure 4.3: Samples of sensor readings for 7 days gathered from five different PIR sensors in a home environment.

are used for reformatting the binary sensor readings as a useful format to segment the activities before a feature extraction process is applied.

In general, various sensors for detecting the user information within the monitoring area using the ambient sensory device-based systems are:

- Passive Infrared Sensors (PIR), known as motion detector sensors, are sensitive to detect the movements of the objects in the ambient environment. PIR sensors respond to detected movements by changing its value to 1, and to 0 when stop detecting the movements. Placing the PIR sensors in the right location is important to capture the movements.

- on/off switch sensors that are used as door/window entry point sensors. The sensors are magnetic sensors used to detect the open and close events of a door/window.

- Electricity power usage plug sensors are utilised to measure/monitor the usage of the electrical devices connoted to them. I particular, they are used to measure the electrical current consumption for a specific electrical device.

- Mat pressure sensors are used to detect the occupancy of the presence in and usage of sofas/beds.

To evaluation the proposed models introduced in Chapter 5 and Chapter 6, two datasets referred to as Dataset A and Dataset B are used in the remaining parts of this thesis. Details of both datasets are provided below:

### 4.3.1 Dataset A

This dataset was collected by our research group from a real home environment representing the ADL of a single user. The dataset was collected at the SmartNTU Home facilities within Nottingham Trent University. A floor plan of the house is shown in Figure 4.4. A list of the used sensors for collecting this dataset is listed in Table 4.1. There are seven daily activities, which are *Sleeping, Toilet, Kitchen, Dining-room, Living-room, Garden*, and *Leaving*.

All binary sensory devices are equipped with batteries and connected wirelessly using a Z-wave technology to a hub; thus they can be installed easily on any place within the testbed environment and connected to a server via a wireless network. Therefore, any events (e.g., detected motion or no motion) are chronologically logged in the server. Each event log contains four parts: date, time-stamp when the sensor is triggered, sensor ID, and sensor status, as

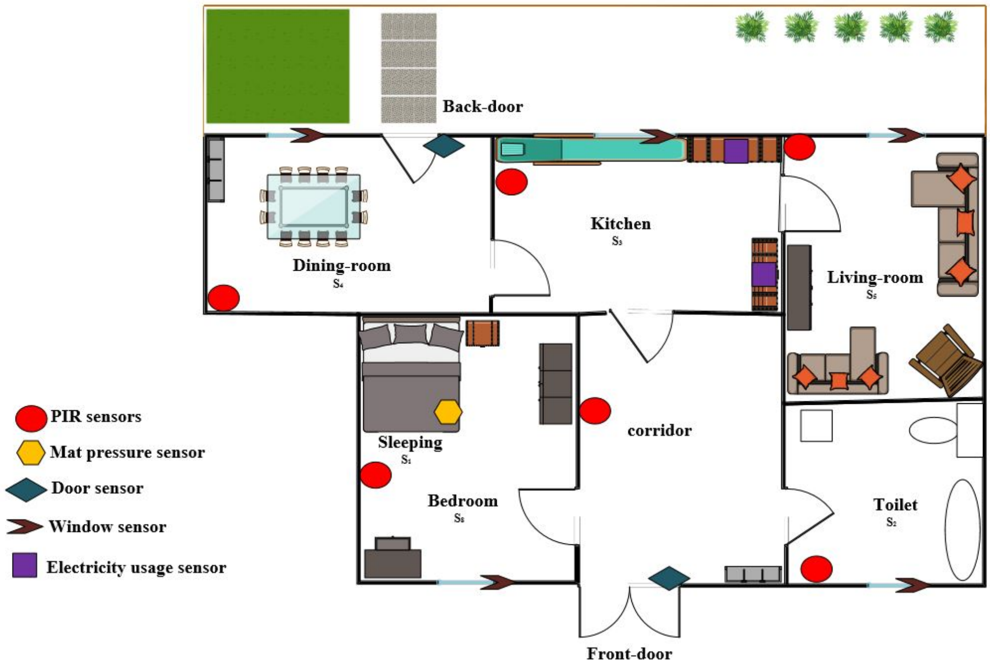



Figure 4.4: Floor plan layout and location of the installed sensors used for data collection in dataset A.

Table 4.1: List of sensors used for collecting the dataset A to measure different conditions and activities. (* denotes the unused sensors in this study).

| Sensor | Purpose of use |
|---|---|
| Passive Infrared (PIR) | Detecting the movement. |
| Entry door(s) on/off switches | Detecting when doors are opened and closed. |
| Mat pressure sensor | Measuring bed and sofa occupancy. |
| Electricity consumption plugs | Measuring electricity consumption. |
| * Indoor temperature sensor | Measuring ambient temperature. |
| * Outdoor temperature sensor | Measuring outdoor temperature. |
| * humidity sensor | Measuring ambient humidity. |
| * Light intensity sensor | Measuring ambient light intensity. |

Table 4.2: A Sample of the gathered ADL dataset.

| Date and Time | Sensor Status | Sensor ID |
|---|---|---|
| 2018-01-04 09:34:45 | 1 | Sleeping |
| 2018-01-04 09:42:18 | 1 | Corridor |
| 2018-01-04 09:43:26 | 1 | Toilet |
| 2018-01-04 09:55:18 | 0 | Sleeping |
| 2018-01-04 09:56:55 | 0 | Corridor |
| 2018-01-04 09:59:07 | 1 | Toilet |
| 2018-01-04 10:00:18 | 1 | Corridor |
| 2018-01-04 10:01:21 | 0 | Toilet |
| 2018-01-04 10:01:34 | 1 | Kitchen |
| 2018-01-04 10:06:37 | 1 | Kitchen |
| 2018-01-04 10:06:54 | 1 | Corridor |
| 2018-01-04 10:07:11 | 1 | Living-room |
| 2018-01-04 10:07:49 | 0 | Kitchen |
| 2018-01-04 10:08:16 | 0 | Corridor |
| ... | ... | ... |

illustrated in Table 4.2. The sensor ID refers to the location where this sensor is installed (e.g., kitchen, toilet or living room) and the sensor status value is either 1 when detecting an event (e.g., movement, a door opens/close, or bed/sofa occupied) or 0 otherwise. Also, the PIR motion sensors are labelled with the names of the places they are installed in, such as toilet, living-room, etc. These motion sensors send a simple 1 message when motion is present under the coverage area, followed by an 0 message shortly after the movement is stopped.

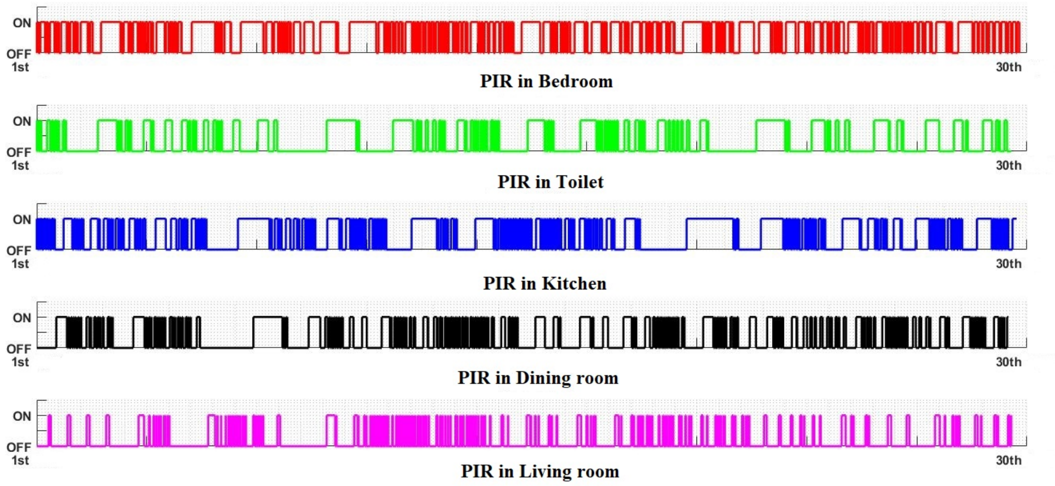
The collected raw dataset containing $8,728$ events (sensor readings) were logged from 6 PIR motion sensors, five-window sensors, two-door switch sensors, and one electricity usage sensor for 28 days. Typical samples from the raw dataset is represented in Figure 4.5. As the main aim of this research is to model and recognise human activity for only a single user, only one user is considered to be at the house while collecting the data. The process of collecting this dataset was done using different human participants (one at the time) to ensure no repetition, and the data pattern will not be similar. The aforementioned dataset events correspond to 7 activities, which are *Sleeping (*87 *instances)*, *Toilet* (161 instances), *Kitchen* (240 instances), *Dining-room* (398 instances), *Living-room* (259 instances), *Garden* (119 instances), and *Leaving-home* (76 instances).

The ambient sensory devices-based systems face hindrances in the multi-occupancy environment due to its sensing methodology, e.g. the PIR sensor can detect human body movement but fails to distinguish individual occupants. Therefore, a hybrid data collection method (e.g., Ambient and wearable sensors) will work much better for such applications. Also, modelling and recognising Concurrent or Simultaneous Activities for a single user is a challenge for the ambient sensory device-based systems. By nature, several activities can be undertaken by a single user at the same time. For example, people can read a book while they are watching TV or eating. In this case, it is not necessary to know which activity started first. That means the existence of concurrent activities when an activity (e.g., eating) starts. In contrast, the other activity is already started (e.g., reading a book). A specialised hybrid approach for collecting data is required to recognise these non-sequential activities.

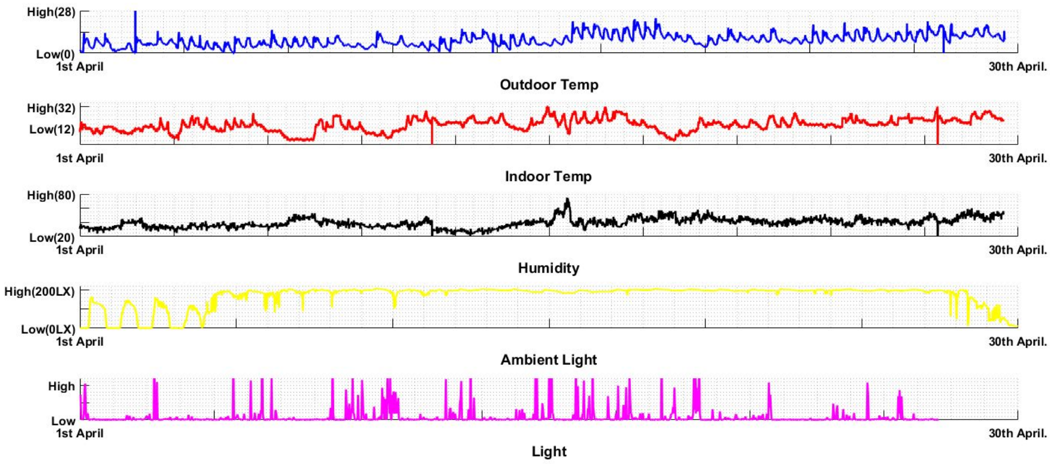
There are some sensors readings such as ambient temperature, humidity and ambient light density which were available to us but not utilised for our study, as mentioned in Table 4.1. Figure 4.5, illustrates a sample of the collected data during one month time from different types of ambient sensory device-based. Figure 4.5-a shows samples of the data gathered from PIR sensors, and Figure 4.5-b shows samples of the data gathered from other ambient sensors such as temperature, humidity and light density sensors.

### 4.3.2 Dataset B

This is a publicly available dataset known as Aruba dataset representing ADL for a single user was collected using the Centre for Advanced Studies in Adaptive System (CASAS) at Washington State University [133]. They used motion, door, and temperature sensors. However, as this work focuses on the ADL, the temperature sensors are excluded, and the other 34 sensors (3 door sensors and 31 motion sensors) are used. Figure 4.6, illustrates the floor plan layout of Aruba testbed that is used to collect dataset B.




(a)




(b)

Figure 4.5: Data collected from a real environment over month time: a) samples of the data gathered from PIR sensors; and b) samples of the data gathered from other ambient sensors such as temperature, humidity and light density sensors.

A single elderly woman lived in the Aruba testbed, and she had received regular visits from her children and grandchildren during the data collection period. The final dataset is saved as a list of sensor-ID, time-stamp, and sensor status.

In this dataset, there are 11 activities performed by the women who was living in the apartment and the data were collected over a period of 224 days.

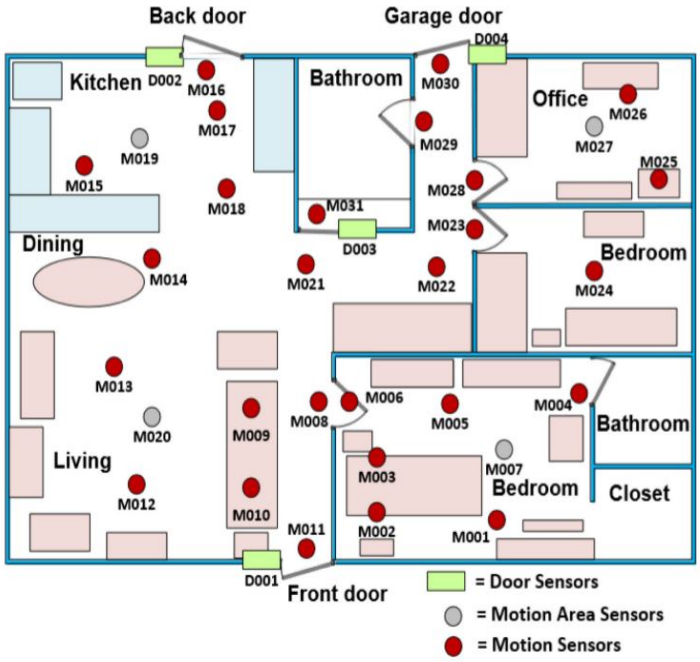



Figure 4.6: Floor plan layout and location of the installed sensors used for data collection in dataset B.

These activities are *Sleeping* (401 instances), *Meal Preparation* (1606 instances), *Relaxing* (2910 instances), *Bed-to-toilet* (157 instances), *Leaving home* (431 instances), *Entering home* (431 instances), *Housekeeping* (33 instances), *Eating* (257 instances), *Washing dishes* (65 instances), *Work* (171 instances) and *Resperate* (6 instances). The *Resperate* activity is excluded from the dataset as it has only 6 instances.

The sensor readings collected from AmI environments in both datasets have different characteristics that could be used for data analysis techniques. It is stored in the central database as a long string of binary data gathered from each sensor which is difficult to understand and analyse manually [139]. This data sometimes includes some noisy if the sensors readings are inaccurate or there may be some missing readings when the sensor fails down [34]. Moreover, the stored values from each sensor are spatial (where the sensor is installed) and temporal (the actual time for the sensor event) [131]. As the majority of the collected information was gathered from low-level sensors, the challenge which may be faced when dealing with such a large volume of the sensor reading is the ability to understand this data and represent it in a proper format.

Different data visualisation and representation techniques are employed with the gathered information in both datasets for handling and reformatting the binary string into a useful format before extracting features from each activity. The process of data handling and visualisation is discussed in the next section.

## 4.4 Data Handling and Visualisation

The gathered sensor readings in Dataset $A$ and Dataset $B$ contain a large volume of binary data. The challenge here is to understand the actual events undertaken by the user from such sensor readings before any feature extraction process applied. The aim of dealing with the data gathered using ambient sensory device-based is to find a flexible and accurate format for representing such sensor readings. Many essential tasks can be performed with the reformatted data exactly the same as the ones would do with the original sensor readings. These tasks could be clustering human activities, classification different activities, activity modelling and recognition, activity prediction, anomalies detection. The sensory readings are often difficult to understand, and it will be even harder when the data collection process has occurred for a long period of time. This becomes even more complicated when the sensory readings gathered from multiple sensors. Due to the fact that only one user is present in the monitored environment, there are no parallel activities in different locations to be detected.

In order to model and recognise human activities for a single user, we need to gather sufficient data representing the daily activities to be able to establish the correlation between different events and activities based on the gathered sensor readings. It should be noted that sensor readings are recorded approximately every 10 seconds and when this repetition occurs for multiple sensors, a difficult challenge would be faced for interpreting and visualising such large amounts of sensor readings. To illustrate the complexity of the sensor data, Figure 4.3 shows the sensor readings gathered from five different PIR sensors over 7 days. It is almost impossible to understand the activities from the sensor data and interpret them correctly.

As mentioned in Chapter 2, different methods in the previous literature work are used to represent and interpret the ADL for a single user living in a smart environment. In this section, several used techniques for interpreting and visualising the gathered sensor readings are presented.

### 4.4.1 Visualisation Using Data Features

It is considered as one of the commonly used and useful visualisation technique which can help to understand the sensory readings that are recorded as a binary string data. To achieve this level of visualisation for a big data set containing large number of low-level data, the gathered binary sensory readings can be transformed and represented differently. Three different data representation approaches could be applied [104, 129, 140]:

- **Raw data**, in this approach, the raw gathered sensor readings are used directly as they were received from the ambient sensory device-based. It gives 1 when the sensor is firing (detecting any movement) and 0 otherwise.
- **Change-point**, the change point representation indicates when a sensor event takes place. This means it indicates when a sensor changes value from 1 to 0 or from 0 to 1. Therefore, it gives 1 when the sensor changes state (i.e. goes from zero to one or vice versa) and 0 otherwise.
- **Last-fired**, the last-fired sensor representation indicates which sensor fired last. The sensor that changed state last continues to give 1 and changes to 0 when another sensor changes state.

By using one of these approaches with the sensor readings, the sequences of the user's movements can be interpreted as who is using a smart environment, and it will show the activities pattern of the person. It will be more difficult to understand human activities when a large scale data set is visualised. For example, Figure 4.3 shows the binary strings and duration recorded from 5 PIR sensors for one person living in a smart home over 7 days. It is clear from the

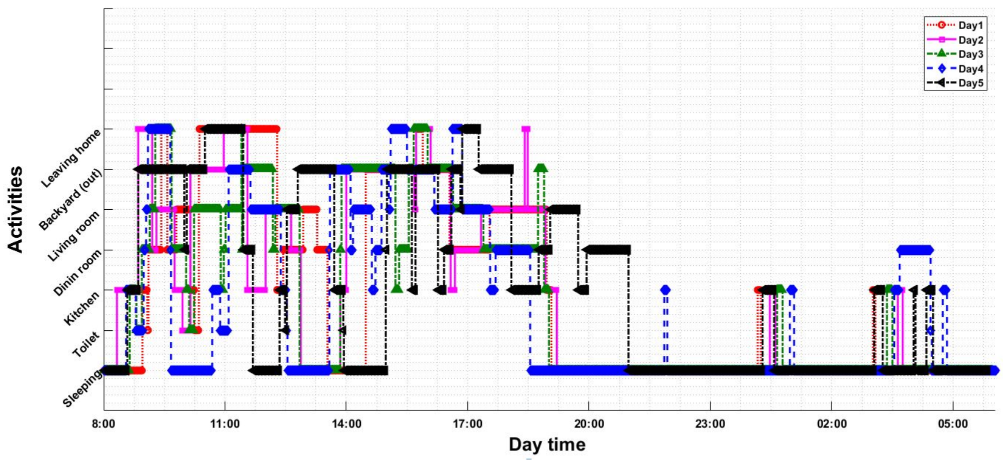



Figure 4.7: Five days' activities represented as an occupancy signal for a single user within a smart home environment extracted from binary sensor readings.

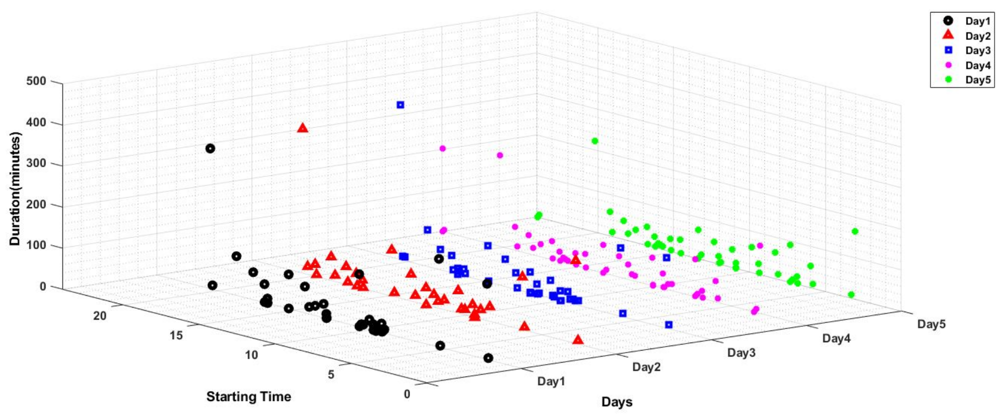



Figure 4.8: Start Time and Duration of five days activities in a smart home.

given example that if the data were for a longer period (e.g. a month), the graph would be more difficult to understand.

In case if the graph shows a short period of the sensors readings such as 1 day, it gives a better idea about the activities in that day, but it will not give a holistic view of the monitored activities. The reformatted binary string data can be visualised as it is depicted in Figure 4.7. This figure illustrates the movements between the daily activities and their duration for 5 days for a single user in a smart home. Also, it could be shown as a scatter plot of a sample data set representing 5 days and the duration of the activities for the same person in

Figure 4.8. The data sets presented in these two graphs are represented and visualised from the sensor readings that gathered using ambient sensory device-based representing the ADL for a single user.

The user's activities can be more easily understood and interpreted in Figures 4.7 and 4.8, than Figure 4.3. For instance, Figure 4.7, shows the user's ADL as an occupancy signal representing the sequences of the movements between different activities for one day for this user and the duration for each time he/she visits any monitored location within the home. It is difficult to achieve this level of visualisation and understanding from the gathered sensor readings displayed in Figure 4.3.

### 4.4.2 Visualisation Using Finite State Machine

One of the techniques to model dynamic processes when data changes over time is the Finite State Machine (FSM) [22]. The FSM contains several states representing different actions and the mechanism of transitions between them. Many researchers have considered using the FSM to model and represent dynamic data. Since data representing a single human activities are restricted to identify one state (activity) at any time based on the gathered binary data, FSM can be employed to interpret such data. In the gathered binary sensor readings, each binary string record from the used sensors contains the information presented as a triple $z = (t, a_{on}, a_{off})$, where $z$ is the sensor's location representing as an ID for each sensor, $t$ is the timestamp when any sensor changes its value $a \in a_{on}, a_{off}$ at the time instant $t$, represented as a binary data 1 or 0. The data must then be represented in a proper format so that it can be more easily accessed, enumerated and represented. This can be achieved by representing the gathered data as time series sequences format. These sequences are represented by a series of sensor values and sensor ID to identify user's activities as occupancy signal. In this case, all sequences are ordered in time and occur sequentially one after another.

Before any feature extraction process occurs, it is important to transform and visualise the collected sensor readings shown in an appropriate format. This could be achieved using practical knowledge-based techniques such as an ontology

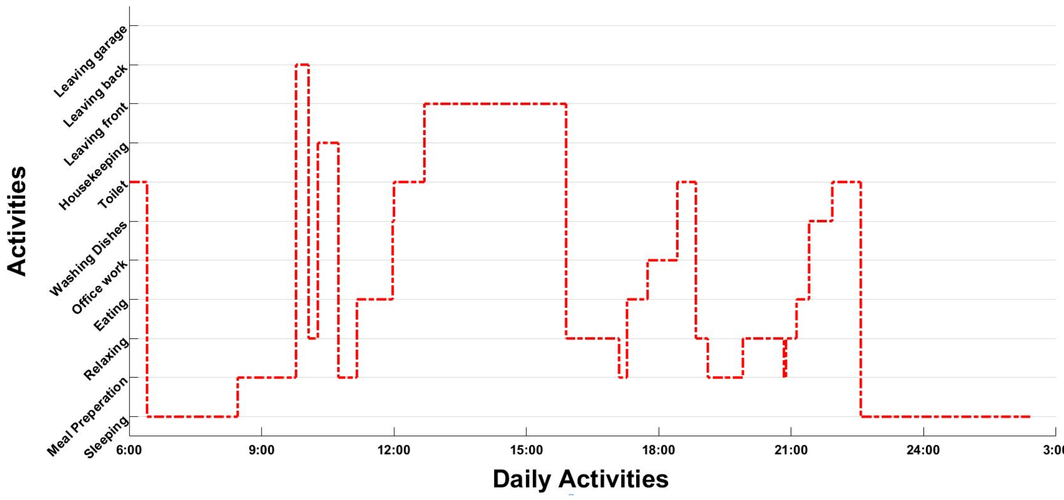



Figure 4.9: Occupancy data representing the daily activities of a single user for one day as a multi-level graph.

