Sequential Learning and Shared Representation for Sensor-Based Human Activity Recognition

Rebeen Ali Hamad

A thesis submitted in partial fulfilment
of the requirements of the
University of Northumbria at Newcastle
for the degree of
Doctor of Philosophy

Research undertaken in the
Department of Computer and Information Sciences
Faculty of Engineering and Environment

October 2022

Abstract

Human activity recognition based on sensor data has rapidly attracted considerable research attention due to its wide range of applications including senior monitoring, rehabilitation, and healthcare. These applications require accurate systems of human activity recognition to track and understand human behaviour. Yet, developing such accurate systems pose critical challenges and struggle to learn from temporal sequential sensor data due to the variations and complexity of human activities. The main challenges of developing human activity recognition are accuracy and robustness due to the diversity and similarity of human activities, skewed distribution of human activities, and also lack of a rich quantity of well-curated human activity data. This thesis addresses these challenges by developing robust deep sequential learning models to boost the performance of human activity recognition and handle the imbalanced class problems as well as reduce the need for a large amount of annotated data.

This thesis develops a set of new networks specifically designed for the challenges in building better HAR systems compared to the existing methods. First, this thesis proposes robust and sequential deep learning models to accurately recognise human activities and boost the performance of the human activity recognition systems against the current methods from smart home and wearable sensors collected data. The proposed methods integrate convolutional neural networks and different attention mechanisms to efficiently process human activity data and capture significant information for recognising human activities.

Next, the thesis proposes methods to address the imbalanced class problems for human activity recognition systems. Joint learning of sequential deep learning algorithms, i.e., long short-term memory and convolutional neural networks is proposed to boost the performance of human activity recognition, particularly for infrequent human activities. In addition to that, we also propose a data-level solution to address imbalanced class problems by extending the synthetic minority over-sampling technique (SMOTE) which we named (iSMOTE) to accurately label the generated synthetic samples. These methods have enhanced the results of the minority human activities and outperformed the current state-of-the-art methods.

In this thesis, sequential deep learning networks are proposed to boost the performance of human activity recognition in addition to reducing the dependency for a rich quantity of well-curated human activity data by transfer learning techniques. A multi-domain learning network is proposed to process data from multi-domains, transfer knowledge across different but related domains of human activities and mitigate isolated learning paradigms using a shared representation. The advantage of the proposed method is firstly to reduce the need and effort for labelled data of the target domain. The proposed network uses the training data of the target domain with restricted size and the full training data of the source domain, yet provided better performance than using the full training data in a single domain setting. Secondly, the proposed method can be used for small datasets. Lastly, the proposed multi-domain learning network reduces the training time by rendering a generic model for related domains compared to fitting a model for each domain separately.

In addition, the thesis also proposes a self-supervised model to reduce the need for a considerable amount of annotated human activity data. The self-supervised method is pre-trained on the unlabeled data and fine-tuned on a small amount of labelled data for supervised learning. The proposed self-supervised pre-training network renders human activity representations that are semantically meaningful and provides a good initialization for supervised fine-tuning. The developed network enhances the performance of human activity recognition in addition to minimizing the need for a considerable amount of labelled data.

The proposed models are evaluated by multiple public and benchmark datasets of sensor-based human activities and compared with the existing state-of-the-art methods. The experimental results show that the proposed networks boost the performance of human activity recognition systems.

Contents

List of Figures

List of Tables

Acknowledgements

In the Name of Allah, the Most Beneficent, the Most Merciful,

First and foremost, I would like to praise Allah the Almighty, the most gracious, and the most merciful for his countless blessings, knowledge, and opportunity have given me throughout my research work to complete my PhD research successfully.

I would like to express my deepest gratitude and appreciation to my supervisors Dr. Bo Wei, Prof. Longzhi Yang and Prof. Wai Lok Woo for the continuous support of my Ph.D study and research, for their patience, inspiration, motivation, enthusiasm, and immense knowledge. Their guidance, extensive understanding of Machine Learning and remarkable enthusiasm for research helped me to overcome many challenging situations in all the time of research and writing of this thesis. I am extremely grateful to my father for his love, prayers, caring and sacrifices for educating and preparing me for my future. I am very much thankful to my wife for her love, understanding, prayers and continuing support to complete this research work.

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by the University Ethics Committee in Tuesday 27 October 2020.