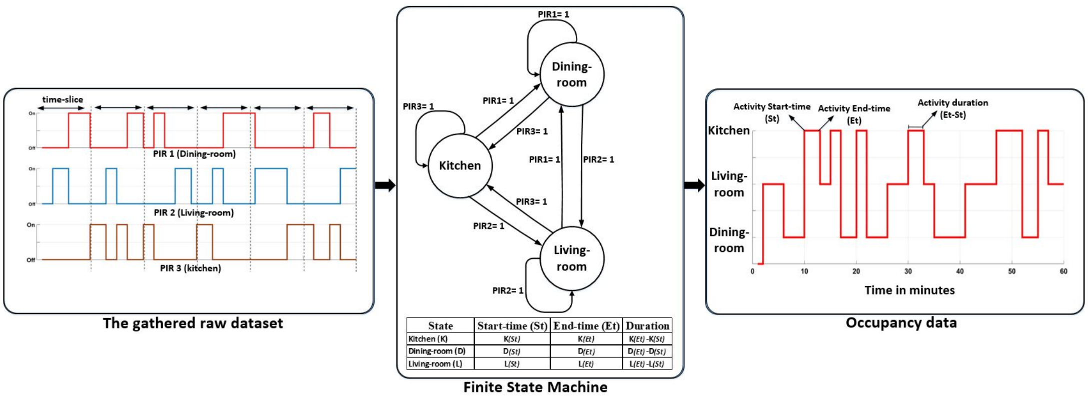



Figure 4.10: An illustrative example of conversion of the time-slice windows of the gathered binary string information into occupancy signal using Finite State Machine.

[35, 36, 140], or using a computational intelligence technique integrated with the collected sensory information [4, 88]. This process is used to visualise the collected sensor readings as occupancy signal containing the user's daily activity patterns as it is shown in Figures 4.7 and 4.9, where each activity is represented as an activity window ($W$). In this work, a form of FSM is used to represent and visualise the collected sensor readings as occupancy data to identify the user's activities.

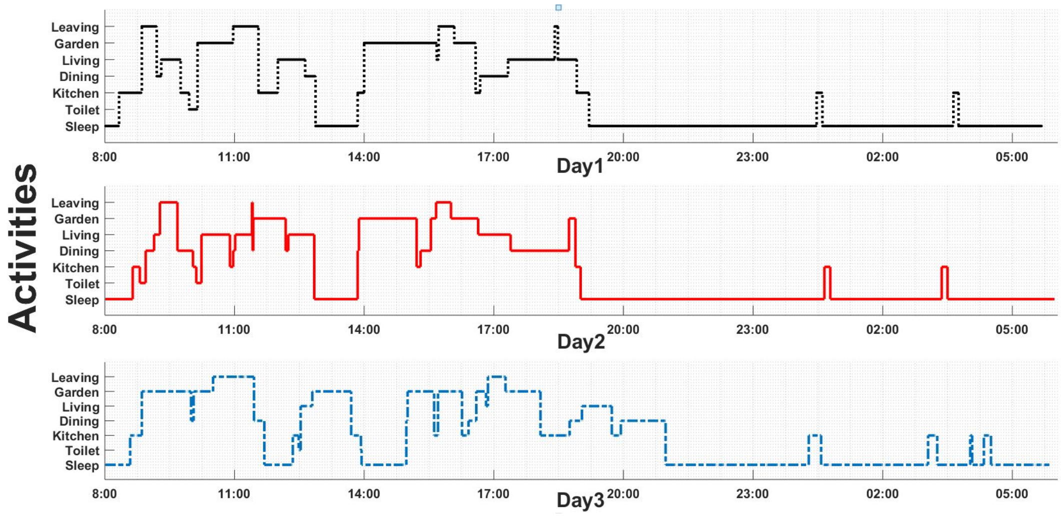



Figure 4.11: Occupancy signal for three days representing the daily activities for a single user that represented as a multi-level graphs.

Considering the gathered time sequence information $z = a(t)$ integrated with the recorded binary string representing the occupancy in a specific location $z$ at time $t$: this signal has two values of either $a = 1$ or $a = 0$, representing the presence and absence from that area; $z(t) \in [0, 1]$. To represent such data in an efficient format, a time-slice chunks approach [129] with a length of 60 seconds is applied to the binary sequence. Then, these time chunks are mapped into $last - fired$ sensor representation using FSM, considering the *Start* and *End* time for each activity, as illustrated in Figure 4.10. The $last - fired$ sensor representation format is used to indicate which sensor is fired last. The sensor that changed its state last continues to give 1, and this changes to 0 when another sensor gives 1 during the specified time slice [129]. The generated occupancy data representing the user's daily activity pattern is shown in Figure 4.11, where a sample of three days' activities representing the ADL for a single user is presented as an occupancy signal.

Once the activities are visualised as an occupancy signal, the next step is to extract the numerical features representing these activities, as explained in Section 4.5.

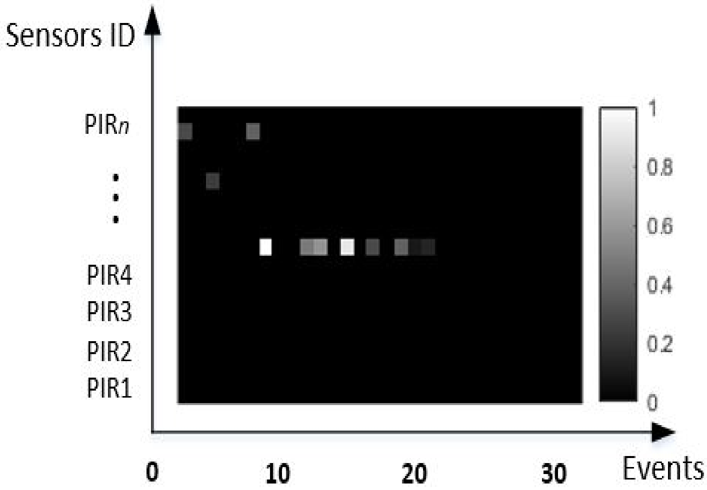



Figure 4.12: Greyscale image generated from binary sensor readings.

### 4.4.3 Visualisation using 2D Greyscale Images

Image representation is also used to visualise the ADL data. As the recorded sensor statuses were either 1 or 0, greyscale colour is used to represent the duration that the sensor was giving the value of 1, where the white colour represents the sensor value of 1 and black colour represents the sensor value 0. In the data preparation stage, binary data is gathered from a real-world environment and then visualised as a greyscale image. The sensors used for gathering data are low-level sensory devices. These devices are considered as a simple kind of sensor that could be used for many different purposes for measuring real parameters, such as object movement, doors opening and closing, and bed/chair occupancy. The data recorded from these sensors are saved as binary string data containing ones (1) when sensors detect any events such as the presence of motion, and zeros (0) shortly after the motion has stopped.

There are many different ways of representing and visualising such data. However, the process of dealing with this kind of data is still a challenging task, due to the lack of information that can be extracted from it [52, 88]. In some recent studies, researchers have reported the use of a time-slice window approach [11, 88] for identifying different events based on a binary string that is gathered using low-level sensory devices. Another way of dealing with this binary data is by applying a manual approach for extracting useful features including start and end times, and the duration for each event [34]. Also, several practical knowledge-based techniques, such as ontology [35, 36], are widely used to deal with such data.

A novel method of representing binary string data as a 2D greyscale image is applied. Each event is visualised as a greyscale image, as shown in Figure 4.12. The generated image is used for extracting unique features that are used to train the proposed model for modelling and recognition of human activity. Each pixel in the x-axis of the generated image represents the sensor status, whereas the sensor ID is denoted on the y-axis. The greyscale range represents the duration of the event in seconds. More details about this method are provided in Chapter 6, when a binary data string representing human activity is visualised as greyscale images and the experiment is conducted.

## 4.5 Feature Extraction and Representation

Feature extraction is an essential process of any activity recognition system as raw data representing the activities do not provide enough information to allow implementing an activity recognition system. Features needed to model and recognise human activities could be computed using several techniques. For example, manual approaches based on the gathered raw data from the used sensory devices [88], or using computational intelligence techniques integrated with the collected sensory data [4, 5, 141, 142]. Once the gathered sensor information is visualised as an occupancy signal containing the user's daily activity patterns, a manual feature extraction approach is applied to obtain a set of useful feature vectors that representing human activities.

In the research presented in this thesis, two different approaches of extracting features representing the human activity from the gathered data are employed. These two approaches are Fuzzy Feature Representation approach, which is applied to the manually extracted numerical information from the occupancy signal; Employed a pre-trained Deep Convolutional Neural Network (CNN) such as AlexNet with the generated greyscale image.

### 4.5.1 Fuzzy Feature Representation

Extraction of the numerical information from the occupancy signal that is generated from acquired raw sensor data is crucial to any learning system as raw data does not provide adequate information that can be used as inputs to the model. Therefore, the numerical features are extracted for each activity-window identified in the occupancy signal. Five different features are extracted; start time, end time, duration, activity count, and activity sequential order are extracted. Therefore, the activity data are extracted for each activity window and represented as a matrix where rows are the length of the activity window and columns are the number of recorded information from the sensors in the window. The extracted numerical feature set is $U = [u_1, u_2, u_3, u_4, u_5]$. where,

- $u_1$ represent the activity start time, which the time when this activity is started.
- $u_2$ represent the activity end time, which the time when this activity is ended.
- $u_3$ denotes the duration of each activity which is calculated by subtracted the end time from the start time for each activity and it is represented in minutes.
- $u_4$ represents the activity count, which defines as a number that represents how many times the activity being occurred per hour/day.
- $u_5$ represents the sequential order of daily activities. As human activity is a sequential matter.

Each input variable is fuzzified to translate the numerical feature to a fuzzy set. The values in fuzzy set are represented as fuzzy Membership Functions (MFs). The associated linguistic labels with each input are then used as fuzzified inputs to the activity modelling and recognition model. The representation of these features as a fuzzy sets are explained in this section.

Fuzzy feature representation approach is designed to convert the collected information into their relevant membership degrees [73]. The resulting membership degrees are taken as features to be used as inputs to train the proposed model. Therefore, fuzzy feature representation is applied to determine the number of Membership Functions (MFs) representing the input data as membership degrees. By replacing each value in the input data with their corresponding degree of memberships; thus, each value in the input data is represented as fuzzified values obtained for each MF as follows:

$$X_{uj} = [\mu_{A^1_{uj}}, \mu_{A^2_{uj}}, ..., \mu_{A^M_{uj}}] \qquad j = 1, ..., P \tag{4.1}$$

$X_{uj}$ is the fuzzified set of the input variable $u_j$. $P$ is the last value in the input variable $u_j$. $\mu_A$ is the degree of MF associated with each linguistic label. The Fuzzy feature representation process can be summarised as follows:

1. Apply the input data to the fuzzifier algorithm consists of $M$ MFs that are represented with the linguistic labels.
2. Define the degree of fuzziness $\mu_A$ that corresponds to each MF.
3. Determine the maximum degree of fuzziness for the variable $u_j$ in each iteration.
4. Create a matrix $q = r \times z$ to store the degree of MFs for each variable $u_j$. Where $r$ is the total number of activity instances in the input data $u_j$, and $z$ is the number of fuzzified values for the variable $u_j$.
5. Update the matrix $q$ after each iteration with the new fuzzified values that corresponding to the next input value.

The final set of the fuzzified features $X_{uj} = [\mu_{A^1_{uj}}, \mu_{A^2_{uj}}, ..., \mu_{A^M_{uj}}]$, will be used as inputs to train the proposed model for learning the relations between the inputs and output data, as it is explained in the next sections. The process of fuzzy feature representation is elaborated in Chapter 5 when modelling and recognising human activity experiments are conducted with a real dataset representing ADL.

### 4.5.2 Deep Leaning Convolutional Neural Network for Feature Extraction

Once the binary data representing an event is presented as a greyscale image, it is important to find a suitable way of extracting features from the generated image [5]. One of the most common approaches for extracting unique features from the images is by employing a Deep Convolutional Neural Network (DCNN) for automatic feature extraction [143]. DCNN can extract image features automatically by tuning the parameters in its convolutional and pooling layers. AlexNet is one of the powerful pre-trained DCNNs which could be used for the purpose of feature extraction. In particular, AlexNet contains several layers, starting from the input later, followed by 5 convolutional layers, 3 pooling layers, and 3 fully-connected layers, with the last layer being the output layer. The output feature vector's dimensional value is 4096, computed by cross-mixing in the first fully connected layer for the two groups' features. This process will be repeated for the next fully-connected layer and so on until reaching the last-fully connected layer, where the two feature groups are combined into a single feature vector [144].

In work presented in this section, a feature vector containing certain features which are suitable for modelling and recognising human activity is extracted from a greyscale image using a pre-trained AlexNet DCNN. The feature vector is obtained directly from fully connected layers and then used as inputs to the proposed models in Chapter 5.

## 4.6 Discussion

In this chapter, environment and data collection system employed in this work are discussed. Ambient sensory device-based is primarily employed for gathering data representing ADL for a single user with is the smart environment. The difficulties involved in processing the large values of the binary string information of the data collected from ambient sensory device-based are also explained. Two different data sets mentioned as dataset A and B are described in this chapter. The approaches employed for representing the gathered data as occupancy signal or as a 2D greyscale image are explained. Once the occupancy signal is generated, the feature extracted processes are applied using fuzzy feature representation approach or by employing a pre-trained AlexNet are discussed.

The fuzzy sets generated using the fuzzy feature representation approach and the feature vectors extracted from the greyscale images for each activity are used in Chapter 5 and Chapter 6 as inputs to the proposed models for human activity modelling and recognition. In particular, they will be used to evaluate the proposed enhanced Fuzzy Finite State Machine and the other proposed deep learning techniques for human activity modelling and recognition.

# Chapter 5

# Enhanced Fuzzy Finite State Machine for Human Activity Modelling and Recognition

## 5.1 Introduction

This chapter is an extension of the explanation provided in Chapter 3, for developing a Fuzzy Finite State Machine (FFSM) used for modelling and recognising human activity. In this chapter, the FFSM is introduced as a means of defining human daily activities and the transition between the states (here, the activities). There are many unknown parameters in the FFSM, representing a model for the ADL, which needs to be identified. The aim of the research reported here is to identify the parameters to represent the real activities of a human subject in an AmI environment accurately. The research presented in this chapter addresses only the challenges in modelling and recognising a single-occupancy at a real-home environment based on a dataset collected from ambient sensory devices. To achieve robust modelling to recognise human activities, the research reported in this chapter has made the following contributions:

- Integrating the advantages of learning capabilities of Neural Network (NN) with the expert knowledge within the FFSM: The results of this integration is a new model named Neuro-Fuzzy Finite State Machine (N-FFSM), and it is described in Section 5.2.

- Integrating the FFSM with Long Short-Term Memory (LSTM) neural networks to enhance the learning capability of the FFSM model in temporal data (such as the human activities dataset). This can lead to accurately generating the fuzzy rules that govern the transition between the system states. The new model is referred to as a Short-Term Memory-Fuzzy Finite State Machine (LSTM-FFSM) and is described in Section 5.3.

- Integrating the FFSM with Convolutional Neural Network (CNN) to add the learning abilities of CNN to model daily human activities based on the numerical and temporal information gathered from the sensory data. The new model is referred to as Convolutional-Fuzzy Finite State Machine (C-FFSM), and it is described in Section 5.4. Also, the results of testing and evaluating the proposed models using the two datasets gathered from real home environments representing ADL for a single user is illustrated. Details of the datasets are already provided in Chapter 4.

This chapter is structured as follows; the process of integrating FSSM with different learning algorithm including NN, Long Short-Term Memory, and Convolutional Neural Networks has led to developing three enhanced FFSM models (N-FFSM, LSTM-FFSM and CNN-FFSM) is explained in Section 5.2, 5.3 and 5.4, respectively. Also, carried out experiments to test and evaluate the enhanced FFSM models for modelling and recognising human activities is conducted in Section 5.5. The obtained results are discussed in Section 5.5.3.

## 5.2 Neuro-Fuzzy Finite State Machine

A common approach in incorporating learning capabilities into fuzzy systems is based on the combination of fuzzy systems and NN leading to the well-known hybrid systems called Neuro-Fuzzy systems [145]. In the standard fuzzy logic systems, the fuzzy rules are generally created based on the expert knowledge rather than the numerical data [146]. Fuzzy logic systems can reason with imprecise information, and they are efficient at explaining the decisions made by the fuzzy systems. However, they seem to lack the ability to attain the decision rules on their own [145]. On the other hand, a neural network is described as a mathematical model which consists of multi-layers with distributed artificial neurons to perform parallel computations and pattern recognition. However, it is difficult to explain how they attain their decisions [147]. The individual disadvantages of these two computational intelligence approaches could be overcome by combining them for developing intelligent hybrid systems. An example of these intelligent hybrid systems is known as Neuro-Fuzzy System (NFS) [148]. This combination is coming up with a human-like reasoning capability of fuzzy systems and the learning abilities of the neural network algorithms to determine the parameters of the fuzzy system.

In this section, the fusion framework between FFSM and NNs is explained. This section discusses an extended version of FFSM used for human activity modelling recognition based on the data gathered from low-level ambient sensors. The original FFSM is extended by integrating the learning capabilities of NN for generating the fuzzy rules that govern the fuzzy states' transitions. Moreover, experts' knowledge is used to identify the number of sates, the number of linguistic labels associated with each input and the general structure of the rules.

Figure 5.1 illustrates the schematic diagram of the proposed Neuro-Fuzzy Finite State Machine (N-FFSM). The proposed N-FFSM model can automatically generate the fuzzy rules representing the state transitions. The fuzzy rules and the associated linguistic labels to each input are automatically derived by the Neuro-Fuzzy. Therefore, it is possible to construct the Membership Functions (MFs) associated with the linguistic labels used in the fuzzy rules.

The N-FFSM system is considered as an adaptive network which is functionally equivalent to the fuzzy systems in term of representing the fuzzy rules linguistically with the capabilities of learning of the neural network. This network is comprised of nodes (neurons) identifying specific functions gathered in layers. The final output of these layers can construct a network that can generate the fuzzy rules.

The system will be initialised by setting up several states based on experts'

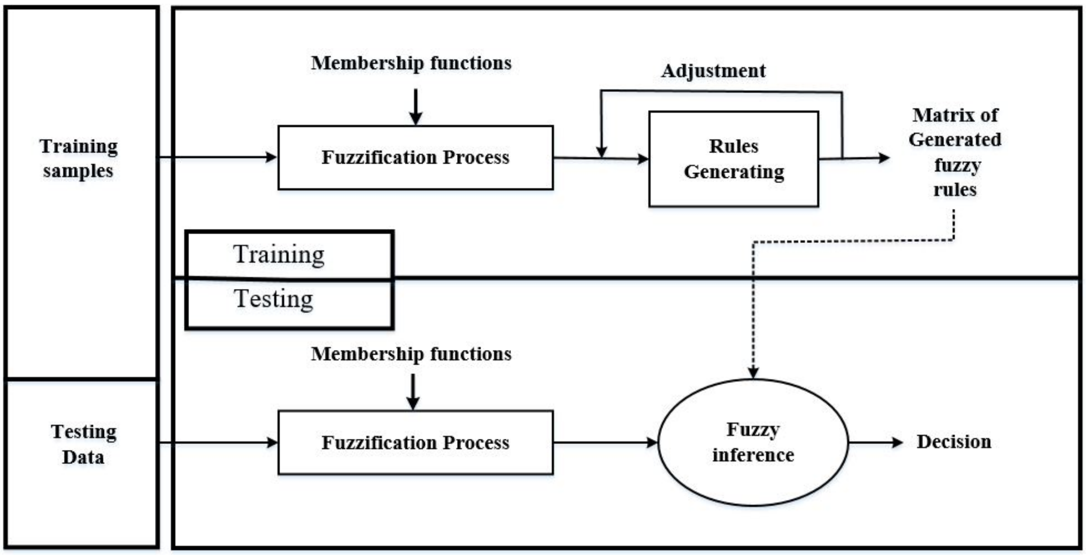



Figure 5.1: A schematic diagram of the proposed Neuro-Fuzzy Finite State Machine.

knowledge to identify the activities that are going to be recognised. The identified states are set up as a set of linguistic variables $S(t) = \{s_1(t), s_2(t), ..., s_i(t), ..., s_N(t)\}$ where $N$ is the number of states at a time $t$. Then as it is illustrated in Figure 5.1, the input variables $U(t) = [u_1(t), \dots, u_j(t), \dots, u_P(t)]$, where $P$ is the last value in the input variable $u_j$, is represented as a fuzzy set $X_{uj}$. Standard NN is employed with the fuzzified data to learn the relations between the inputs sets. The learnt parameters during the training stage, are used to generate the matrix of fuzzy rules $R^{\lambda}_{mn}$ that is used to govern the transition between the system's state. Once the system is trained, the generated matrix of fuzzy rules is applied to a new data set in the testing stage to evaluate the performance of the modelling process using this new data set. Considering the complexity of modelling a large scale dataset based on only the expert's knowledge, it is difficult to identify the rules analytically. The transition between any two states $s_m$ and $s_n$ in the Neuro-FFSM is govern by fuzzy rule, which identified based on expert's knowledge and the learning capabilities of the Neural Network as:

$$R^{\lambda}_{mn}\textbf{: IF } (S(t) \textbf{ is } s_m) \textbf{ AND } H_{mn} \textbf{ THEN } S(t+1) \textbf{ is } s_n$$

where, $(S(t)$ **is** $s_m)$ is used to determine the current state $S(t)$ as $s_m$ at time $t$,

$H_{mn}$ represents all constraints imposed on the input variables that are required to either remain in state $s_m$ (when, $m = n$) or change to state $s_n$. The part of the given rule after the **THEN** determines the value of the next state $S(t+1)$ for being in state $s_n$.

During the training phase, each rule can be interpreted as a training pattern for a multi-layer neural network, where the antecedent term of the rule has two parts; the current state of the system $S(t)$, and the output of the neural network $H_{mn}$, which used for the next state $S(t+1)$ at the consequence part of the rule. There is no need to use separate outputs for the next state $S(t+1)$ and the NN output $H_{mn}$ of the FFSM, since a Moore type of FFSM is implemented.

Based on the explanation given in this section, a new model is proposed to automatically generate the rules representing the transition based on learning from the sensors' data.

## 5.3 Long Short-term Memory Fuzzy Finite State Machine

To improve the learning capability of the FFSM, an integration of Long Short-Term Memory and Fuzzy Finite State Machine is proposed. A detailed explanation about the LSTM is provided in Chapter 3, whereas the enhanced LSTM-FFSM is introduced in this section. Long Short-Term Memory (LSTM) is a recurrent neural network is highly recommended for time-series sequential problems (e.g., human activity modelling and recognition), as it has the ability for remembering the previous input for each iteration throughout the training phase. Its architecture has a unique characteristic represented as a forget gate, that allows it either to remember or forget the previous input information and consider them in the next loop calculation. Comparing the LSTM with some other existing probabilistic approaches such as Hidden Markov Models (HMM), HMM is much simpler than LSTM, and rely on strong assumptions which may not always be true. However, LSTM performs better when it applies to a very large dataset since the extra complexity can take better advantage of the information in the used data.

Figure 5.2 illustrates the schematic diagram of the proposed FFSM using LSTM, which consists of three different stages; data collection process, fuzzy feature representation, and the fuzzy finite state machine modelling. In the data collection stage, the data from sensors in an AmI environment representing the ADL is collected. The fuzzy feature representation stage is designed to transform the data into fuzzy features to be used as inputs to the proposed FFSM model. In the third stage, the proposed FFSM model will generate the fuzzy rules employing the capabilities of learning algorithms in LSTM to representing the states' transitions.

As the LSTM is a particular kind of Recurrent Neural Network (RNNs), it was designed to solve vanishing and gradients problems in pattern recognition and predictions [72, 123]. The LSTM is a powerful tool for learning the sequential tasks that are represented as temporal data. It can also remember previous information for long periods. These characteristics make LSTM especially useful for temporal data classification problems [23]. The LSTM cell consists of three gating mechanisms to provide the ability to remove or add information to the memory cell. These three gates are used to regulate the impact of the input through an input gate, the previous cell state through a forget gate and an output through the output gate. The essential gate in the LSTM cell is the forget gate as it decides if the information is going to be remembered or be forgotten from the previous states.

The LSTM-FFSM is an enhanced version of FFSM, allowing the system to learn the temporal relations in the data by storing the information through the time-sequential steps. The learned relations are then used to formulate the fuzzy rules that control the transitions between the system's states and to identify the currently activated states at any given time. In this approach, the experts are also allowed to introduce their knowledge over the whole system by defining the following aspects:

- Defining the system states.
- Specifying the general structure of the fuzzy rules that represent the state transitions.

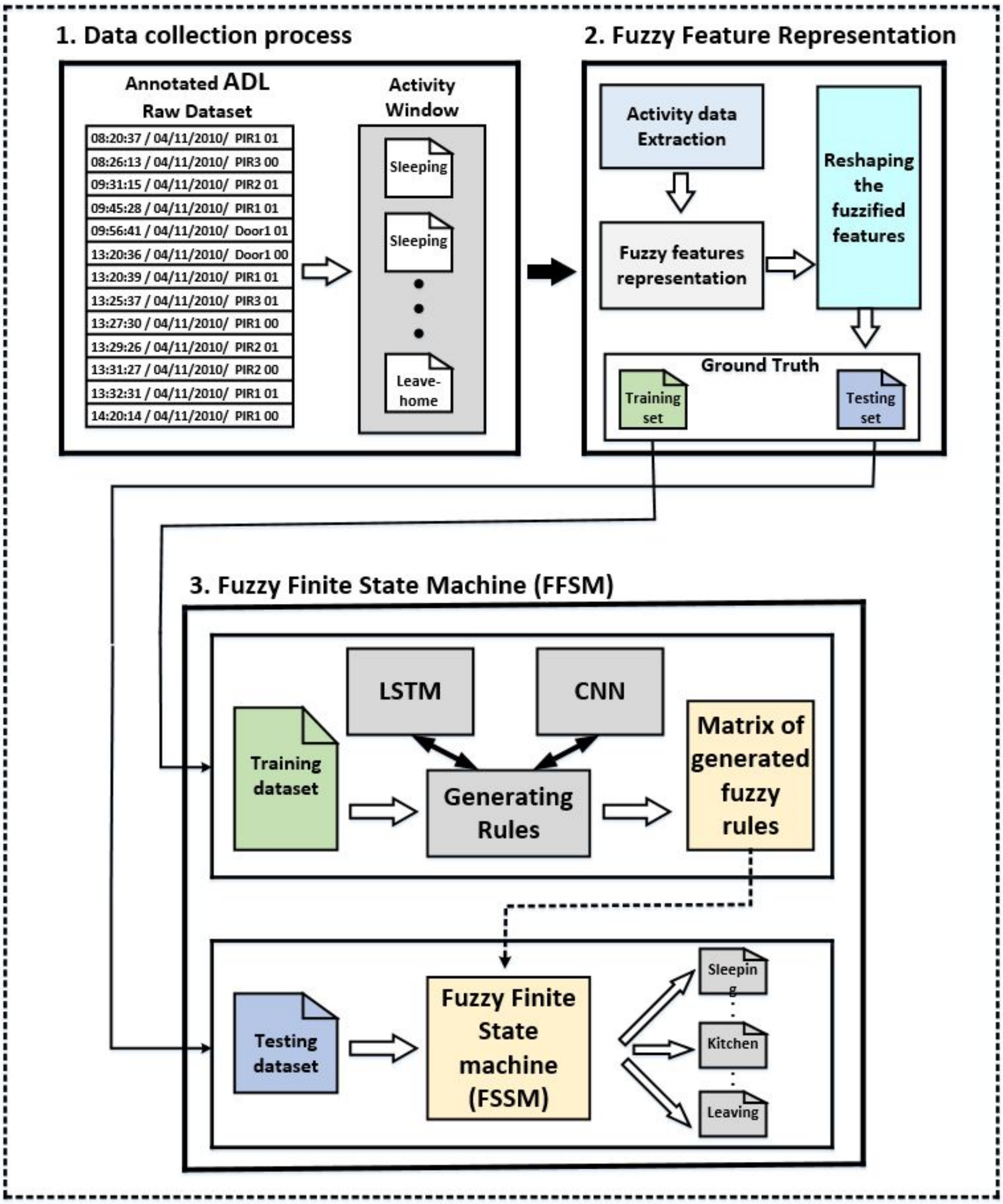



Figure 5.2: A schematic diagram of the enhanced Fuzzy Finite State Machine models using Long Short-term Memory neural network and Convolutional neural network.

- Specifying the number of linguistic labels that are associated with each input variable.

In a typical FFSM, a rule to identify the transition between state $m$ and state $n$ was presented as $R^{\lambda}_{mn}$ in Chapter 3. This demonstrates the relation between the system's current state $S(t)$ and the input variables that are represented as $H_{mn}$ to identify the next state $S(t+1)$.

In this version of the LSTM-FFSM, the term $H_{mn}$, involves the output values from the LSTM $Z_t$ to learn the temporal relation between the inputs and outputs, as well as the fuzzy input set $X_{uj}$. At this point, LSTM is employed with the fuzzy set input $X_{uj}$ at time $t$ to learn the temporal relations in the data by storing the inputs variables $X_{uj}$ in a time-sequential manner $t$. Thus, the term $H_{mn}$ will be represented as:

$$H_{mn} = X_{uj}(t) + Z_t \tag{5.1}$$

where $Z_t$ is the output of the LSTM at time $t$ and $X_{uj}(t) = [\mu_{A^1_{uj}}, \mu_{A^2_{uj}}, \ldots, \mu_{A^M_{uj}}]$ ; $M \neq 0$ is the fuzzified input at time $t$. $M$ is the number of associated linguistic labels that represented as membership degrees $\mu A_{uj}$ with the input $uj$. During the training phase of the LSTM, one input of the LSTM is used for the fuzzy set input $X_{uj}$ at time $t$, and another input is used for the current state $S(t)$. The output of the long short-term memory $(Z_t)$ is used for determining the second part of the antecedent term of the rule $H_{mn}$ for the next state $S(t+1)$ based on the general structure of the fuzzy rule as:

$R^{\lambda}_{mn}$**: IF** $(S(t)$ **is** $s_m)$ **AND** $H_{mn}$ **THEN** $S(t+1)$ **is** $s_n$

The output $Z_t$ is fed back to the current state input $S(t)$ through a delay unit $(D)$, once the training is completed. Thus, the LSTM-FFSM system's architecture is changed from a feed-forward structure to a recurrent architecture during the recall mode. Also, in this version of the LSTM-FFSM, there is no need to use separate outputs for the next state $S(t+1)$ and the LSTM output $H_{mn}$, since a Moore type of FFSM is implemented.

Based on the explanation introduced in this section, LSTM-FFSM is proposed to generate the fuzzy rules representing the transition based on learning the relations in the sequential temporal data. Therefore, the obtained output sequence from the calculation process using the LSTM is used with fuzzy set of input $X_{uj}$ at the time $t$ to compute the term $H_{mn}$. The obtained parameters in the term $H_{mn}$ are used to demonstrate the fuzzy rule $R^{\lambda}_{mn}$ that governs the transition between state $m$ and state $n$.

## 5.4 Convolutional Fuzzy Finite State Machine

The integration of CNN with FFSM is also proposed for enhancing the learning capabilities of the FFSM. This is achieved by selecting the most effective features to learn the relationship between the inputs and outputs data. Following the detailed explanation about CNN provided in Chapter 3, the Convolutional-Fuzzy Finite State Machine (CNN-FFSM) is introduced in this section. Figure 5.2 illustrates the schematic diagram of the proposed FFSM using CNN. This consists of three stages; data collection using ambient sensory device-based, fuzzy feature representation, and the fuzzy finite state machine. In the data collection stage, the data from sensors in the AmI environment representing the ADL is collected. The fuzzy feature representation stage is designed to represent the extracted features from the occupancy signal as a fuzzy set to be used as inputs to the proposed FFSM model. In the third stage, the proposed FFSM model will generate the fuzzy rules employing the capabilities of learning algorithms in CNN to representing the states' transitions.

Generally, the input data to a CNN is a matrix $c$ in dimensions of $h \times w \times d$, where $h, w$ and $d$ are the height, width and the number of channels in the input matrix $c$ [5, 88]. When the input window has only one class, the number of channel $d$ is 1.

The typical CNN architecture has two convolutional layers or more and one fully-connected layer. Each convolutional layer contains multiple feature filters to optimise convolutional values during the training phase. Each convolutional layer is followed by a max-pooling layer that has a window in a specific size to ensure the outputs from each conventional layer are smaller than the inputs. Rectified Linear Unit (ReLU) is added after each convolutional layer that operates as an activation function. The fully-connected layer in this architecture is a traditional Multi-Layer Perceptron (MLP) that operates a softmax activation function for the output layer. By using the softmax activation function for the output layer, the CNN classifier model will be able to classify the input features into various classes based on the learned relations during the training stage.

In the case of highly complex input data, the CNN architecture can contain more than one pair of the convolutional and max-pooling layers with different sizes of border filters to process the data [88]. Also, the top convolutional layer is followed by one or more fully-connected layers for the final classification purpose. During the training phase, the standard forward and backward propagation algorithms are used to estimate the values of the CNN parameters. The selected features are mapped by the convolutional operator [5] as follows:

$$V_t = \frac{1}{1 + exp(d_\eta + \sum_\iota \kappa_{\iota\eta} \vartheta x_\iota)} \tag{5.2}$$

where $\vartheta$ denotes the convolutional operator, $\kappa_{\iota\eta}$ is the convolutional filter for the $\iota$-th input, $V_t$ is the generated $\eta$-th output which is achieved by selecting the most effective features over the non-overlapping pooling regions from the input data $x_\iota$ and $d_\eta$ denotes the bias.

As the system's states are identified, the extracted features from the occupancy signal representing the ADL will be fuzzified. The generated fuzzy sets are used as inputs to CNN. CNN is employed for allowing the system to select the most effective features from the inputs (fuzzy sets) and then learn the temporal relations from the selected features by storing the information through the time-sequential steps. The learned relations are used to formulating the fuzzy rules used to control the system's state transitions and identify the current activated states at any given time $t$. In this approach, the experts are also allowed to introduce their knowledge over the whole system by defining system's states, the general structure of the fuzzy rules, and the number of associated linguistic labels where each input variable.

In this version of the FFSM, the system's states transition will be govern by fuzzy rules identified as follows:

$$R^{\lambda}_{mn}\textbf{: IF } (S(t) \textbf{ is } s_m) \textbf{ AND } H_{mn} \textbf{ THEN } S(t+1) \textbf{ is } s_n$$

Before staying at the current state $s_m$ or moving to the next state $s_n$ is finalised. During the training phase, the used CNN is employed with the fuzzy set $X_{uj}$. The output the used CNN is $(V_t)$ is obtained by selecting the most effective features over the non-overlapping pooling regions from the fuzzy input set $X_{uj}$. The output of the convolutional neural network $(V_t)$ is used for determining the second part of the antecedent term of the FFSM rule $H_{mn}$ for the next state $S(t+1)$. The output $V_t$ is fed back to the current state input $S(t)$ once the training is completed.

As mentioned in the FFSM explanation earlier, the rule $R^{\lambda}_{mn}$ is used to control the transition between state $m$ and $n$. The rule represents the relation between the system's current state $S(t)$ and the input variables that are represented as $H_{mn}$. The final value obtained from this calculation is used to identify the next state $S(t+1)$. As each input variable involved in the term $H_{mn}$ is calculated using the used CNN to learn the relations in the inputs (features) and outputs (labels) data by selecting and mapping the most effective features. Therefore, the term $H_{mn}$ will be represented as:

$$H_{mn} = X_{uj}(t) + V_t \tag{5.3}$$

where $V_t$ is the generated $\eta$-th output using CNN at time $t$ and $X_{uj}(t) = [\mu_{A^1_{uj}}, \mu_{A^2_{uj}}, \ldots, \mu_{A^M_{uj}}]$ $M \neq 0$ is the fuzzified input dataset at time $t$. $M$ is the number of membership degrees $\mu A$ that are associated with the input $uj$. Based on that, CNN is used as state-of-the-art to select the most effective features from the input dataset and learn the relations between the inputs (selected features) and the outputs (labels). At this stage, the parameters that are learned from employing the CNN will be used to generate the fuzzy rules $R^{\lambda}_{mn}$ for representing the transition between the system's states (activities) in the proposed CNN-FFSM.

The proposed N-FFSM, LSTM-FFSM, and CNN-FFSM are employed to learn the unknown parameters for generating the fuzzy rules representing the transition between the FFSM's states. The next section introduces experiments with the proposed approaches, which integrates the learning abilities of the NN, LSTM and CNN by selecting and mapping the most effective features in the temporal

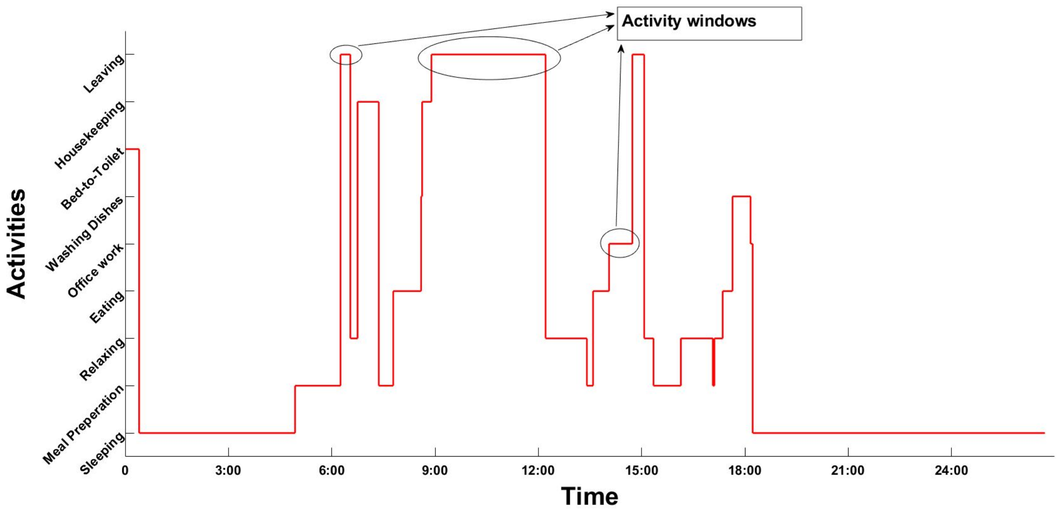



Figure 5.3: An illustration of activity windows for one-day activities.

datasets representing daily human activities.

## 5.5 Experimental Setup

To evaluate the performance of the proposed N-FFSM, LSTM-FFSM and CNN-FFSM explained in Sections 5.2, 5.3 and 5.4, respectively, experimental works are conducted where two different datasets representing the ADL for a single user are used for modelling and recognising the user's activities. The experimental setup is presented below, and all results are presented in the Experimental Setup and Experimental Results sections respectively. The two datasets referred to as Dataset A and Dataset B are used to evaluate the proposed approaches for human activity modelling and recognition. Details of these datasets are provided in Chapter 4.

### 5.5.1 Feature Extraction and Fuzzification

Extracting the numerical information from the acquired raw sensor data is crucial to any learning system since raw data does not provide adequate information that can be used as inputs to the model. The collected data was gathered from low-level ambient sensory devices; it will be saved to a database as time-stamped binary data. The gathered raw data would be represented and interpreted using FSM approach explained in Chapter 4, to convert the collected binary data into an occupancy signal representing each activity as an activity window as it is shown in Figure 5.3. To fuzzify the activity data for each activity-window start time, end time, duration, activity count and activity sequential order are extracted. Therefore, the activity data are obtained for each activity window and represented as a matrix where rows are the length of the activity window and columns are the number of recorded information from the sensors in the window.

The extracted information from each activity-window is mapped into a numerical activity data representing the input variables $U(t) = [u_1(t), \ldots, u_j(t), \ldots, u_P(t)]$, where $P$ is the number of input variables. An overall framework of the used approach for representing the activity data as fuzzy features is illustrated in Figure 5.4. In this figure, the gathered information using ambient sensory devices is presented as binary strings. These binary strings are converted into an occupancy signal where each activity can be segmented as an activity window $W$. The number of numerical features is extracted from each activity window, including the activity's start time, end time and duration. A fuzzy feature representation approach is employed to represent each numerical value as a fuzzy set. In particular, each value in the input variable $u_j$ is represented with the relevant membership values to each fuzzy set. The process of fuzzifing these extracted numerical values are occurred by using Gaussian MFs. Five different MF degrees are used to convert each value in the start $u_1$ time, end time $u_2$ and duration $u_3$ variables into their relevant membership degrees as illustrated in Figure 5.4. Therefore, every single value from these variables is represented by the relevant number of belonging degrees to each MF. The linguistic labels associated with each input are represented as MFs. These MFs are described as follows:

$$X_U(t) = \begin{cases} X_{u1}(t) \rightarrow \{EM_{u_1}, M_{u_1}, AF_{u_1}, EV_{u_1}, NI_{u_1}\} \\ X_{u2}(t) \rightarrow \{EM_{u_2}, M_{u_2}, AF_{u_2}, EV_{u_2}, NI_{u_2}\} \\ X_{u3}(t) \rightarrow \{VS_{u_3}, SH_{u_3}, ME_{u_3}, LO_{u_3}, VLO_{u_3}\} \end{cases} \tag{5.4}$$

where $X_U(t)$ is the input vector of fuzzified variables $\{X_{u1}(t), X_{u2}(t), X_{u3}(t)\}$ to

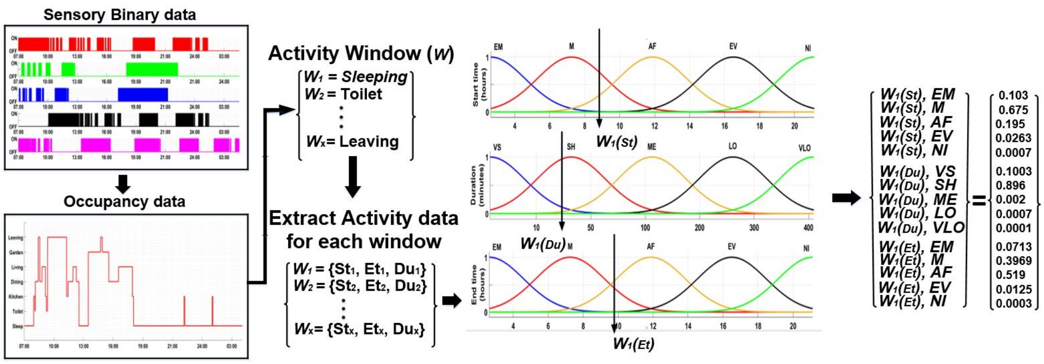



Figure 5.4: Overall framework of the used fuzzy feature representation approach.

the system at time $t$. The linguistic labels that are associated with the MFs are explained as:

- The MFs representing activity start time for the input variable $u_1$ are represented as $\{EM_{u_1}, M_{u_1}, AF_{u_1}, EV_{u_1}, NI_{u_1}\}$. Where $EM$, $M$, $AF$, $EV$ and $NI$ are MF labels corresponding to Early Morning, Morning, Afternoon, Evening and Night respectively.

- The MFs representing activity end time for the input variable $u_2$ are represented as $\{EM_{u_2}, M_{u_2}, AF_{u_2}, EV_{u_2}, NI_{u_2}\}$. Where $EM$, $M$, $AF$, $EV$ and $NI$ are the MF labels corresponding to Early Morning, Morning, Afternoon, Evening and Night respectively.

- The MFs representing activity duration for the input variable $u_3$ are represented as $\{VS_{u_3}, SH_{u_3}, ME_{u_3}, LO_{u_3}, VLO_{u_3}\}$. Where $VS$, $SH$, $ME$, $LO$ and $VLO$ are MF labels corresponding to Very Short, Short, Medium, Long and Very long respectively.

The other two variables representing activity count $u_4$ and activity sequential order $u_5$ will not be fuzzified together with the other activity data. They will be normalised and then added to the fuzzy represented features before the entire set of the input data $X_U(t)$ is fed into the models. The final set of the fuzzified features $X_{uj} = [\mu A^1_{uj}, \mu A^2_{uj}, \ldots, \mu A^M_{uj}]$, will be used as inputs to train the proposed models for modelling and recognising the activities.

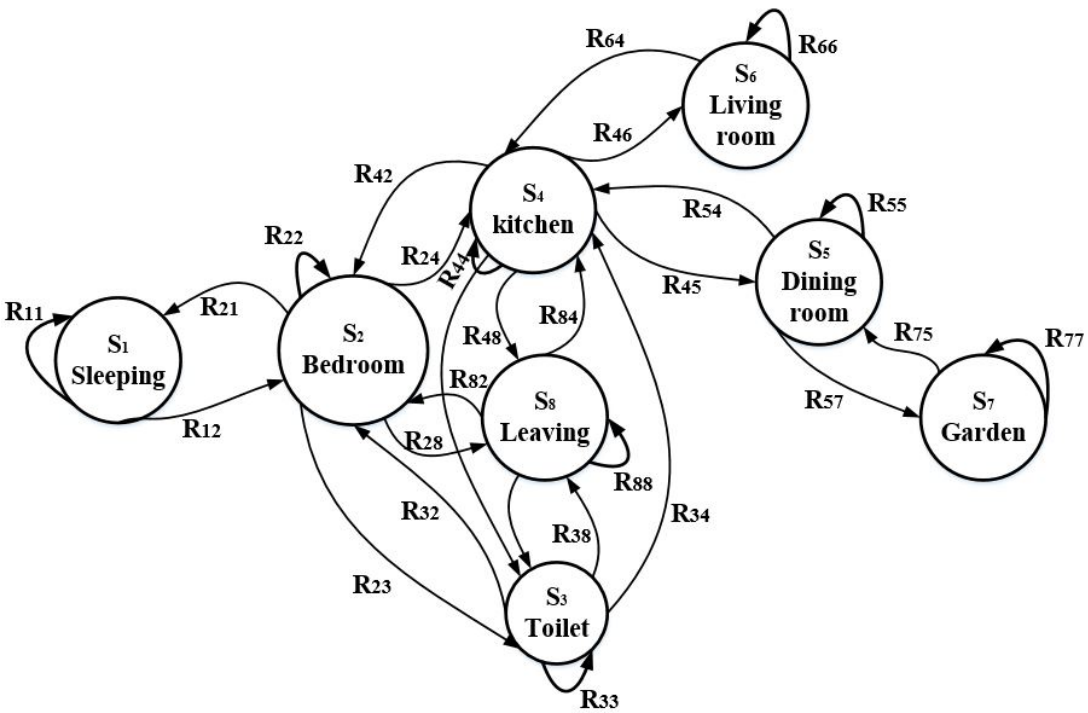



Figure 5.5: A state diagram of human activity's based on experimental datasets A.

### 5.5.2 System definition

The collected data represent 7 and 11 different activities in Dataset A and Dataset $B$, respectively. Each activity is represented as one state in the FFSM model. These states are defined based on the experts' knowledge. This is easily represented using the proposed state diagram illustrated in Figures 5.5, and 5.6 for data set A and B, respectively. These states in Datasets $A$ and $B$ are defined as follows:

**States representing activities in Dataset A:**

- $s_1$: Sleeping State, represents sleeping activity either during the night or taking a nap during the daytime. Intuitively, the collected starting time and the duration of this activity could vary depending on the days of the week, even for the same user. Also, the state can be interrupted by other activities such as going to the toilet, etc.

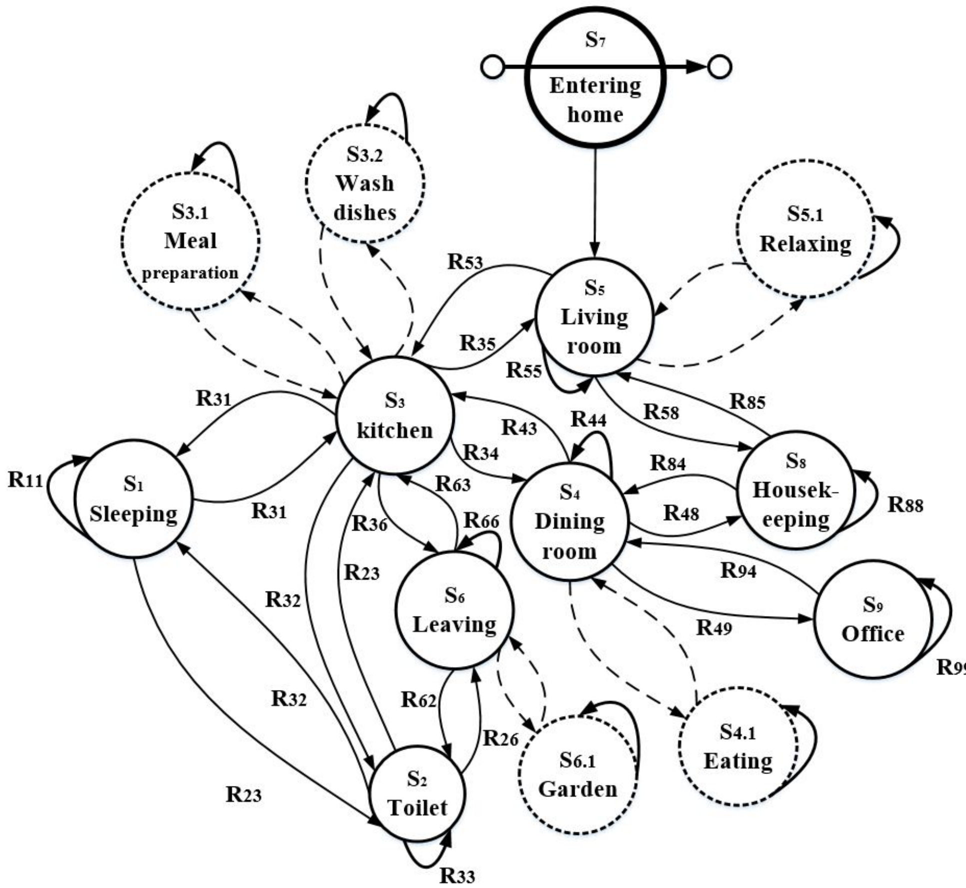



Figure 5.6: A state diagram of human activity's based on experimental Datasets B.

- $s_2$: Bedroom State, this state is used to represent the other duties in the bedroom except for the sleeping activity. This state will be considered as sleeping state $S_1$ in this experiment.
- $s_3$: Toilet State, this state represents the times when the user is using the toilet.
- $s_5$: Kitchen State, where the user spends time in the kitchen for preparing food or for cleaning.
- $s_5$: Dining Room State, this state usually comes after the kitchen state, when the user stays in the dining room to eat the prepared food.

- $s_6$: Living Room State, this state corresponds to the time spent in the living room for watching TV or other social activities.
- $s_7$: Garden State, this state is used when the user uses the back door to go to the garden.
- $s_8$: Leaving Home State, this state becomes active when the individual leaves home from the front door. This can be for any of outdoor duties such as shopping. This state might be occurred regularly at a specific time (in case of the individual having a daily job) or irregularly (in the case of shopping and social visiting).

**States representing activities in Dataset B:**
States representing the Dataset B are as follows:

- $s_1$ : Sleeping states to represent the sleeping activities.
- $s_2$ : Bed-to-Toilet state to represent the times of using the toilet in between the sleeping time.
- $s_{3.1}$ : Meal preparation state to represent the event of preparing food. This state is the first part of the kitchen state.
- $s_{3.2}$ : Washing dishes state to represent the event of washing dishes in the kitchen area. This state is the second part of the kitchen state.
- $s_{4.1}$ : Eating state to represent the time when the user at the dining room. This state is usually activated after the Meal preparation state.
- $s_{5.1}$ : Relaxing state to represent the time spent in the living room.
- $s_6$ : Leaving state to represent the time when the user leaves the house. As the house has three different doors, this state will be activated when any of these doors are used.
- $s_7$ : Entering home state to represent the time when the user comes back home.

- $s_8$ : Housekeeping state to represent cleaning work, e.g., hoovering the carpet.

- $s_9$ : Office-work state to represent the event of doing some homework in the office room.

Once the system's states are created, a set of fuzzy rules is required to control the transition between the system's states. In the standard FFSM, these rules are defined based on the experts' knowledge only. In this contribution, as the generated data is temporal data representing sequential order events, NN, LSTM and CNN are employed to learn the relationships in the data through the time steps. The learned relations are used to generate fuzzy rules in the system. The final output for this model, $Y(t)$, is represented as the degree of belonging to each state in the system.

### 5.5.3 Experimental Results

The results obtained from the conducted experiments are presented here. As humans behave with some unpredictability and uncertainty in their environment, datasets representing the human activities are usually imbalanced, where some activities appear more dominant than the other activities. In that case, if the dominant activities are identified with a high degree of certainty, the performance over the whole system will be high even if the other activities are not well identified. Therefore, each activity will be evaluated separately, then the performance over the whole system will be calculated. The enhanced proposed models, namely N-FFSM, LSTM-FFSM and CNN-FFSM, are tested and evaluated using the two earlier mentioned Dataset $A$ and Dataset $B$.

The confusion matrices in Figures 5.7, 5.8 and 5.9 show the recall (known as sensitivity) and precision scores obtained using the proposed N-FFSM, LSTM-FFSM and CNN-FFSM models respectively, for each activity using Dataset A and Dataset B. As well as the accuracy over the whole models. In each of the figures, the obtained results are labelled as (a) when the proposed models are applied to Dataset A, and (b) when they are applied to Dataset B.

The information given in the confusion matrices is explained as follows:

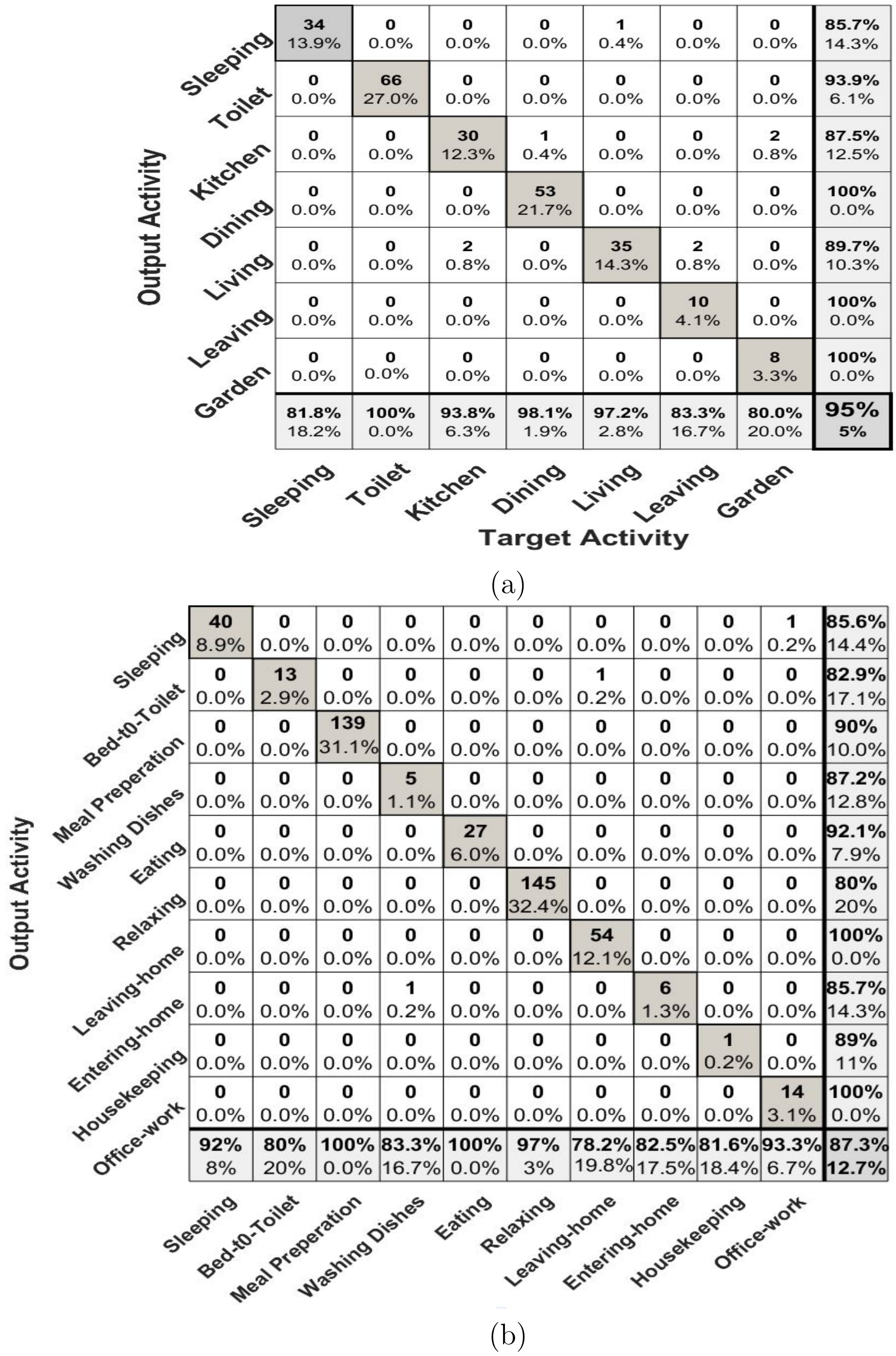
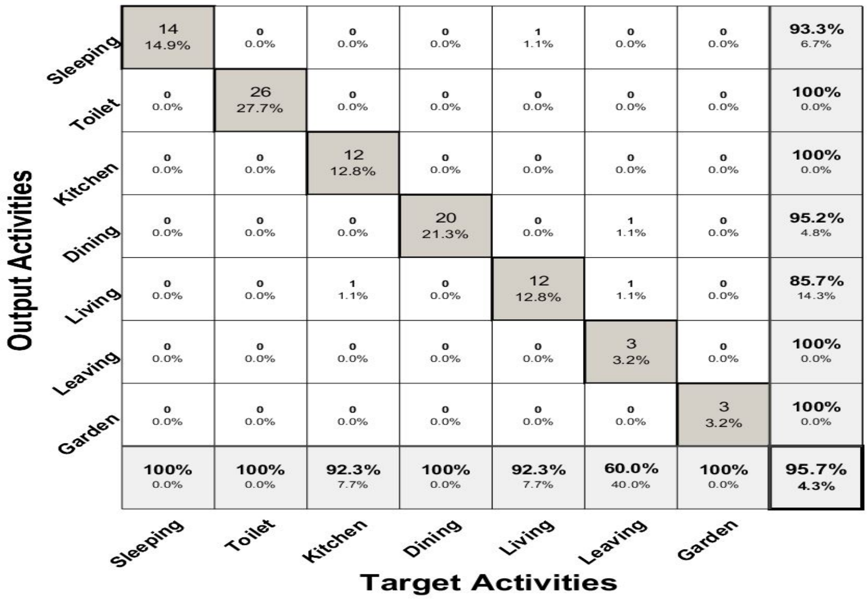
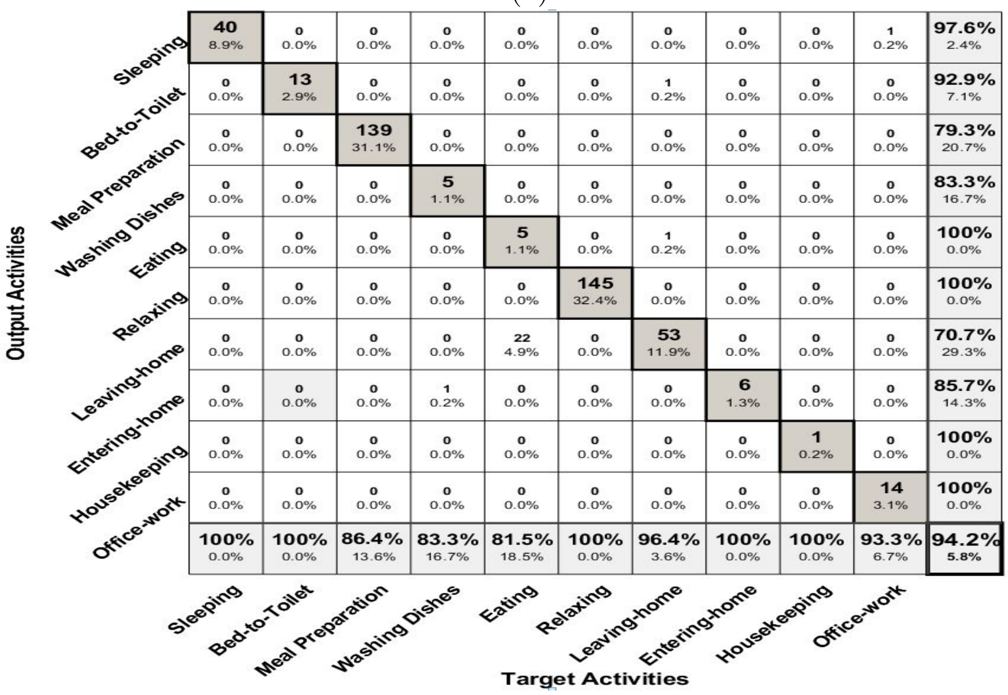



Figure 5.7: Confusion matrix for ADL modelling and recognition results using Neuro Fuzzy Finite State Machine (N-FFSM); a) using Dataset A, b) using Dataset B.




(a)




(b)

Figure 5.8: Confusion matrix for ADL modelling and recognition results using Long Short-term Memory Fuzzy Finite State Machine (LSTM-FFSM); a) using Dataset A, b) using Dataset B.

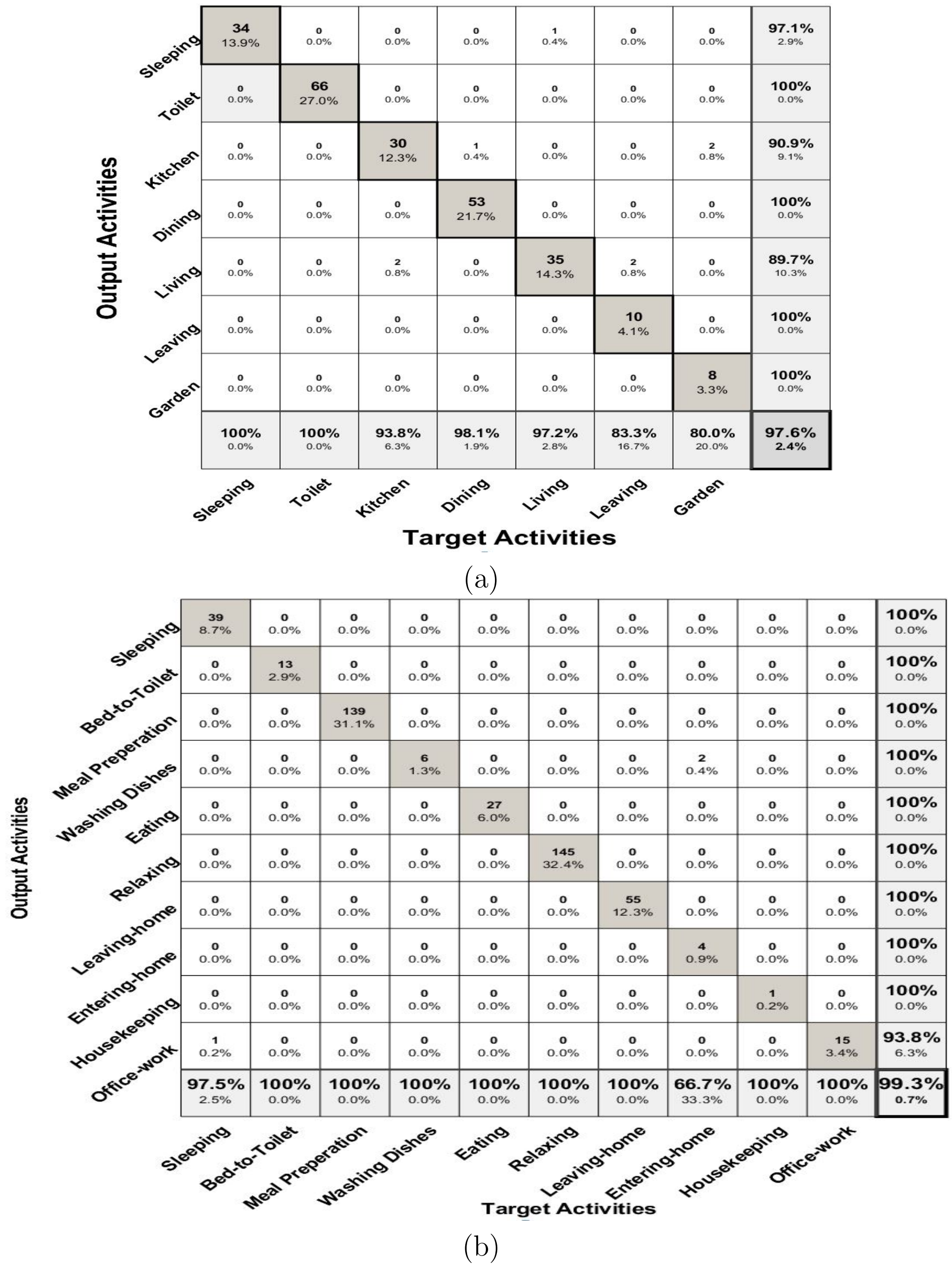


Figure 5.9: Confusion matrix for ADL modelling and recognition results using Convolutional Neural Network Fuzzy Finite State Machine (CNN-FFSM); a) using Dataset A, b) using Dataset B.

- The rows represent the output activities, and columns represent the target activities. The activities in Dataset A are named as *Sleeping, Toilet, Kitchen, Dining, Living, Leaving home, and Garden.* The activities in Dataset B are named as *Sleeping, Bed-to-Toilet, Meal preparation, Washing dishes, Eating, Relaxing, Leaving-home, Entering-home, Housekeeping, and Office-work.*
- The diagonal cells from the upper left to the lower right illustrate the activities that are correctly recognised.
- The off-diagonal cells present the incorrectly recognised activities.
- The precision for each activity is presented in the last column in the right.
- The recall for each activity is presented in the last row at the bottom.
- The accuracy over the whole model is illustrated in the bottom-right cell.

As it can be seen from these figures, CNN-FFSM model is more efficient when it is applied to a larger dataset containing uncertain activities such as those that could occur at the same place (e.g., *Meal Preparation activity* and *Washing Dishes activity*), both of which are being undertaken in the kitchen. In Figure 5.9(b), nine out of ten activities, including in Dataset B, are recognised with 100% precision.

To evaluate the proposed models, the results obtained in the experiments employing the different version of the enhanced FFSM models are compared with the results obtained using six different existing approaches for modelling and recognising human activities such as LSTM, SVM and NNs with both Datasets A and B. These datasets are divided into 70% for training stage and 30% testing stage randomly on a daily basis. The comparison between the performance of the proposed N-FFSM, LSTM-FFSM and CNN-FFSM with the other existing methods has been made for the accuracy over the whole models in Table 5.1. The overall performance for the proposed models based on Dataset A's testing subset is 95.7% and 94.2% obtained by applying LSTM-FFSM and CNN-FFSM models, respectively. 97.6% and 99.3% are the obtained results based on Dataset B's testing subset using LSTM-FFSM and CNN-FFSM models, respectively. On the other hand, the SVM and NNs models tend to have the lowest performance compared to other evaluated models. Hence, the optimised SVM uses the default kernel function, which is linear and this could be the reason of this low-performance ability with collected data. Similarly, the number of examples for the complexity problem could be not enough in the provided training dataset for the NN model.

Furthermore, the generalisation ability of the proposed approaches has been verified with different data splitting configuration. In this validation experiment, the dataset is divided into training and testing subsets using k-folds cross-validation. Table 5.2 shows the performance of the proposed models with 5-folds cross validation. It can be concluded from these results that the proposed models have a high-generalisation ability, and data overfitting is avoided.

The expressions that are used to calculate accuracy, precision and recall for each activity are given below:

$$Recall = \frac{1}{C}\sum_{i=1}^{C}\frac{TP_i}{TP_i + FN_i} \tag{5.5}$$

Table 5.1: The overall accuracy of the proposed LSTM-FFSM and CNN-FFSM compared with the existing approaches based on Datasets A and B.

| Methods | Dataset A | Dataset B |
|---|---|---|
| CNN-FFSM | 94.2% | **99.3**% |
| LSTM-FFSM | **95.7**% | 97.6% |
| N-FFSM | 95.2% | 94.5% |
| LSTM | 80.6% | 78.1% |
| Bi-LSTM | 83.7% | 81.4% |
| NNs | 78.7% | 76.8% |
| FFSM | 46.6% | 41.3% |
| SVM | 62.8% | 60.9% |

$$Precision = \frac{1}{C} \sum_{i=1}^{C} \frac{TP_i}{TP_i + FP_i} \tag{5.6}$$

$$Accuracy = \frac{1}{N} \sum_{i=1}^{C} TP_i \tag{5.7}$$

where $TP_i$, $TN_i$, $FN_i$ and $FP_i$ are the number of true positives, true negatives, false negatives and false positives of $i^{th}$ activity respectively. $N$ is the number of the values $TP_i + TN_i + FP_i + FN_i$ for $i^{th}$ activity. $C$ is the activity of which its recall, precision and accuracy are calculated.

Considering the interpretability point of view, the most commonly used approaches for modelling and recognising human activities are the approaches based on mathematics, e.g., NNs and SVM. These models are well known as black-box approaches because of the complexity of understanding their underlying calculations and concepts. This complexity will be more challenging when a large number of input and output variables are expected. Moreover, designing a model that is only based on the linguistic information assigned by human experts is not enough for a successful and robust human activities modelling and recognition. Therefore, the advantages of integrating the experts' knowledge with the learning capabilities in LSTM and CNN can be integrated into the proposed LSTM-FFSM and CNN-FFSM models for generating a successful and robust model.

The obtained results are compared with some previous works, such as the research presented in [149]. They have used a dataset collected using ambient

Table 5.2: Overall accuracy performance for the proposed models using 5-folds cross validation.

| Methods | Dataset A | Dataset B |
|---|---|---|
| CNN-FFSM | 90.7% | **97.4%** |
| LSTM-FFSM | **98.1%** | 96.2% |
| N-FFSM | 92.3% | 96.7% |

sensory devices to discover 5 different activities; *Telephone use, Hand washing, Meal preparation, Eating*, and *Cleaning*. A new mining method, called Discontinuous Varied-Order Sequential Miner (DVSM), is used in their research with the collected dataset to find frequent patterns that may be discontinuous and might have variability in the ordering of this activities. The achieved results based on the DVSM is 77.3%.

In a recent publication for recognition of interleaved human activities, researchers have proposed a new human activity model containing three phases, namely Preprocessing, Discovery Method for Varying Patterns (DMVP) and Predictive Modelling [93]. The first phase is used to convert the collected raw sensor data into event sequences, which are then fed to DMVP in the second phase to discover the frequent activities that naturally happen during the regular daily routine, and then a classification model is applied to predict the activities in the third phase. The achieved result from this approach is 87.94% once it is evaluated with the CASAS dataset.

## 5.6 Discussion

In this chapter, novel integration of the learning techniques with the standard FFSM for enhancing the performance of human activity modelling and recognition. The main goal is to enable the modelling and recognition of the activities based on binary data collected by ambient sensory device-based representing ADL for a single user in a smart home environment. Due to the dynamic nature that human behaves, there are uncertainties associated with modelling and recognising their activities. This work focused on proposing different versions of FFSM models capable of adapting to and learning the variations that exist in the human activity sequential data. The use of Fuzzy Feature Representation approach to obtain the fuzzy sets representing the human activity makes it possible to acquire representations of activities for learning such process.

We leverage the ability of NN, LSTM and CNN techniques in learning essential parameters from the activity sequences for generating the FFSM's rules to govern the states' transition. Also, human experts' knowledge still in use to create the fuzzy states (here are the activities) and the number of MFs used to fuzzify the input data. Furthermore, considering the results obtained from the conducted experiments, it can be concluded that the LSTM-FFSM and CNN-FFSM models exhibit a high score for accuracy, recall, and precision when its performance is tested for each activity separately. Also, the overall activity recognition performance, when it is over the whole system, demonstrates the effectiveness of the proposed approaches. The CNN-FFSM model shows a more robust and reliable performance once applied to a larger dataset (e.g., Dataset B) representing ten activities over 240 days. This is mainly the case when this dataset contains some activities that could be happening at the same place (such as *Washing Dishes activity* and *Meal Preparation activity* in the kitchen). In real-life scenarios, it is hard to know which activity is the current activity based on the data collected from the PIR sensory devices. Thus a fuzzy feature representation approach with CNN-FFSM can be used to deal with such cases as it can detect the changes in the fuzzy feature patterns. The essential feature of the proposed approach is that it integrates the available expert's knowledge with the learned information from the deep learning techniques. The LSTM-FFSM has shown better performance for a simple scenario once it is applied to a short period dataset (e.g., Dataset A). The CNN-FFSM achieved more accurate results to detect the ADL activities for a more extended period dataset (e.g., Dataset B). Besides, it can be seen how the proposed models can follow the proper sequence of states with the correct state activation degree.

# Chapter 6

# Employing Deep Learning Techniques for Human Activity Modelling and Recognition

## 6.1 Introduction

This chapter proposes the use of a Deep Convolutional Neural Network (DCNN) for human activity recognition using binary ambient sensors such as Passive Infrared (PIR) and door entry sensors. Each activity is represented as a binary string converted into a greyscale image. Unique features are selected and then used as inputs to an Adaptive Boosting (AdaBoost) and Fuzzy C-means (FCM) classifiers for recognising Activities of Daily Living (ADL). The performance of the proposed model is evaluated using the Dataset A explained in Chapter 4, representing the ADL for a single user. The results from the experiments in this chapter using the extracted features from the greyscale image representing ADL with AdaBoost and FCM algorithms achieved accuracy of 99.5% and 86.4%, respectively. Also, another deep learning technique such as Bidirectional Long Short-term Memory (Bi-LSTM) and CNN are employed for modelling and recognising human activities.

The proposed human activity classification and recognition model in this chapter employ a DCNN with fully annotated real data representing ADL for a single user living alone. Binary ambient sensors such are employed for the data collection process. First, the collected Dataset A is divided into training and testing sets on day-by-day basis, and then the activities in the training set are segmented and converted into greyscale activity image representing each activity. The greyscale image is considered as a 2D image used for representing the numerical measurements such as sensor data. A pre-trained DCNN is employed to extract a unique feature vector from the generated greyscale activity image. The extracted feature vector is used as an input for two different classifiers to classify and recognise the ADL for that user. The obtained results are then compared. The reason behind choosing the pre-trained CNN for extracting the feature vector representing each activity's unique features is because each activity is considered an image with spatial features. However, more simplistic approaches can be used for these applications, such as standard ANN. Still, the pre-trained CNN is more recommended for the data that has spatial features.

This chapter is structured as follows; In Section 6.2.1, the conducted experiments using the proposed DCNN for extracting the feature vector from the generated 2D greyscale image, and the proposed AdaBoost and FCM are explained. The final results obtained from the conducted experiments are discussed in Section 6.2.2. Other deep learning techniques including Bi-LSTM and CNN employed for human activity modelling and recognition are explained in Section 6.3 and Section 6.4, respectively. This is followed by pertinent conclusions drawn in Section 6.5.

## 6.2 Deep Learning Human Activity Modelling and Recognition Based on Greyscale Image

Recently, there are many research works on the application of Convolutional Neural Networks (CNNs) for Human Activity Recognition (HAR). Most of the published research is based on vision data (images or video stream) [150, 151], wearable devices [11] and binary sensors [4, 5]. The framework of the proposed method is shown in Figure 6.1. The proposed method contains three stages;

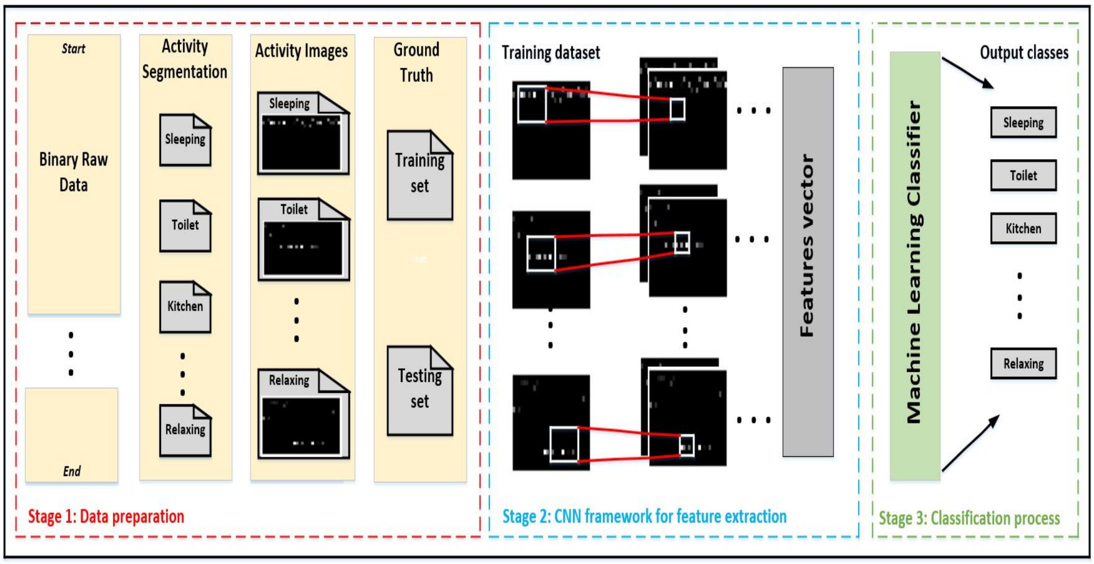


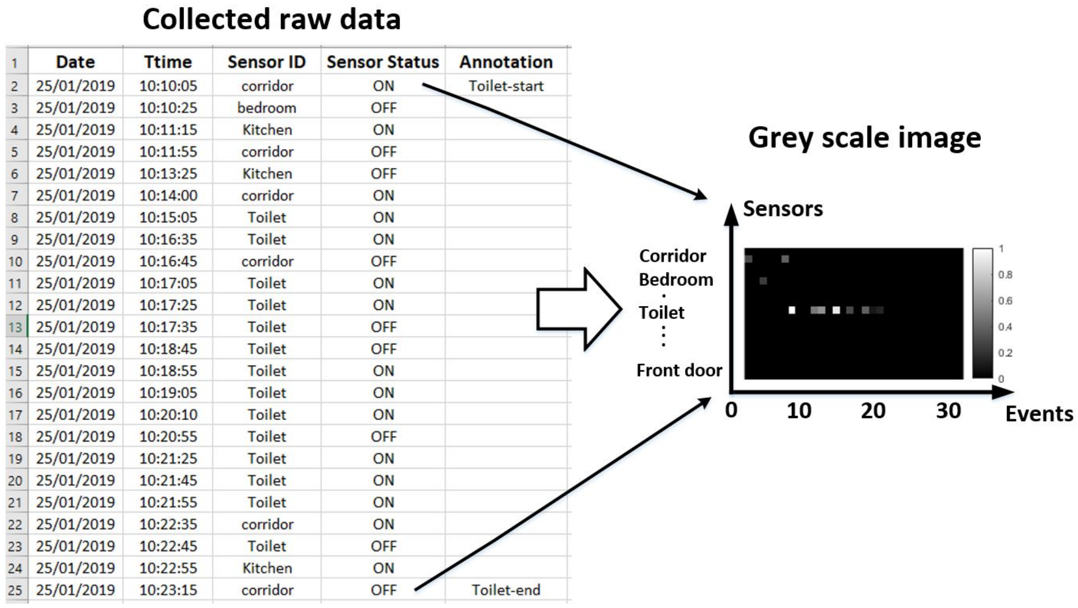

Figure 6.1: Framework for the proposed method including data preparation, feature extraction and the classification process.

data preparation, feature extraction, and the classification process. The data preparation stage includes the process of converting binary string data into greyscale images by representing each activity as an image with a black background and a grey range of pixels corresponding to each "0" and "1" sensory information, respectively. Then the second stage is the CNN framework that is used for extracting the important features from the generated images by employing a pre-trained AlexNet for extracting a feature vector that is used as an input to the proposed model in the next stage for classifying the activities. Finally, the used classification approaches are AdaBoost and FSM.

### 6.2.1 Experimental Setup

To evaluate the performance of the proposed approach presented in this chapter, an experiment for recognising and classifying a single user's activities is conducted. In particular, the techniques explained in the previous sections are employed with a real dataset, representing the ADL for a single user, recognising human activities based on the proposed techniques.

The dataset used in this experiment is Dataset A that was explained in




| 1 | Date | Ttime | Sensor ID | Sensor Status | Annotation |
|---|---|---|---|---|---|
| 2 | 25/01/2019 | 10:10:05 | corridor | ON | Toilet-start |
| 3 | 25/01/2019 | 10:10:25 | bedroom | OFF | |
| 4 | 25/01/2019 | 10:11:15 | Kitchen | ON | |
| 5 | 25/01/2019 | 10:11:55 | corridor | OFF | |
| 6 | 25/01/2019 | 10:13:25 | Kitchen | OFF | |
| 7 | 25/01/2019 | 10:14:00 | corridor | ON | |
| 8 | 25/01/2019 | 10:15:05 | Toilet | ON | |
| 9 | 25/01/2019 | 10:16:35 | Toilet | ON | |
| 10 | 25/01/2019 | 10:16:45 | corridor | OFF | |
| 11 | 25/01/2019 | 10:17:05 | Toilet | ON | |
| 12 | 25/01/2019 | 10:17:25 | Toilet | ON | |
| 13 | 25/01/2019 | 10:17:35 | Toilet | OFF | |
| 14 | 25/01/2019 | 10:18:45 | Toilet | OFF | |
| 15 | 25/01/2019 | 10:18:55 | Toilet | ON | |
| 16 | 25/01/2019 | 10:19:05 | Toilet | ON | |
| 17 | 25/01/2019 | 10:20:10 | Toilet | ON | |
| 18 | 25/01/2019 | 10:20:55 | Toilet | OFF | |
| 19 | 25/01/2019 | 10:21:25 | Toilet | ON | |
| 20 | 25/01/2019 | 10:21:45 | Toilet | ON | |
| 21 | 25/01/2019 | 10:21:55 | Toilet | ON | |
| 22 | 25/01/2019 | 10:22:35 | corridor | ON | |
| 23 | 25/01/2019 | 10:22:45 | Toilet | OFF | |
| 24 | 25/01/2019 | 10:22:55 | Kitchen | ON | |
| 25 | 25/01/2019 | 10:23:15 | corridor | OFF | Toilet-end |

Figure 6.2: The process of generating a greyscale image by converting raw data for the toilet activity.

Chapter 4. As this dataset was collected using ambient sensory devices, it contains a large volume of binary data. The challenge here is to represent this dataset in a suitable format that can be used for extracting features to train the classifier on a later stage.

Before any feature extraction process occurs, the collected sensor readings are represented as a greyscale image. The greyscale image is a 2D visual representation of the binary string data collected during a particular activity (e.g. sleeping, kitchen, etc.). The greyscale image representing a particular activity is an image with a black background and a grey range of pixels corresponding to each "ON" and "OFF" signal of the motion sensors and "OPEN" and "CLOSE" signals of the door sensors respectively. Figure 6.2 illustrates the process of generating a greyscale image from the segmented *Toilet* activity. The two dimensions of the greyscale image are the x-axis (length) and y-axis (height), representing temporal events of the activity and the sensor patterns of the activity, respectively. The greyscale range corresponds to the duration of each sensor event. Figure 6.3 illustrates random samples for three different activities.

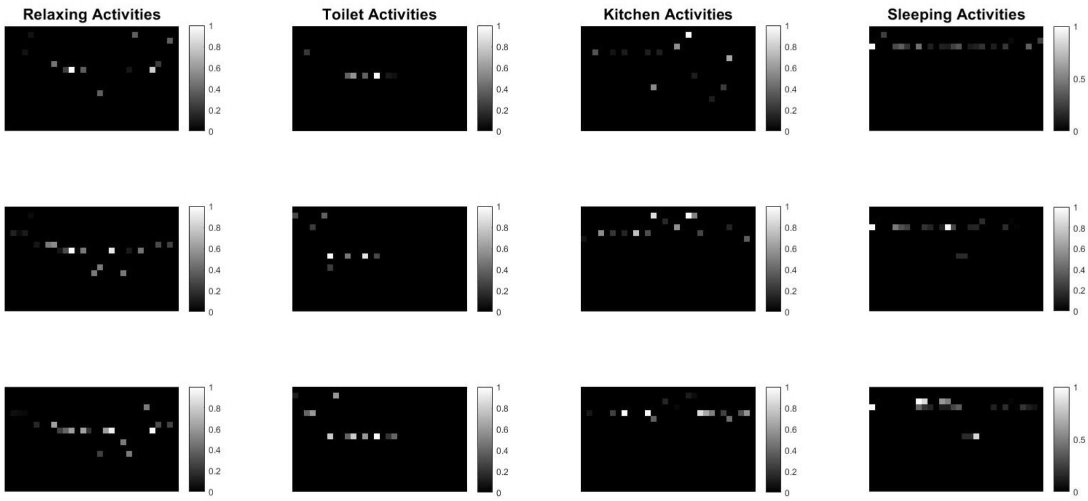



Figure 6.3: Three greyscale image samples for 4 different activities generated form the binary collected data using the proposed approach.

The AlexNet convolutional neural network [152], is applied to the generated greyscale images for the feature extraction process. The size of the feature extracted vector from each greyscale image representing one activity is 4096, computed at the first connected layer. The extracted feature vector is used as an input to AdaBoost and the FCM classifier.

Once the feature vector is extracted using the DCNN, the vector is used as inputs for the classification process. The extracted feature vector is divided into training and testing sets. The classifier is trained first using the training set, and then it will be tested and evaluated based on the testing set.

The last stage of the proposed system is where the extracted features vector is used as inputs to two different classification techniques. The gathered dataset represents eight different activities, and each activity is labelled separately as one class. These classes are defined as follows:

- *Class 1* → represents sleeping activity, either night sleeping or daytime napping.
- *Class 2* → represents the events when the user is using the toilet.
- *Class 3* → represents the time when the user is using the kitchen for preparing food or washing the dishes.

- *Class 4* → comprises relaxing activities, representing when the user spends time in the living room area for relaxing or watching TV.
- *Class 5* → represents dining activity, which usually comes after kitchen activity when the user is in the dining room to eat the prepared meal.
- *Class 6* → is bedroom activity, which represents when the user is in the bedroom for anything except sleeping (e.g., dressing).
- *Class 7* → is the garden activity, which represents the time when the user leaves home to go to the garden through the back door.
- *Class 8* → is leaving the home activity, which represents when the user leaves home through the front door.

Once the feature vectors are extracted from the generated greyscale image, the corresponding output classes are labelled for each activity. A 4096-dimensional feature vector is computed by cross-mixing in the first fully connected layer for the two groups' features. This process is repeated for the next fully-connected layer until the last fully-connected layer is reached. This is where the two feature groups are combined into a single feature vector for each greyscale image. The extracted feature vectors will be used as inputs to train the AdaBoost and FCM classifiers, based on the explanation provided in this section. The achieved results are presented and discussed in the next section.

### 6.2.2 Experimental Results

To test and evaluate the proposed approach, a real data set representing ADL for a single user is used. The results obtained from the conducted experiments are presented here. The gathered binary string data for each activity is represented as a greyscale image. AlexNet DCNN is applied to the generated images to extract a feature vector that will be used as input to train the AdaBoost and FCM classifiers in classifying and recognising human activity. The results obtained based on AdaBoost are illustrated as a confusion matrix plot in Figure 6.4. The same extracted feature vector is used as an input to the FCM classifier in classifying the ADL for a single user, and the obtained results are shown in Figure 6.5.

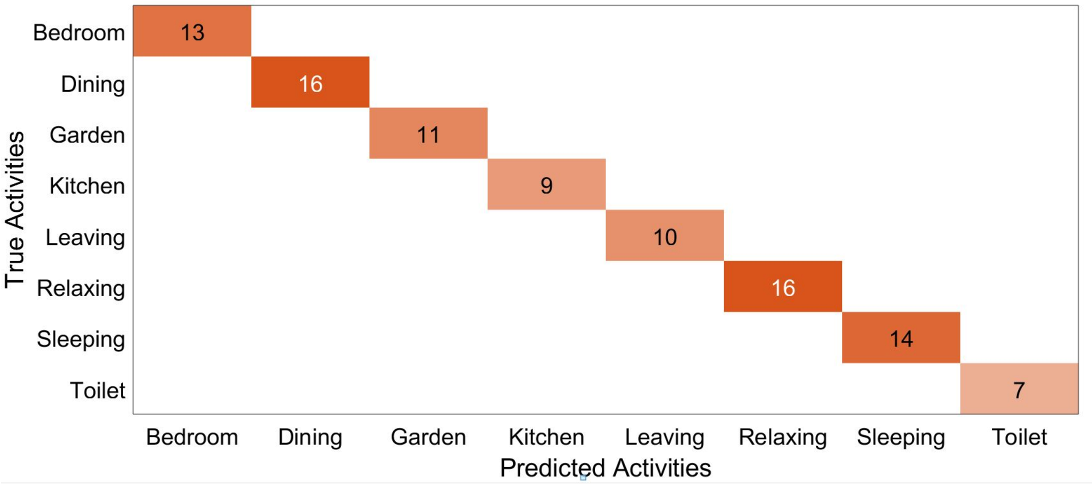



Figure 6.4: Confusion matrix plot representing the results achieved using AdaBoost approach.

As humans behave in unpredictable ways in their environment, datasets representing human activities are usually imbalanced, and therefore, some activities appear more dominant than the others. In this scenario, if one of the dominant activities are identified with a high degree of accuracy, the performance over the whole system will be high even if the other activities are not well identified. Based on that, the accuracy of each activity will be calculated separately, and then the performance over the whole system will be computed. Tables 6.1 and 6.2 show the accuracy, recall and precision for each

Table 6.1: The results of accuracy, recall and precision for each activity obtained based on AdaBoost.

| **Activities** | **Accuracy** | **Recall** | **Precision** |
|---|---|---|---|
| (1) Sleeping | 99.4% | 97.8% | 100% |
| (2) Toilet | 100% | 24.0% | 100% |
| (3) Kitchen | 100% | 100% | 100% |
| (4) Relaxing room | 99.1% | 98.9% | 100% |
| (5) Dining room | 100% | 100% | 100% |
| (6) Bedroom | 100% | 98.3% | 99.1% |
| (7) Garden | 98.9% | 97.8% | 99% |
| (8) Leaving home | 98.4% | 97.2 | 98.2 |

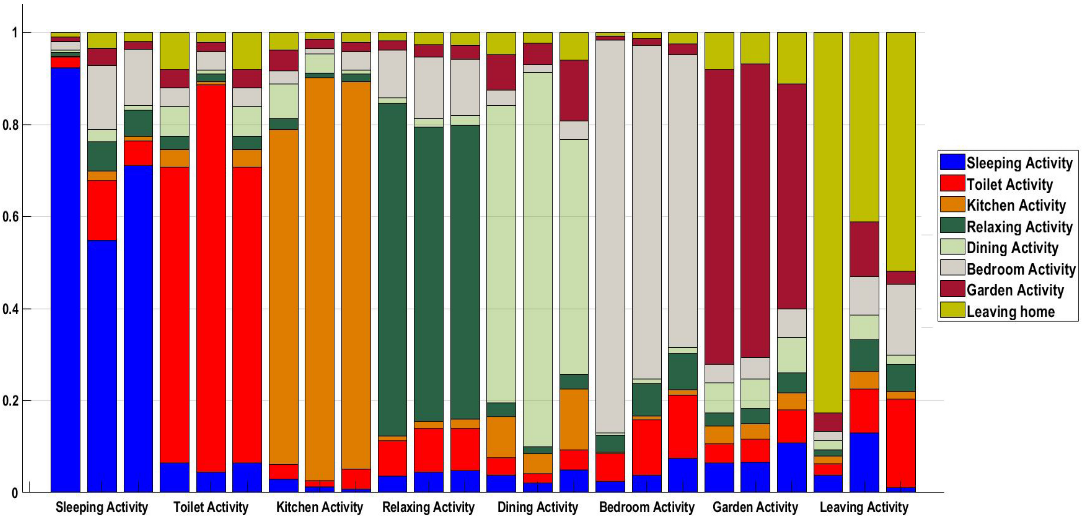



Figure 6.5: The results of the classification process using the FCM algorithm with the features extracted from the greyscale images representing three samples of each activity.

activity using AdaBoost and FCM, based on the feature vector extracted from the respective greyscale images. Based on the obtained classification results, it can be concluded that the idea behind improving the performance of decision trees in the AdaBoost technique by converting several weak classifiers to one single strong learner has achieved the best accuracy performance. The decision tree works to specify the input and the corresponding output during the training mode. Then the input data set is continuously split according to a

Table 6.2: The results of accuracy, recall and precision for each activity obtained based on Fuzzy C-Means.

| **Activities** | **Accuracy** | **Recall** | **Precision** |
|---|---|---|---|
| (1) Sleeping | 100% | 98.4% | 86.4% |
| (2) Toilet | 89.1% | 88.7% | 100% |
| (3) Kitchen | 93.7% | 92.5% | 93.7% |
| (4) Relaxing room | 85.7% | 85.1% | 88.8% |
| (5) Dining room | 86.3% | 91.2% | 87.7% |
| (6) Bedroom | 95.6% | 95.2% | 86.6% |
| (7) Garden | 71.2% | 70.9% | 72.1% |
| (8) Leaving home | 81% | 83.2% | 75.9% |

certain parameter to learn the relations between the inputs and their corresponding output by adjusting the weights during the training mode.

## 6.3 Bidirectional Long Short-term Memory with Fuzzy Feature Representation

The overall framework of the proposed method comprises three stages, as illustrated in Figure 6.6. Stage one is the data collection process. The data used in this experiment is Dataset A. At this stage, the gathered sensor information is the time-stamps of the binary data representing the ADL. Stage two is the fuzzy feature representation, which is designed to fuzzify the data. The resulting fuzzy membership degrees are taken as features to be used in modelling and recognition of activities, as explained in Chapter 4. The final stage in the proposed framework is the Bi-LSTM model for activity modelling and recognition, where the features obtained from stage two are used to train a Bi-LSTM to learn the relationship between the features and their corresponding outputs for modelling and recognising human activities. A detailed description of each stage is given in Section 6.3.1.

### 6.3.1 Experimental Setup

To evaluate the proposed method, it is applied to the data set A that gathered from a real environment representing the ADL of a single user. As mentioned in Chapter 4, the dataset used in this experiment represents seven different activities for a single user corresponding to; Sleeping, Toileting, Meal-preparation, Dining, relaxing, Garden and Leaving home.

The fuzzy feature representation stage is applied to represent the numerical features that extracted from the occupancy signal as explained in Chapter 4. In the fuzzy feature representation stage, the selected numerical information representing each activity will be used. The selected variables used in this experiment are;

- Activity start time ($u_1$) - which represents the start time for each activity.

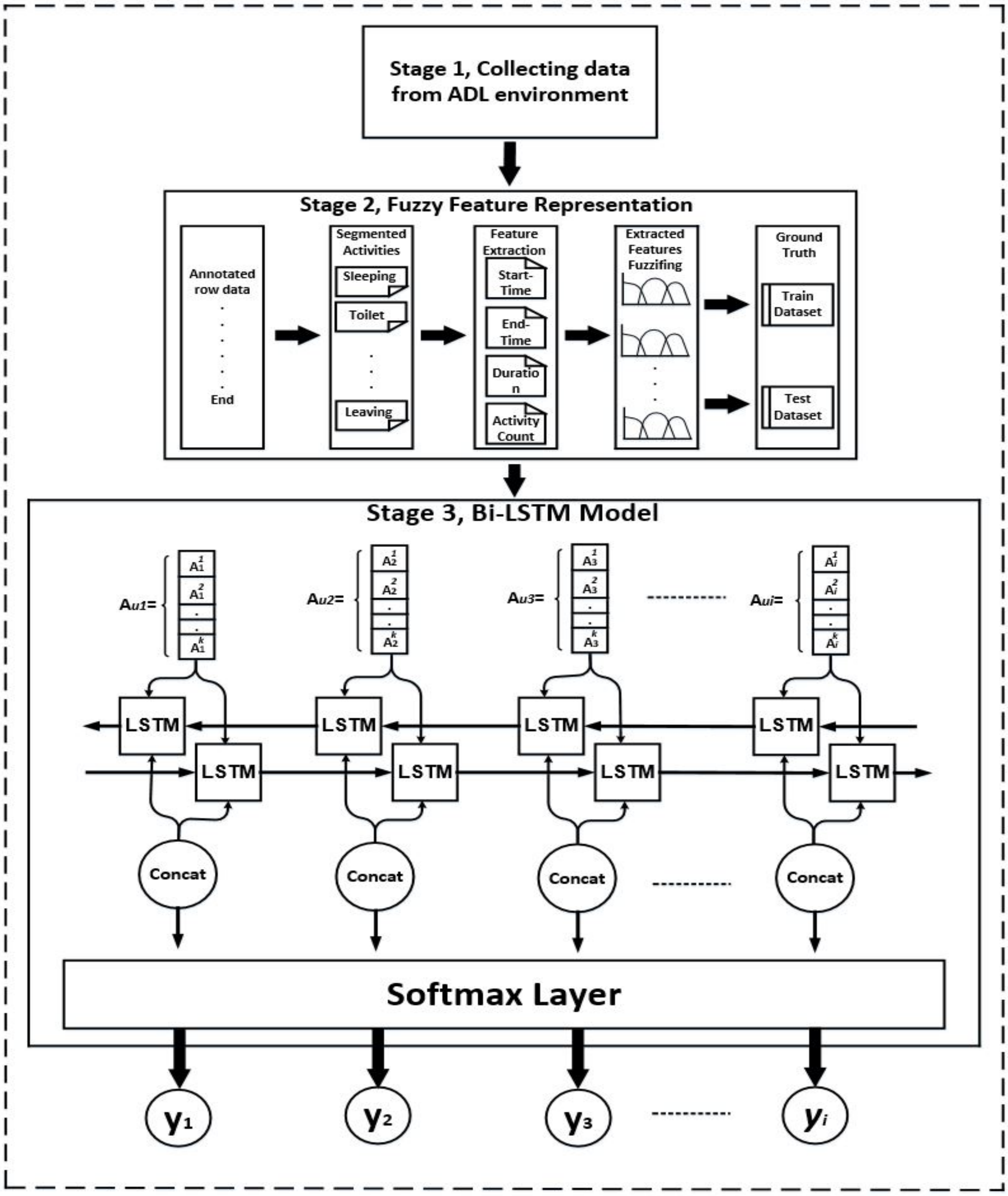



Figure 6.6: Proposed framework for fuzzy feature representation and the Bi-LSTM model for human activity modelling and recognition.

- Activity end time ($u_2$) - which is the end time for each activity.
- Activity duration ($u_3$) - which is the time spent for each activity.
- Activity count ($u_4$) - which is the number of activity attempts per hour.

Each value $x$ in the selected variables is fuzzified using Gaussian MFs to transform the data to relevant membership degrees. The membership labels associated with each input variable are represented using the following MFs:

$$U(t) = \begin{cases} u_1(x) \rightarrow \{EM_{u_1}, M_{u_1}, AF_{u_1}, EV_{u_1}, NI_{u_1}\} \\ u_2(x) \rightarrow \{EM_{u_2}, M_{u_2}, AF_{u_2}, EV_{u_2}, NI_{u_2}\} \\ u_3(x) \rightarrow \{VS_{u_3}, SH_{u_3}, ME_{u_3}, LO_{u_3}, VLO_{u_3}\} \\ u_4(x) \rightarrow \{VRU_{u_4}, RU_{u_4}, MU_{u_4}, HU_{u_4}, VHU_{u_4}\} \end{cases} \tag{6.1}$$

where each value $x$ in the selected input features $\{u_1, u_2, u_3, u_4\}$ is represented with 5 MFs as follows:

- The MFs representing activity start time for the input variable $u_1$ are represented as $\{EM_{u_1}, M_{u_1}, AF_{u_1}, EV_{u_1}, NI_{u_1}\}$, where $EM$, $M$, $AF$, $EV$ and $NI$ are MF labels corresponding to Early Morning, Morning, Afternoon, Evening and Night respectively.
- The MFs representing activity end time for the input variable $u_2$ are represented as $\{EM_{u_2}, M_{u_2}, AF_{u_2}, EV_{u_2}, NI_{u_2}\}$, where $EM$, $M$, $AF$, $EV$ and $NI$ are the MF labels corresponding to Early Morning, Morning, Afternoon, Evening and Night respectively.
- The MFs representing activity duration for the input variable $u_3$ are represented as $\{VS_{u_3}, SH_{u_3}, ME_{u_3}, LO_{u_3}, VLO_{u_3}\}$, where $VS$, $SH$, $ME$, $LO$ and $VLO$ are MF labels corresponding to Very Short, Short, Medium, Long and Very long respectively.
- The MFs representing activity count for the input variable $u_4$ are represented as $\{VRU_{u_4}, RU_{u_4}, MU_{u_4}, HU_{u_4}, VHUu_4\}$, where $VRU$, $RU$, $MU$, $HU$ and $VHU$ are the MF labels corresponding to Very Rare Usage,

Rare Usage, Medium Usage, Heavy Usage and Very Heavy Usage respectively.

In the next step, the Bi-LSTM model will be trained with the fuzzy represented features to learn the sequential relationships between the input data (activity data) and their corresponding output (activity labels).

### 6.3.2 Experimental Results

This experiment focuses on modelling and recognising human activities using fuzzy feature representation with Bi-LSTM model. Based on the explanation provided in the previous sections, the Bi-LSTM model is employed to model and recognise the ADL for a single user within a smart home environment. This section presents the results obtained with the proposed framework using the experimental dataset. Humans behave in their daily life in an imbalanced way as they would perform some activities more than the others. Therefore, if the dominant activities are well modelled and recognised, the overall accuracy for the whole system will still be high even though the other activities are not well modelled and recognised. Based on that, each activity will be evaluated separately, and then the accuracy performance of the whole system will be computed. Figure 6.7 shows the results obtained from the conducted experiment. This shows the confusion matrix illustrating the recall and

Table 6.3: Recall and precision for each activity obtained based on Bi-LSTM, standard LSTM and SVM.

| **Activities** | **Bi-LSTM** | | **LSTM** | | **SVM** | |
|---|---|---|---|---|---|---|
| | Recall | Precision | Recall | Precision | Recall | Precision |
| Sleeping | 100% | 94.4% | 91.7% | 87.50% | 91.7% | 68.8% |
| Toilet | 100% | 100% | 100% | 90.9% | 100% | 78.9% |
| Meal-preparation | 90.6% | 100% | 83.1% | 78.3% | 58.3% | 89.2% |
| Dining | 98.1% | 100% | 100% | 83.3% | 77.3% | 85.0% |
| Relaxing | 91.7% | 89.2% | 76.9% | 50% | 50% | 72.2% |
| Garden | 83.3% | 83.3% | 75% | NaN | NaN | 15.2% |
| Leaving | 100% | 90.9% | 35.9% | 42.6% | NaN | NaN |
| Overall accuracy | **96.3%** | | 80.8% | | 74% | |

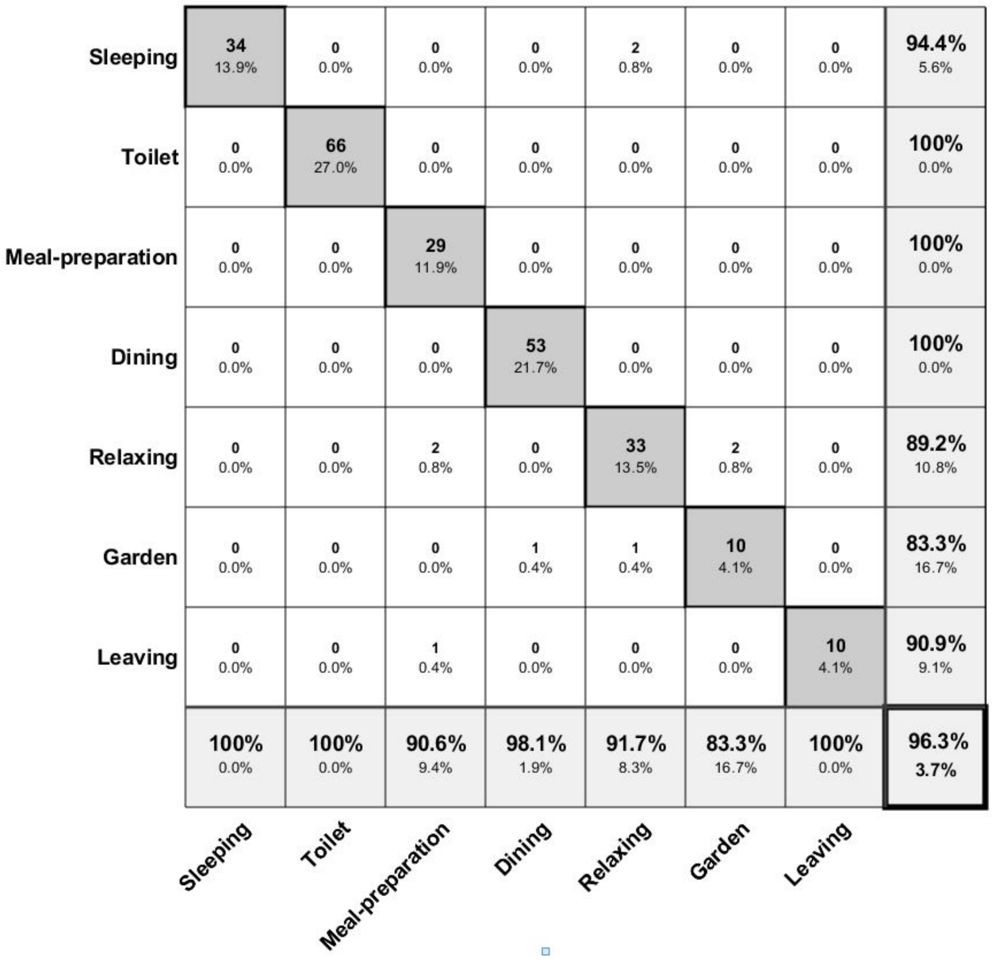



Figure 6.7: Confusion matrix plot for the obtained results by using Bi-LSTM for human activity modelling and recognition

precision scores for each activity as well as over the whole system. The information given in the confusion matrix is as follows:

- The rows represent the output activities, and columns represent the target activities. The activities are named as *Sleeping, Toilet, Meal-preparation, Dining, Relaxing, Garden and Leaving home.*
- The diagonal cells from the upper left to the lower right illustrate the activities that are correctly recognised.
- The off-diagonal cells show the incorrectly recognised activities.

- The last column in the right represents the precision for each activity.
- The last row in the bottom shows the recall for each activity.
- The bottom right cell represents the accuracy over the whole model.

To evaluate the proposed method, the results obtained in the proposed framework employing the Bi-LSTM model are compared with the results obtained using the standard LSTM and the Support Vector Machine (SVM) with the same dataset. The comparison between the performance of Bi-LSTM and the other two methods has been made for each activity in terms of recall, precision and the overall accuracy of the model as presented in Table 6.3.

## 6.4 Convolutional Neural Network with Fuzzy Feature Representation

The overall framework of the proposed CNNC for human activity modelling is shown in Figure 6.8. This framework consists of two stages; the data collection and presentation stage, and the Convolutional Neural Network Classification. In the data collection and presentation stage, binary sensor information gathered from a real-world environment in data set A is used. This sensory information is converted into an occupancy signal represented as fuzzy features. The fuzzy set thus generated will be used as input to train the proposed CNNC in the second stage of the proposed framework.

### 6.4.1 Experimental Setup

In this section, an experiment for modelling and classifying a single user's activities are conducted using the Dataset A with the proposed CNNC.

A commonly used CNNC architecture containing two conventional layers or more, dependent on the complexity of the input data, and one fully-connected layer is used in this experiment. Each one of these convolutional layers has multiple feature filters for accurately optimising values during the training phase. Additionally, each convolutional layer is followed by a max-pooling layer that has

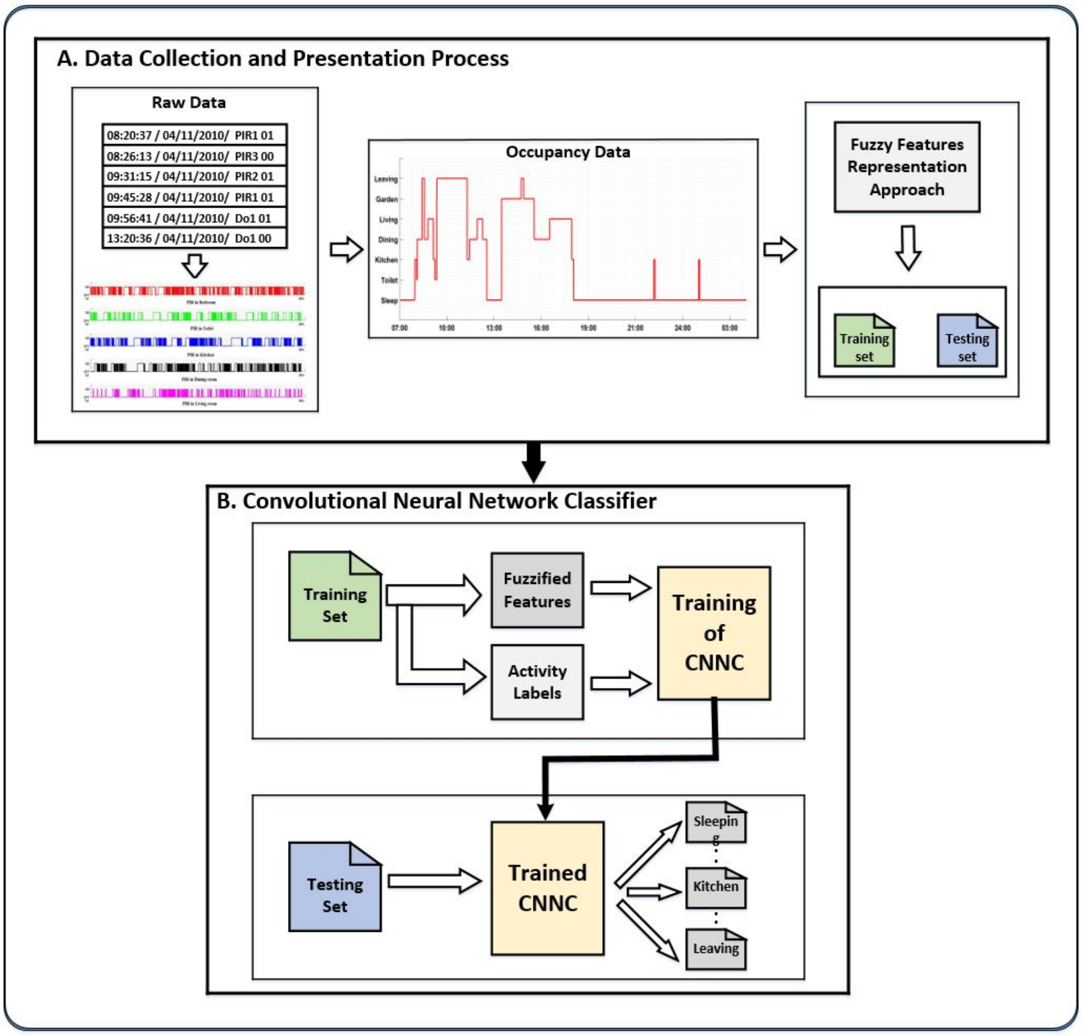



Figure 6.8: Architecture framework of the proposed Convolutional Neural Network Classifier with fuzzy feature representation approach for human activity modelling.

a window of a specific size to ensure that the outputs from each conventional layer are smaller than the inputs. The fully-connected layer in this architecture is a traditional Multi-Layer Perceptron (MLP) that contains a softmax activation function for the output layer. By operating the softmax activation function in the output layer, the CNNC can be used to classify the input features into various classes based on relations learned during the training stage [88]. In a scenario where complex input data is expected, the CNNC architecture may be modified to contain more than two of the convolutional and max-pooling layers with different sizes of border filters to process such data [4]. Also, more than one fully-connected layer could be considered after the top convolutional layer. In the training process of the CNNC, the standard forward and backward propagation algorithms are employed to select the values of the CNNC parameters.

As the Dataset A was collected using ambient sensory devices for a single user, it may contain a large volume of binary information. Before any data processing occurs, it is important to transform and visualise the collected sensor readings in an appropriate and useful format. In this experiment, a form of Finite State Machine (FSM) is used to represent and visualise the collected sensor readings as occupancy signal to identify the user's activities, as explained in Chapter 4. Once a set of activity data is extracted from the occupancy signal generated for each activity, they are fuzzified and represented as fuzzy sets, which are used as inputs for the proposed CNNC. This CNNC is applied to the fuzzy set in order to learn the relations between the input data (fuzzy sets) and their corresponding labels (activities).

The fuzzy feature representation approach applied to the activity data to represent the unique features in the user's daily activities is explained previously in Chapter 4. Three different unique features are extracted from each activity window ($W$) in the occupancy data. These three features are:

- *Start-time (St)*, which represents the time when the current activity starts.
- *End-time (Et)*, which identifies the ending time for the current activity.
- *Duration (Du)*, which represents the duration undertaking for a specific activity and is calculated as $Du = Et - St$.

Therefore, the extracted activity data vector for each activity window is expressed as:

$$U_x = \{St_x, Et_x, Du_x\} \tag{6.2}$$

where $U$ is the activity data vector extracted from the activity $x$, and $St, Et$ and $Du$ are the start time, end time and duration variables for the activity $x$ respectively. Each value in these three variables is represented with their relevant membership degrees for each fuzzy set. The generated set of fuzzified features for each input variable $uj$ is expressed as:

$$X_{uj} = [\mu A_{uj}^1, \mu A_{uj}^2, \ldots, \mu A_{uj}^M] \tag{6.3}$$

Since the data used is temporal data representing human activities for a single user, a sequential order is employed with the generated fuzzy sets to learn the relations in the data through the time steps using the proposed CNNC. The learned relationships are then used for modelling and recognising the user's activities.

### 6.4.2 Experimental Results

The proposed method of using the fuzzified feature with a CNNC was implemented to model ADL for a single user based on low-level sensory data gathered from a smart home environment. A binary dataset containing seven different activities are used to test and evaluate the proposed method. Three full days' sensor readings are used for testing and evaluating the proposed method and the remaining four days' data is used for the training process. Also, each activity is evaluated separately, and then the performance over the whole system is calculated. Datasets representing human activities are usually imbalanced because people behave in a random and unpredictable manner during their daily life. Thus, some activities are more dominant than other activities. Therefore, in this case, if the dominant activities are identified with a high degree of accuracy, the overall performance of the whole system will be high even if some activities are not well identified. Figure 6.9 illustrates the recall and precision for each activity, as well as the performance accuracy over the whole model.

The overall performance accuracy for the whole system based on the obtained results is 97.8%. The information given in the confusion matrix represented in Figure 6.9 is as follows:

- The rows and columns represent the output activities and target activities respectively. The activities are identified as $1, 2, ..., 7$ for Sleeping, Toilet,

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| 1 | **14** 14.9% | **0** 0.0% | **0** 0.0% | **0** 0.0% | **0** 0.0% | **0** 0.0% | **0** 0.0% | **100%** 0.0% |
| 2 | **0** 0.0% | **26** 27.7% | **0** 0.0% | **0** 0.0% | **0** 0.0% | **0** 0.0% | **0** 0.0% | **100%** 0.0% |
| 3 | **0** 0.0% | **0** 0.0% | **12** 12.8% | **0** 0.0% | **0** 0.0% | **0** 0.0% | **0** 0.0% | **100%** 0.0% |
| 4 | **0** 0.0% | **0** 0.0% | **0** 0.0% | **19** 20.2% | **0** 0.0% | **0** 0.0% | **0** 0.0% | **100%** 0.0% |
| 5 | **0** 0.0% | **0** 0.0% | **1** 1.1% | **0** 0.0% | **13** 13.8% | **1** 1.1% | **0** 0.0% | **86.7%** 13.3% |
| 6 | **0** 0.0% | **0** 0.0% | **0** 0.0% | **1** 1.1% | **0** 0.0% | **4** 4.3% | **0** 0.0% | **80.0%** 20.0% |
| 7 | **0** 0.0% | **0** 0.0% | **0** 0.0% | **0** 0.0% | **0** 0.0% | **0** 0.0% | **3** 3.2% | **100%** 0.0% |
| | **100%** 0.0% | **100%** 0.0% | **92.3%** 7.7% | **95.0%** 5.0% | **100%** 0.0% | **80.0%** 20.0% | **100%** 0.0% | **97.8%** **2.2%** |

Figure 6.9: Confusion matrix for ADL modelling and recognition results using fuzzy feature representation and the CNNC.

Meal-preparation, Dining, Relaxing, Garden and Leaving home respectively.

- The diagonal cells from the upper left to the lower right indicate the activities that are correctly recognised.
- The off-diagonal cells represent the incorrectly recognised activities.
- The right-most column shows the accuracy for each activity.

- The last row at the bottom shows the precision for each activity.
- The bottom right cell represents the accuracy over the whole model.

## 6.5 Discussion

In this work, we present novel approaches to model and recognise human activities based on a dataset collected using the ambient sensory device-based system. The method we propose facilitates faster and accurate learning abilities of human activities by generating a greyscale image from the binary string collected. The method uses a combination of pre-trained CNN (AlexNet) to extract features representing the human activity with FSM and AdaBoost for classifying the activities based on the extracted feature vector. AlexNet is used to extract and determine the number of high-level feature vectors for each activity. This approach is experimented to model and recognise different ADL for a single user with a dataset representing simple human activity such as sleeping, toilet and kitchen.

The work presented in this chapter has proposed a new method for visualising binary string data as a greyscale image. The dataset was gathered from a real-home environment using ambient sensory device-based system representing ADL activity for a single user. The generated greyscale images are used to extract a unique feature vector using DCNN technique. The extracted feature vector is then used as inputs to two different classifiers for recognising human activity. The results obtained from the conducted experiments (Table 6.1), show that the use of the extracted feature vector with the AdaBoost model exhibits a high score for accuracy, recall and precision when its performance is tested for each activity separately. Also, the overall performance of activity classification and recognition, when calculated over the whole system, demonstrates the effectiveness of this approach. The benefit of using the AdaBoost approach is that it improves the accuracy of the classification decision by integrating several weak classifiers in one single strong learner that has achieved the best accuracy performance.

The AdaBoost model is considerably better for ADL classification and recognition based on data gathered from low-level ambient sensors and is represented as a greyscale image. Besides, it can be seen that the AdaBoost model can recognise human activity in a proper sequence with the correct classification performance. The AdaBoost algorithm achieved more accurate results (99.5%) compared with the FCM (86.4%) for detecting the ADL. Both algorithms were tested and evaluated using the same feature vector extracted from the greyscale image representing 8 different ADL for a single user.

Also, a model for fuzzy feature representation of ADLs with a focus on modelling and recognising the activities is proposed. Fuzzy feature representation method makes it possible for the extraction of unique features from sparse features set of ADL and capture underlying uncertainties. To achieve the aim of this research, a Bi-LSTM model is used. The Bi-LSTM model is used because of its unique capability of modelling sequential data from the past and the future observations, which makes this beneficial in human ADL applications such as trend analysis. The obtained results show a high performance of the Bi-LSTM model on recognition of different activities compared with the results that are obtained from a standard LSTM and SVM methods for the same dataset. For future work, more experiments will be conducted with the proposed method in order to evaluate the robustness in modelling ADLs. This will be compared with state-of-the-art methods for feature representation, such as feature extraction by employing DCNN techniques.

Besides, the same fuzzy feature representation approach was applied to the extracted activity data, representing them as membership degrees and generated fuzzy sets. These sets were used along with the CNNC in modelling and recognising the user's activity. The proposed model was tested and evaluated using a "leave one day" approach, where once the proposed model is trained with the training set of data, one full day of sensor readings was used to test the proposed method and the remaining days were used for training the model. The fuzzy feature representation approach makes it possible to extract unique features from activity data representing ADL and to capture underlying uncertainties in such data. For future work, different datasets will be explored using the proposed method to evaluate its performance robustness in human activity modelling and recognition.

# Chapter 7

# Conclusion and Future Work

## 7.1 Thesis Summary

This thesis presents novel techniques to enhance the performance of the Fuzzy Finite State Machine (FFSM) for modelling and recognising human activities. The proposed approach will support independent living, particularly for older adults living within a home environment. Based on the results obtained from the conducted experiments, it can be concluded that the ability to model and recognise human activities using the enhanced FFSM depends on two different aspects. Firstly, the quality of the inputs applied to the model, as the data representing human activities is gathered using ambient sensory device-based, it should be represented in useful format before it will applied to the model. Therefore, the fuzzy feature representation approach of the extracted human activities play a crucial role in the effectiveness and robustness of modelling and recognising human activities. Secondly, the proposed learning approaches combined with the FFSM and their ability in learning the critical relations between the inputs (fuzzy sets representing human activity data) and the outputs (activities) during the training mode.

As part of the reported research, another approach was investigated to use a Deep Convolutional Neural Network (DCNN) for converting the acquire binary string of each activity into a greyscale image. Unique features are selected and extracted using pre-trained DCNN such as AlexNet. The extracted feature vectors are then used as inputs to different classifiers, including Adaptive Boosting (AdaBoost) and Fuzzy C-means (FCM) algorithms for recognising the Activities of Daily Living for a single user. The performance of the proposed model is evaluated using a real collected dataset.

This thesis contributes a novel framework for enhancing FFSM for human activity modelling and recognition. The existing version of FFSM is enhanced by integrating the learning abilities of the Neural Network (NN), Long Short-term Memory (LSTM), and Convolutional Neural Network (CNN) to generate N-FFSM, LSTM-FFSM, and CNN-FFSM, respectively. The learning capability in the NN, LSTM, and CNN allows the system to learn the relationship in the temporal human activity data and to identify the parameters of the rule-based system as building blocks of the FFSM through time steps in the learning mode. The learned parameters are then used for generating the fuzzy rules that govern the transitions between the system's states representing activities. The motivation for the work is from the perspective of the uncertain and unpredicted way that humans behave.

## 7.2 Concluding Remarks

This thesis attempts to provide an enhanced version of the existing FFSM for human activity modelling and recognition in ADL. Conclusions for various aspects of the project are presented below.

### 7.2.1 Activity Data Representation and Visualisation

To develop a robust model for human activity modelling and recognition, the need for flexible and efficient techniques to visualise and represent large scale binary string data representing human activities is highlighted and investigated. As the data sets were collected using ambient sensory device-based from real home environment representing ADL for a single user, the challenge was to find a proper approach to convert such binary sensor readings into occupancy data in a time-series format. A form of Finite State Machine (FSM) is used to represent and visualise the collected sensor readings as occupancy data to identify the user's activities. Five different numerical features were extracted for each activity. A fuzzy feature representation approach is employed to the extracted numerical features for representing them as fuzzy sets. The obtained fuzzy sets are used as inputs to the enhanced FFSM for human activity modelling and recognition.

### 7.2.2 Enhanced Fuzzy Finite State Machine

The work presented in the thesis has proposed a novel method for recognising and modelling human activities using data gathered from ambient sensory device-based, representing the ADL for a single user within a smart home environment. A practical application of utilising FFSM to model and recognise human activities using NN, LSTM and CNN with human expert knowledge. The principal elements of the FFSM were explained in details as well as the developed NN, LSTM and CNN learning techniques for generating the fuzzy rules and MFs associated with the linguistic labels in the inputs and outputs of the FFSM. The expert knowledge can still be used to define the system's states and the general structure of the state transitions. Experimental results are presented to demonstrate the effectiveness of the proposed methods. The advantage of the proposed system in integrating the experts' knowledge with the information derived from the automatic learning process. The results obtained from the proposed models show that human activities could be modelled/learnt with a high degree of accuracy based on the data gathered from ambient sensory device-based.

The results obtained from the conducted experiments in this chapter, three different aspects related to the proposed models: i.e. accuracy, interpretability and the importance of using expert knowledge are discussed as follows:

1. *Accuracy*: The results illustrated in Table 5.1 show that the LSTM-FFSM model achieves more accurate results for modelling and recognising human activities with a shorter period data (e.g., Dataset A), whereas, CNN-FFSM model is more effective for modelling and recognising human activities within a longer period data (e.g., Dataset B). Moreover, The results presented in Table 5.1 show the overall activity recognition performance when it is compared with some of the existing approaches

includes FFSM, LSTM, and NNs in terms of overall accuracy performance. According to the achieved results, generally, the three enhanced versions of FFSM model are considerably better for modelling and recognising ADL based on data gathered from ambient sensory device-based systems. Also, it can be seen how they are able to follow the proper sequence of states with the correct state activation degree.

2. *Interpretability Discussion*: From interpretability point of view, the most commonly used approaches in human activities recognition research works are Neural Network, Convolutional Neural Network and Hidden Markov Module on it is own. These models are considered as black-box approaches because of the complexity of understanding their underlying concepts. This complexity increases when a large number of input and output variables are used. Nevertheless, the proposed enhanced FFSM models are described linguistically using a different number of linguistic states (e.g., 8 and 10 states representing 8 or 10 various activities), as well as fuzzy rules associated with the linguistic inputs.

3. *The Importance of Using Human Expert Knowledge*: In order to achieve a robust model for representing human activities, the advantages of using expert knowledge with the learning capabilities in different deep learning techniques such as NNs, LSTM and CNN can be integrated with the FFSM model. Designing an FFSM only based on the linguistic information assigned by human experts is not enough for successful human activities modelling and recognition model. On the other hand, only information derived from the gathered sensor data is not usually enough to achieve a high-performance model. The expert knowledge is used to define the fuzzy rules, as well as to distinguish the system's current state(s). This allows obtaining a linguistic description of the ADL, i.e., the final set of fuzzy rules that control the transition between states.

Considering the results obtained from the conducted experiments, it can be concluded that the LSTM-FFSM and CNN-FFSM models exhibit a high score for accuracy, recall, and precision when its performance is tested for each activity separately. Also, the overall activity recognition performance, when it is over the whole system, demonstrates the effectiveness of the proposed approaches. The CNN-FFSM model shows a more robust and reliable performance once applied to a larger dataset (e.g., Dataset B) representing ten activities over 240 days. In particular, when this dataset contains some activities that could be happening at the same place (such as *Washing Dishes* and *Meal Preparation* in the kitchen). In real-life scenarios, it is hard to know which activity is the current activity based on the data collected from the PIR sensory devices. Thus using a fuzzy feature representation approach with CNN-FFSM will be used to deal with such cases as it can detect the changes in the fuzzy feature patterns. The essential feature of the proposed approach is that it integrates the available expert's knowledge with the learned information from the deep learning techniques. The LSTM-FFSM has shown better performance for a simple scenario once it is applied to a short period dataset (e.g., Dataset A). The CNN-FFSM achieved more accurate results to detect the ADL activities for a longer period dataset (e.g., Dataset B). Besides, it can be seen how the proposed models can follow the proper sequence of states with the correct state activation degree.

### 7.2.3 Employing Deep Learning Techniques with Fuzzy Feature Representation

In the framework proposed in this thesis, a fuzzy feature representation method is applied for the representation of human activities information obtained from an AmI environment. This approach is used as a means to improve the limited feature set of activities. The features obtained after the fuzzy representation process are used as input to a Bi-LSTM and CNN models which are trained for human activity modelling and recognition. The fuzzy feature representation approach is used mainly to fuzzify the data representing human activities. The membership degrees obtained for each activity in the data are used as features in the proposed model, and they are fed into Bi-LSTM and CNN model which learns the sequential relationship over a long-term period of activities data.

### 7.2.4 Employing Deep Leaning Technique for Feature Extraction from a Greyscale Images

The work presented in this thesis related a new method for visualising binary string data as a greyscale image. The dataset was gathered from a real-home environment using ambient sensory device-based representing ADL activity for a single user. The generated greyscale images are used to extract a unique feature vector using DCNN technique. The extracted feature vector is then used as inputs to two different classifiers for modelling and recognising human activity. Considering the results obtained result from the conducted experiments, it shows that the use of the extracted feature vector with the AdaBoost model exhibits a high score for accuracy, recall and precision when its performance is tested for each activity separately. Also, the overall performance of activity classification and recognition, when calculated over the whole system, demonstrates the effectiveness of this approach. The essential feature of the AdaBoost approach is that it improves the accuracy of the classification decision by integrating several weak classifiers in one single strong learner that has achieved the best accuracy performance.

The AdaBoost model is considerably better for ADL classification and recognition based on data gathered from low-level ambient sensors and is represented as a greyscale image. In addition, it can be seen that the AdaBoost model is able to recognise human activity in a proper sequence with the correct classification performance.

## 7.3 Directions for Future Works and Recommendations

Some suggestions for future work and improving our approach are presented below:

- A further investigation is required to implement the enhanced FFSM approaches presented in this thesis for modelling and recognising multiple occupancy activities in the same environment, i.e. more than one occupant at home. The presented FFSM approaches were not tested in the presence of visitors or even when the user has a pet companionship which is true for some cases. In the real-life, this task is considered as a complex scenario, as it is not enough to model and recognise only the user, but it must also monitor whether if the user alone or has visitors/pet at the same place in a specific time.

- The data sets used in this thesis were collected using ambient sensory device-based, which a combination of low-level sensors to measure user's activities (e.g., PIR sensors, door entry sensors and mat pressure sensors) and continuous sensors to monitor the environment conditions (e.g., ambient temperature sensors, humidity sensors and ambient light density sensors). The conducted experiments have only examined the binary string collected from the low-sensors measuring human activities. A future study is recommended to investigate the impact of using the non-binary string data gathered from the continuous sensors (e.g. temperature and light density) that may be found in a smart home environment. This can provide larger data samples for the experiments would be very interesting.

- The feature extraction and representation approach presented in this thesis focused on using a form of FSM to convert the binary string data gathered from ambient sensory devices into a time-series occupancy signal. 5 numerical features were extracted for each activity from the occupancy signal and represented as fuzzy sets. Future work should consider more features that have not been considered in this research. Another suggestion has been presented in the last chapter of this thesis to visualise the binary string data as a 2D greyscale image. Then a numerical features vector was extracted from the generated grey image using pre-trained CNN. This approach would be more interesting when it applied with a data set gathered using a higher number of sensory devices to provide more information in each image. This could also be used to provide information such as the temperature and light density for the environment.

- The proposed approaches presented in this thesis are based on offline data collection mode. Therefore, future work could be undertaken to extend the proposed approach for a real-time based system by enabling the suggested models to act as "AI-as-a-service".

# References


[1] Debraj Basu, Giovanni Moretti, Gourab Sen Gupta, and Stephen Marsland, "Wireless sensor network based smart home: Sensor selection, deployment and monitoring", in *IEEE Sensors Applications Symposium (SAS) Proceedings*. 2013, pp. 49–54, Galveston, TX. 1

[2] Liming Chen, Jesse Hoey, Chris D Nugent, Diane J Cook, and Zhiwen Yu, "Sensor-based activity recognition", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 790–808, 2012. 1, 18

[3] 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. 1, 16

[4] Munkhjargal Gochoo, Tan-Hsu Tan, Fu-Rong Jean, Shih-Chia Huang, and Sy-Yen Kuo, "Device-free non-invasive front-door event classification algorithm for forget event detection using binary sensors in the smart house", in *IEEE International Conference on Systems, Man, and Cybernetics*, 2017, pp. 405–409. 1, 14, 16, 26, 53, 73, 76, 108, 122

[5] Munkhjargal Gochoo, Tan-Hsu Tan, Shing-Hong Liu, Fu-Rong Jean, Fady S Alnajjar, and Shih-Chia Huang, "Unobtrusive activity recognition of elderly people living alone using anonymous binary sensors and dcnn", *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 693–702, 2019. 1, 14, 16, 27, 53, 76, 79, 89, 90, 108

[6] Artur Jordao, Antonio C Nazare Jr, Jessica Sena, and William Robson Schwartz, "Human activity recognition based on wearable sensor data: A standardization of the state-of-the-art", *arXiv preprint arXiv:1806.05226*, 2018. 1, 26

[7] Reshma Khemchandani and Sweta Sharma, "Robust parametric twin support vector machine and its application in human activity recognition", in *Proceedings of International Conference on Computer Vision and Image Processing*. Springer, 2017, pp. 193–203. 1, 21, 29

[8] Yu Zhao, Rennong Yang, Guillaume Chevalier, Ximeng Xu, and Zhenxing Zhang, "Deep residual bidir-lstm for human activity recognition using wearable sensors", *Mathematical Problems in Engineering*, vol. 2018, no. 7316954, pp. 13, 2018. 1, 24

[9] Odongo Steven Eyobu and Dong Han, "Feature representation and data augmentation for human activity classification based on wearable imu sensor data using a deep lstm neural network", *Sensors*, vol. 18, no. 9, Aug 2018. 1, 2

[10] 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, 16, 17

[11] Munkhjargal Gochoo, Shing-Hong Liu, Damdinsuren Bayanduuren, Tan-Hsu Tan, Vijayalakshmi Velusamy, and Tsung-Yu Liu, "Deep convolutional neural network classifier for travel patterns using binary sensors", in *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)*. IEEE, 2017, pp. 132–137. 1, 26, 29, 75, 108

[12] Panpan Ge, Jun He, Shuhua Zhang, Liwei Zhang, and Jiangfeng She, "An integrated framework combining multiple human activity features for land use classification", *ISPRS International Journal of Geo-Information*, vol. 8, no. 2, pp. 90, 2019. 1, 56

[13] Ahmed Nait Aicha, Gwenn Englebienne, and Ben Kröse, "Unsupervised visit detection in smart homes", *Pervasive and Mobile Computing*, vol. 34, pp. 157–167, 2017. 1, 16, 18, 28

[14] Jiadong Zhu, Rub San-segundo, and José M. Pardo, "Feature extraction for robust physical activity recognition", *Human-centric Computing and Information Sciences*, vol. 7, no. 1, pp. 1–16, 06 2017. 2

[15] Yue Deng, Zhiquan Ren, Youyong Kong, Feng Bao, and Qionghai Dai, "A hierarchical fused fuzzy deep neural network for data classification", *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 4, pp. 1006–1012, 2017. 2, 3

[16] Alberto Alvarez-Alvarez, Gracian Trivino, and Oscar Cordón, "Body posture recognition by means of a genetic fuzzy finite state machine", in *IEEE 5th International Workshop on Genetic and Evolutionary Fuzzy Systems (GEFS)*, 2011, pp. 60–65. 3, 4, 25, 43

[17] Jie Yin, Qiang Yang, and Jeffrey Junfeng Pan, "Sensor-based abnormal human-activity detection", *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1082–1090, 2008. 3, 26

[18] Caroline Langensiepen, Ahmad Lotfi, and Saifullizam Puteh, "Activities recognition and worker profiling in the intelligent office environment using a fuzzy finite state machine", in *IEEE International Conference on Fuzzy Systems (FUZZ)*, 2014, pp. 873–880. 3, 4, 20, 25, 30, 38, 42, 44

[19] Mingsheng Ying, "A formal model of computing with words", *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 5, pp. 640–652, 2002. 3

[20] Yongzhi Cao, Mingsheng Ying, and Guoqing Chen, "Retraction and generalized extension of computing with words", *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 6, pp. 1238–1250, 2007. 3

[21] Alberto Alvarez and Gracian Trivino, "Comprehensible model of a quasi-periodic signal", in *2009 Ninth International Conference on Intelligent Systems Design and Applications*. 2009, vol. 10, pp. 450–455, IEEE. 3

[22] Alberto Alvarez-Alvarez, Gracian Trivino, and Oscar Cordon, "Human gait modeling using a genetic fuzzy finite state machine", *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 2, pp. 205–223, 2012. 3, 16, 25, 38, 42, 43, 44, 45, 72

[23] Javier Medina-Quero, Shuai Zhang, Chris Nugent, and M Espinilla, "Ensemble classifier of long short-term memory with fuzzy temporal windows on binary sensors for activity recognition", *Expert Systems with Applications*, vol. 114, pp. 441–453, 2018. 4, 29, 86

[24] Faith A Unal and Emdad Khan, "A fuzzy finite state machine implementation based on a neural fuzzy system", in *proceedings of the Third IEEE World Congress on Computational Intelligence*, 1994, pp. 1749–1754. 4, 42, 44

[25] KP Sridhar, S Baskar, P Mohamed Shakeel, and VR Sarma Dhulipala, "Developing brain abnormality recognize system using multi-objective pattern producing neural network", *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 8, pp. 3287–3295, 2019. 4, 29, 40, 46

[26] Vito Magnanimo, Matteo Saveriano, Silvia Rossi, and Dongheui Lee, "A bayesian approach for task recognition and future human activity prediction", in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 2014, pp. 726–731. 4, 18, 19

[27] Ane Alberdi, Alyssa Weakley, Maureen Schmitter-Edgecombe, Diane J Cook, Asier Aztiria, Adrian Basarab, and Maitane Barrenechea, "Smart home-based prediction of multidomain symptoms related to alzheimer's disease", *IEEE journal of biomedical and health informatics*, vol. 22, no. 6, pp. 1720–1731, 2018. 4, 14, 16, 17

[28] Pau-Choo Chung and Chin-De Liu, "A daily behavior enabled hidden markov model for human behavior understanding", *Pattern Recognition*, vol. 41, no. 5, pp. 1572–1580, 2008. 4, 18

[29] Zhelong Wang, Ming Jiang, Yaohua Hu, and Hongyi Li, "An incremental learning method based on probabilistic neural networks and adjustable

fuzzy clustering for human activity recognition by using wearable sensors", *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 4, pp. 691–699, 2012. 4, 23, 25, 43, 45, 47

[30] Vincent Bombardier and Emmanuel Schmitt, "Fuzzy rule classifier: Capability for generalization in wood color recognition", *Engineering Applications of Artificial Intelligence*, vol. 23, no. 6, pp. 978–988, 2010. 4, 43

[31] Andrey Ignatov, "Real-time human activity recognition from accelerometer data using convolutional neural networks", *Applied Soft Computing*, vol. 62, pp. 915–922, 2018. 4, 19

[32] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu, "Deep learning for sensor-based activity recognition: A survey", *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019. 4, 14

[33] 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. 4, 13, 58

[34] 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, 2013. 4, 14, 68, 76

[35] Konlakorn Wongpatikaseree, Mitsuru Ikeda, Marut Buranarach, Thepchai Supnithi, Azman Osman Lim, and Yasuo Tan, "Activity recognition using context-aware infrastructure ontology in smart home domain", in *2012 Seventh International Conference on Knowledge, Information and Creativity Support Systems*. IEEE, 2012, pp. 50–57. 6, 73, 76

[36] Shaukat Ali, Shah Khusro, Irfan Ullah, Akif Khan, and Inayat Khan, "Smartontosensor: ontology for semantic interpretation of smartphone sensors data for context-aware applications", *Journal of Sensors*, vol. 2017, 2017. 6, 73, 76

[37] Kévin Viard, Maria Pia Fanti, Gregory Faraut, and Jean-Jacques Lesage, "Human activity discovery and recognition using probabilistic finite-state automata", *IEEE Transactions on Automation Science and Engineering*, 2020. 13, 21

[38] Diane J Cook, Juan C Augusto, and Vikramaditya R Jakkula, "Ambient intelligence: Technologies, applications, and opportunities", *Pervasive and Mobile Computing*, vol. 5, no. 4, pp. 277–298, 2009. 14, 59

[39] Giovanni Acampora, Diane J Cook, Parisa Rashidi, and Athanasios V Vasilakos, "A survey on ambient intelligence in healthcare", *Proceedings of the IEEE*, vol. 101, no. 12, pp. 2470–2494, 2013. 14

[40] Parisa Rashidi and Alex Mihailidis, "A survey on ambient-assisted living tools for older adults", *IEEE journal of biomedical and health informatics*, vol. 17, no. 3, pp. 579–590, 2013. 14, 22

[41] Qin Ni, Ana Belén García Hernando, and Iván Pau de la Cruz, "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. 14

[42] Oscar D Lara and Miguel A Labrador, "A survey on human activity recognition using wearable sensors.", *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013. 14

[43] Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and Moussa Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications", *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015. 14

[44] Hongnian Yu, Shuang Cang, and Yan Wang, "A review of sensor selection, sensor devices and sensor deployment for wearable sensor-based human activity recognition systems", in *2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA)*. IEEE, 2016, pp. 250–257. 14

[45] Yu Kong and Yun Fu, “Human action recognition and prediction: A survey”, *arXiv preprint arXiv:1806.11230*, 2018. 14, 18, 19

[46] Chun Zhu, Weihua Sheng, and Meiqin Liu, “Wearable sensor-based behavioral anomaly detection in smart assisted living systems”, *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 4, pp. 1225–1234, 2015. 16

[47] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. W. Baik, “Action recognition in video sequences using deep bi-directional lstm with cnn features”, *IEEE Access*, vol. 6, pp. 1155–1166, 2018. 16

[48] Yekta Said Can, Bert Arnrich, and Cem Ersoy, “Stress detection in daily life scenarios using smart phones and wearable sensors: A survey”, *Journal of biomedical informatics*, p. 103139, 2019. 16

[49] Prafulla Dawadi, Diane Cook, Carolyn Parsey, Maureen Schmitter-Edgecombe, and Miya Schneider, “An approach to cognitive assessment in smart home”, *Proceedings of the 2011 workshop on Data mining for medicine and healthcare*, pp. 56–59, 2011. 16, 18, 28

[50] Sumi Helal, Jae Woong Lee, Shantonu Hossain, Eunju Kim, Hani Hagras, and Diane Cook, “Persim-simulator for human activities in pervasive spaces”, in *Intelligent Environments (IE), 2011 7th International Conference on*. IEEE, 2011, pp. 192–199. 16

[51] Nam Thanh Nguyen, Dinh Q Phung, Svetha Venkatesh, and Hung Bui, “Learning and detecting activities from movement trajectories using the hierarchical hidden markov model”, in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, vol. 2, pp. 955–960. 18

[52] 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. 18, 27, 28, 75

[53] Lakmini P Malasinghe, Naeem Ramzan, and Keshav Dahal, "Remote patient monitoring: a comprehensive study", *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 1, pp. 57–76, 2019. 18

[54] 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. 18

[55] Masaya Inoue, Sozo Inoue, and Takeshi Nishida, "Deep recurrent neural network for mobile human activity recognition with high throughput", *Artificial Life and Robotics*, vol. 23, no. 2, pp. 173–185, 2018. 19

[56] Dhruv Mahajan, Nipun Kwatra, Sumit Jain, Prem Kalra, and Subhashis Banerjee, "A framework for activity recognition and detection of unusual activities", in *Indian Conference on Computer Vision, Graphics and Image Processing*, 2004. 20

[57] Hoang Trinh, Quanfu Fan, Pan Jiyan, Prasad Gabbur, Sachiko Miyazawa, and Sharath Pankanti, "Detecting human activities in retail surveillance using hierarchical finite state machine", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 1337–1340. 20

[58] Antonio Fernández-Caballero, José Carlos Castillo, and José María Rodríguez-Sánchez, "Human activity monitoring by local and global finite state machines", *Expert Systems with Applications*, vol. 39, no. 8, pp. 6982–6993, 2012. 20

[59] Nattapon Noorit and Nikom Suvonvorn, "Human activity recognition from basic actions using finite state machine", in *Proceedings of the first international conference on advanced data and information engineering (DaEng-2013)*. Springer, 2014, pp. 379–386. 20

[60] Youngwook Kim and Hao Ling, "Human activity classification based on micro-doppler signatures using a support vector machine", *IEEE*

*Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328–1337, 2009. 21, 59

[61] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L Reyes-Ortiz, “Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine”, in *International workshop on ambient assisted living*. Springer, 2012, pp. 216–223. 21

[62] Jie Yang, Jian Cheng, and Hanqing Lu, “Human activity recognition based on the blob features”, in *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, 2009, pp. 358–361. 22

[63] Diego R Faria, Mario Vieira, Cristiano Premebida, and Urbano Nunes, “Probabilistic human daily activity recognition towards robot-assisted living”, in *2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 2015, pp. 582–587. 22

[64] Asma Benmansour, Abdelhamid Bouchachia, and Mohammed Feham, “Modeling interaction in multi-resident activities”, *Neurocomputing*, vol. 230, pp. 133–142, 2017. 23, 29, 33

[65] K Subramanian and Sundaram Suresh, “Human action recognition using meta-cognitive neuro-fuzzy inference system”, *International journal of neural systems*, vol. 22, no. 06, pp. 1250028, 2012. 23, 24, 30

[66] Ce Li, Zhenjun Han, Qixiang Ye, and Jianbin Jiao, “Visual abnormal behavior detection based on trajectory sparse reconstruction analysis”, *Neurocomputing*, vol. 119, pp. 94–100, 2013. 23

[67] Zhengyou Zhang, Michael Lyons, Michael Schuster, and Shigeru Akamatsu, “Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron”, in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, 1998, pp. 454–459. 23

[68] Stefan Oniga and Jozsef Suto, “Human activity recognition using neural networks”, in *Control Conference (ICCC), 2014 15th International Carpathian*. IEEE, 2014, pp. 403–406. 23, 26, 45

[69] Michael C Mozer, "An intelligent environment must be adaptive", *IEEE Intelligent Systems and their applications*, vol. 14, no. 2, pp. 11–13, 1999. 23

[70] David Ada Adama, Ahmad Lotfi, and Caroline Langensiepen, "Key frame extraction and classification of human activities using motion energy", in *Advances in Computational Intelligence Systems*. 2019, pp. 303–311, Springer International Publishing. 24

[71] Alex Graves and Jurgen Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures", *Neural Networks*, vol. 18, no. 5, pp. 602 – 610, 2005. 24

[72] Intan Nurma Yulita, Mohamad Ivan Fanany, and Aniati Murni Arymurthy, "Fuzzy clustering and bidirectional long short-term memory for sleep stages classification", in *International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT)*, 2017, pp. 11–16. 24, 40, 49, 50, 86

[73] Intan Nurma Yulita, Mohamad Ivan Fanany, and Aniati Murni Arymuthy, "Bi-directional long short-term memory using quantized data of deep belief networks for sleep stage classification", *Procedia computer science*, vol. 116, pp. 530–538, 2017. 24, 77

[74] Vikramaditya Jakkula, Diane J Cook, et al., "Anomaly detection using temporal data mining in a smart home environment", *Methods of information in medicine*, vol. 47, no. 1, pp. 70–75, 2008. 24

[75] Michael S Ryoo and Jake K Aggarwal, "Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities", in *Computer vision, 2009 - 12th international conference on*. IEEE, 2009, pp. 1593–1600. 24

[76] Li Liu, Li Cheng, Ye Liu, Yongpo Jia, and David S Rosenblum, "Recognizing complex activities by a probabilistic interval-based model.", in *AAAI*, 2016, vol. 30, pp. 1266–1272. 25

[77] Hamid Medjahed, Dan Istrate, Jerome Boudy, and Bernadette Dorizzi, "Human activities of daily living recognition using fuzzy logic for elderly home monitoring", in *Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on*. IEEE, 2009, pp. 2001–2006. 25, 61

[78] M Javad Akhlaghinia, Ahmad Lotfi, Caroline Langensiepen, and Nasser Sherkat, "Fuzzy predictor model in an inhabited intelligent environment, in proceeding of ieee world congress on computational intelligence (wcci)", *Hong Kong*, pp. 939–946, 2008. 25

[79] Jianbo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiaoli Li, and Shonali Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition.", in *IJCAI*, 2015, pp. 3995–4001. 26, 30

[80] Seyed Ali Rokni, Marjan Nourollahi, and Hassan Ghasemzadeh, "Personalized human activity recognition using convolutional neural networks", *arXiv preprint arXiv:1801.08252*, 2018. 26

[81] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu, "3d convolutional neural networks for human action recognition", *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 221–231, 2013. 26

[82] Song-Mi Lee, Sang Min Yoon, and Heeryon Cho, "Human activity recognition from accelerometer data using convolutional neural network", in *Big Data and Smart Computing (BigComp), 2017 IEEE International Conference on*. IEEE, 2017, pp. 131–134. 26, 59

[83] Tang Lu-An, Han Jiawei, and Jiang Guofei, "Mining sensor data in cyberphysical systems", *Tsinghua Science and Technology*, vol. 19, no. 1, pp. 225–234, 2015. 26

[84] Madhuri Panwar, S Ram Dyuthi, K Chandra Prakash, Dwaipayan Biswas, Amit Acharyya, Koushik Maharatna, Arvind Gautam, and Ganesh R Naik, "Cnn based approach for activity recognition using a wrist-worn accelerometer", in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, 2017, pp. 2438–2441. 26

[85] Rui Hou, Chen Chen, and Mubarak Shah, "Tube convolutional neural network (t-cnn) for action detection in videos", in *IEEE international conference on computer vision*, 2017. 26

[86] Artur Jordao, Leonardo Antônio Borges Torres, and William Robson Schwartz, "Novel approaches to human activity recognition based on accelerometer data", *Signal, Image and Video Processing*, pp. 1–8, 2018. 26

[87] 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. 26

[88] 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. 27, 53, 54, 73, 75, 76, 89, 90, 121

[89] Urbano Miguel Nunes, Diego R Faria, and Paulo Peixoto, "A human activity recognition framework using max-min features and key poses with differential evolution random forests classifier", *Pattern Recognition Letters*, vol. 99, pp. 21–31, 2017. 27

[90] Nagender Kumar Suryadevara, Subhas C Mukhopadhyay, Ruili Wang, and RK Rayudu, "Forecasting the behavior of an elderly using wireless sensors data in a smart home", *Engineering Applications of Artificial Intelligence*, vol. 26, no. 10, pp. 2641–2652, 2013. 27

[91] Naomi Irvine, Chris Nugent, Shuai Zhang, Hui Wang, and Wing WY NG, "Neural network ensembles for sensor-based human activity recognition within smart environments", *Sensors*, vol. 20, no. 1, pp. 216, 2020. 29

[92] David M Burns and Cari M Whyne, "Personalized activity recognition with deep triplet embeddings", *arXiv preprint arXiv:2001.05517*, 2020. 29

[93] Mahsa Raeiszadeh, Hooman Tahayori, and Andrea Visconti, "Discovering varying patterns of normal and interleaved adls in smart homes", *Applied Intelligence*, pp. 1–14, 2019. 29, 105

[94] Mohammed Mehedi Hassan, Md Zia Uddin, Amr Mohamed, and Ahmad Almogren, "A robust human activity recognition system using smartphone sensors and deep learning", *Future Generation Computer Systems*, vol. 81, pp. 307–313, 2018. 29, 39

[95] Nils Y Hammerla, Shane Halloran, and Thomas Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables", *arXiv preprint arXiv:1604.08880*, 2016. 29

[96] Francisco Javier Ordóñez and Daniel Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition", *Sensors*, vol. 16, no. 1, pp. 115, 2016. 29

[97] Mohammad Abu Alsheikh, Ahmed Selim, Dusit Niyato, Linda Doyle, Shaowei Lin, and Hwee-Pink Tan, "Deep activity recognition models with triaxial accelerometers", in *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence*, 2016. 30

[98] Charissa Ann Ronao and Sung-Bae Cho, "Human activity recognition with smartphone sensors using deep learning neural networks", *Expert systems with applications*, vol. 59, pp. 235–244, 2016. 30

[99] Licheng Zhang, Xihong Wu, and Dingsheng Luo, "Recognizing human activities from raw accelerometer data using deep neural networks", in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 2015, pp. 865–870. 30

[100] Wenchao Jiang and Zhaozheng Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks", in *Proceedings of the 23rd ACM international conference on Multimedia*, 2015, pp. 1307–1310. 30

[101] Charissa Ann Ronao and Sung-Bae Cho, "Deep convolutional neural networks for human activity recognition with smartphone sensors", in *International Conference on Neural Information Processing*. Springer, 2015, pp. 46–53. 30

[102] Charissa Ann Ronaoo and Sung-Bae Cho, "Evaluation of deep convolutional neural network architectures for human activity recognition with smartphone sensors", in *ICONIP*, 2015, pp. 858–860. 30

[103] Thomas Plötz, Nils Y Hammerla, and Patrick L Olivier, "Feature learning for activity recognition in ubiquitous computing", in *Twenty-second international joint conference on artificial intelligence*, 2011. 30

[104] Asma Benmansour, Abdelhamid Bouchachia, and Mohammed Feham, "Multioccupant activity recognition in pervasive smart home environments", *ACM Computing Surveys (CSUR)*, vol. 48, no. 3, pp. 1–36, 2015. 33, 70

[105] Sara Amendola, Rossella Lodato, Sabina Manzari, Cecilia Occhiuzzi, and Gaetano Marrocco, "Rfid technology for iot-based personal healthcare in smart spaces", *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 144–152, 2014. 33

[106] C Ronald Kube, "Task modelling in collective robotics", *Autonomous Robots*, vol. 4, no. 1, pp. 53–72, 1997. 38

[107] Kevin Viard, Maria Pia Fanti, Gregory Faraut, and Jean-Jacques Lesage, "Recognition of human activity based on probabilistic finite-state automata", in *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE, 2017, pp. 1–7. 38

[108] Kwang-Ting Cheng and Avinash S Krishnakumar, "Automatic functional test generation using the extended finite state machine model", in *30th ACM/IEEE Design Automation Conference*, 1993, pp. 86–91. 38

[109] John E Seem, Gaylon M Decious, Carol Lomonaco, and Alex Bernaden, "Hybrid finite state machine environmental system controller", June 18 2002, US Patent 6,408,228. 38

[110] Anastasia Mavridou and Aron Laszka, "Designing secure ethereum smart contracts: A finite state machine based approach", in *International Conference on Financial Cryptography and Data Security*. Springer, 2018, pp. 523–540. 38, 39

[111] George Larionov, Zachary Kaden, Hima Varsha Dureddy, Gabriel Bayomi T Kalejaiye, Mihir Kale, Srividya Pranavi Potharaju, Ankit Parag Shah, and Alexander I Rudnicky, "Tartan: A retrieval-based socialbot powered by a dynamic finite-state machine architecture", *arXiv preprint arXiv:1812.01260*, 2018. 38, 39

[112] Francois Mouton, Alastair Nottingham, Louise Leenen, and HS Venter, "Finite state machine for the social engineering attack detection model: Seadm", *SAIEE Africa Research Journal*, vol. 109, no. 2, pp. 133–148, 2018. 38

[113] Alberto Fernández-Isabel, Paulo Peixoto, Isaac Martín de Diego, Cristina Conde, and Enrique Cabello, "Combining dynamic finite state machines and text-based similarities to represent human behavior", *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 504–516, 2019. 39

[114] Mehrbakhsh Nilashi and Othman Bin Ibrahim, "A model for detecting customer level intentions to purchase in b2c websites using topsis and fuzzy logic rule-based system", *Arabian Journal for Science and Engineering*, vol. 39, no. 3, pp. 1907–1922, 2014. 40

[115] Shabnam Yazdani and Morteza Montazeri-Gh, "A novel gas turbine fault detection and identification strategy based on hybrid dimensionality reduction and uncertain rule-based fuzzy logic", *Computers in Industry*, vol. 115, pp. 103131, 2020. 40

[116] L-X Wang and Jerry M Mendel, "Generating fuzzy rules by learning from examples", *IEEE Transactions on systems, man, and cybernetics*, vol. 22, no. 6, pp. 1414–1427, 1992. 43

[117] Shirin Enshaeifar, Ahmed Zoha, Andreas Markides, Severin Skillman, Sahr Thomas Acton, Tarek Elsaleh, Masoud Hassanpour, Alireza Ahrabian, Mark Kenny, Stuart Klein, et al., "Health management and pattern analysis of daily living activities of people with dementia using in-home sensors and machine learning techniques", *PloS one*, vol. 13, no. 5, pp. e0195605, 2018. 45

[118] Jose Antonio Iglesias, Plamen Angelov, Agapito Ledezma, and Araceli Sanchis, "Human activity recognition based on evolving fuzzy systems", *International journal of neural systems*, vol. 20, no. 05, pp. 355–364, 2010. 45

[119] Trupti Katte, "Recurrent neural network and its various architecture types", *International Journal of Research and Scientific Innovation (IJRSI)*, vol. 5, pp. 124–129, 2018. 48

[120] Orlando De Jesus and Martin T Hagan, "Backpropagation algorithms for a broad class of dynamic networks", *IEEE Transactions on Neural Networks*, vol. 18, no. 1, pp. 14–27, 2007. 48

[121] Yong Du, Wei Wang, and Liang Wang, "Hierarchical recurrent neural network for skeleton based action recognition", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1110–1118. 48

[122] Vivek Veeriah, Naifan Zhuang, and Guo-Jun Qi, "Differential recurrent neural networks for action recognition", in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4041–4049. 48

[123] Martin Jenckel, Sourabh Sarvotham Parkala, Syed Saqib Bukhari, and Andreas Dengel, "Impact of training lstm-rnn with fuzzy ground truth", in *ICPRAM*. 2018, pp. 388–393, IEEE. 49, 86

[124] Nabil Zerrouki, Fouzi Harrou, Ying Sun, and Amrane Houacine, "Vision-based human action classification using adaptive boosting algorithm", *IEEE Sensors Journal*, vol. 18, no. 12, pp. 5115–5121, 2018. 54

[125] Kishor H Walse, Rajiv V Dharaskar, and Vilas M Thakare, "A study on the effect of adaptive boosting on performance of classifiers for human activity recognition", in *Proceedings of the International Conference on Data Engineering and Communication Technology*. Springer, 2017, pp. 419–429. 54

[126] Jigar Surana, C Sweetlin Hemalatha, V Vaidehi, S Ananth Palavesam, and MJ Adith Khan, "Adaptive learning based human activity and fall

detection using fuzzy frequent pattern mining", in *2013 International Conference on Recent Trends in Information Technology (ICRTIT)*. IEEE, 2013, pp. 744–749. 56

[127] Kairui Guo, Rifai Chai, Henry Candra, Ying Guo, Rong Song, Hung Nguyen, and Steven Su, "A hybrid fuzzy cognitive map/support vector machine approach for eeg-based emotion classification using compressed sensing", *International Journal of Fuzzy Systems*, vol. 21, no. 1, pp. 263–273, 2019. 56

[128] Gheorghe Sebestyen, Ionut Stoica, and Anca Hangan, "Human activity recognition and monitoring for elderly people", in *2016 IEEE 12th international conference on intelligent computer communication and processing (ICCP)*. IEEE, 2016, pp. 341–347. 59

[129] Tim LM van Kasteren, Gwenn Englebienne, and Ben JA Kröse, "Human activity recognition from wireless sensor network data: Benchmark and software", in *Activity recognition in pervasive intelligent environments*, pp. 165–186. Springer, 2011. 59, 70, 74

[130] Dorothy Monekosso, Francisco Florez-Revuelta, and Paolo Remagnino, "Ambient assisted living [guest editors' introduction]", *IEEE Intelligent Systems*, vol. 30, no. 4, pp. 2–6, 2015. 59, 61

[131] Sawsan M Mahmoud, Ahmad Lotfi, and Caroline Langensiepen, "Occupancy pattern extraction and prediction in an inhabited intelligent environment using narx networks", in *2010 Sixth International Conference on Intelligent Environments*. IEEE, 2010, pp. 58–63. 59, 68

[132] Sawsan Mousa Mahmoud, *Identification and prediction of abnormal behaviour activities of daily living in intelligent environments*, PhD thesis, Nottingham Trent University, 2012. 59

[133] Diane J Cook, "Learning setting-generalized activity models for smart spaces", *IEEE intelligent systems*, vol. 2010, no. 99, pp. 1, 2010. 60, 66

[134] Shulai Yin, Hui Fang, and Xuehua Hou, "Athlete's respiratory frequency and physical energy consumption model based on speech recognition

technology", *International Journal of Speech Technology*, pp. 1–9, 2020. 60

[135] Dario Ortega Anderez, Ahmad Lotfi, and Caroline Langensiepen, "A novel crossings-based segmentation approach for gesture recognition", in *UK Workshop on Computational Intelligence*. Springer, 2018, pp. 383–391. 60

[136] Dario Ortega Anderez, Ahmad Lotfi, and Amir Pourabdollah, "Eating and drinking gesture spotting and recognition using a novel adaptive segmentation technique and a gesture discrepancy measure", *Expert Systems with Applications*, vol. 140, pp. 112888, 2020. 60

[137] Sawsan Mahmoud, Ahmad Lotfi, and Caroline Langensiepen, "Behavioural pattern identification and prediction in intelligent environments", *Applied Soft Computing*, vol. 13, no. 4, pp. 1813–1822, 2013. 61

[138] Samih Eisa and Adriano Moreira, "A behaviour monitoring system (bms) for ambient assisted living", *Sensors*, vol. 17, no. 9, pp. 1946, 2017. 61

[139] Sawsan M Mahmoud, Ahmad Lotfi, and Caroline Langensiepen, "User activities outlier detection system using principal component analysis and fuzzy rule-based system", in *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*, 2012, pp. 1–8. 68

[140] Juan Ye, Graeme Stevenson, and Simon Dobson, "Kcar: A knowledge-driven approach for concurrent activity recognition", *Pervasive and Mobile Computing*, vol. 19, pp. 47–70, 2015. 70, 73

[141] Selda Güney and Çağatay Berke Erdaş, "A deep lstm approach for activity recognition", in *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, 2019, pp. 294–297. 76

[142] Enrique Garcia-Ceja and Ramon F Brena, "An improved three-stage classifier for activity recognition", *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 01, pp. 1860003, 2018. 76

[143] Siyuan Lu, Zhihai Lu, and Yu-Dong Zhang, "Pathological brain detection based on alexnet and transfer learning", *Journal of computational science*, vol. 30, pp. 41–47, 2019. 79

[144] Arif Budiman, Mohamad Ivan Fanany, and Chan Basaruddin, "Adaptive convolutional elm for concept drift handling in online stream data", *arXiv preprint arXiv:1610.02348*, 2016. 79

[145] Jyh-Shing Roger Jang, Chuen-Tsai Sun, and Eiji Mizutani, "Neuro-fuzzy and soft computing; a computational approach to learning and machine intelligence", *International Journal of Intelligent Systems*, 1997. 82, 83

[146] Detlef Nauck, Frank Klawonn, and Rudolf Kruse, *Foundations of neuro-fuzzy systems*, John Wiley & Sons, Inc., 1997. 83

[147] Madan M Gupta, "Fuzzy logic and neural networks", in *[Proceedings 1992] IEEE International Conference on Systems Engineering*. IEEE, 1992, pp. 636–639. 83

[148] Victor Boskovitz and Hugo Guterman, "An adaptive neuro-fuzzy system for automatic image segmentation and edge detection", *IEEE Transactions on fuzzy systems*, vol. 10, no. 2, pp. 247–262, 2002. 83

[149] Parisa Rashidi, Diane J Cook, Lawrence B Holder, and Maureen Schmitter-Edgecombe, "Discovering activities to recognize and track in a smart environment", *IEEE transactions on knowledge and data engineering*, vol. 23, no. 4, pp. 527–539, 2010. 104

[150] Shreyank N Gowda, "Human activity recognition using combinatorial deep belief networks", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 1–6. 108

[151] Lu Xia, Chia-Chih Chen, and Jake K Aggarwal, "View invariant human action recognition using histograms of 3d joints", in *Computer vision and pattern recognition workshops (CVPRW), 2012 IEEE computer society conference on*, 2012, pp. 20–27. 108

[152] Zheng-Wu Yuan and Jun Zhang, "Feature extraction and image retrieval based on alexnet", in *Eighth International Conference on Digital Image Processing (ICDIP 2016)*. International Society for Optics and Photonics, 2016, vol. 10033, p. 100330E. 111