I declare that the Word Count of this Thesis is 37240 words

Name: Rebeen Ali Hamad

Date: 30 September 2022

Chapter 1

Introduction

Human activity recognition (HAR) has emerged as an active key research field of pervasive computing aiming at providing information on human physical activities based on sensor observation data [1]. The major objective of HAR systems based on sensor data is not only to provide information about a user's behaviour but also to serve as an essential step to monitor and analyse human activities in several applications. Sensor-based HAR has a wide range of applications including assisted living and healthcare [2]. Furthermore, HAR can potentially be used to monitor and analyse physical human activities to manage and reduce the risk of multiple conditions including diabetes, cardiovascular and obesity [3]. Besides, HAR as an automatic system can be utilised to better understand human behaviour by continuously and remotely monitoring the motions of a resident to mitigate the caregiver's burden and reduce the economic and social pressures on families. Thus, HAR systems can support caregivers by remotely analysing the health status of users who stay independently and sending notifications to the caregiver whenever the users are in need of care [4].

1.1 Motivation

HAR has become an active and essential research topic in ubiquitous computing due to its usability in a variety of applications including healthcare, interactive gaming, transportation mode, sports, and monitoring systems for general purposes [5, 6]. HAR as a useful tool provides information to better understand people's behaviour based on sensing technology. This allows computing systems to automatically and remotely monitor and analyse individuals' movements to assist them in their day-to-day tasks [1]. As an example, one of the applications of HAR is elderly monitoring systems due to the increasing ageing population in the healthcare sector.

Population ageing is becoming one of the main world's immediate and severe problems and causing economic and social challenges [7]. According to a report provided by The World Health Organization (WHO), the statistics demonstrate that the aged population is growing faster compared to the previous decades [8, 9]. Hence, WHO predicts that the global population of people aged 65 and over, i.e. elderly people will grow from 461 million to 2 billion by 2050 [8]. Besides, a large number of elderly people are particularly vulnerable to living alone and in social isolation for a variety of reasons including the effect of stressful life changes like divorce or widowhood that change living arrangements. Further reasons that have increased the number of elderly people staying alone in their homes are the deaths of spouses, weakness, illnesses and disabilities. However elderly people often require more physical assistance and support in their activities of daily living (ADLs) due to cognitive and physical impairment. Also, independent living results in poor vision and health, as well as difficulties in conducting ADLs citekharicha2007health. ADLs are regular daily activities such as walking, showering, and sleeping which are performed for self-care [10]. In addition to that, a survey conducted by WHO in 2011 shows that over 650 million people of working age are disabled in the world. Adequate facilities are yet not available to meet the needs of people with disabilities. The aforementioned issues impose several problems on society and inflate healthcare costs. One of the promising solutions to address these challenges is monitoring the physical activities of elderly people or people with disabilities to support them from dangerous situations and detect deviations of behaviour to improve care alert systems [5]. HAR is emerging as a powerful tool to track and monitor ADLs of elderly people in their own homes. The main purpose of HAR as a challenging and highly dynamic research topic is to automatically recognise human physical daily activities in uncontrolled or controlled settings [11].

HAR systems have been conducted through different sensing technology approaches, i.e., video-based systems and sensor-based systems that include ambient sensors and wearable sensors. Data about human movements are collected using these sensing approaches for HAR systems. Video-based systems use cameras and rely on images to collect information about individuals' behaviours or physical activities. However, privacy matters and the high costs of installing cameras in residents' areas to monitor and record their physical daily activities are the most common concerns. Even though video-based approaches have obtained satisfactory performance in numerous applications including games, safety surveillance and security, the camera services are not welcomed in many scenarios, particularly when privacy matters are considered. Moreover, residents may perceive cameras as a threat to their privacy and thus mostly they are reluctant to share their information through cameras in their private

spaces. Besides, video processing methods are computationally intensive and require huge computational resources. Therefore the aforementioned problems hinder video-based systems from recognizing human activities.

The limitations of video-based systems have led to a shift towards using sensors-based approaches as a major tool to capture ADLs for HAR systems [12]. Sensor-based approaches include wearable sensors and smart home environments as an application of ambient assisted living. Wearable sensors including accelerometers and gyroscopes are enabled by recent technological developments to be embedded into smartphones and smartwatches. Wearable sensor-based solutions such as accelerometers are one of the most popular methods and have been used by a large number of studies to monitor human activities as people most commonly wear or carry these motion sensors that are embedded into smartphones or smartwatches [13]. Typically, wearable sensor-based approaches have been widely used in different applications to their miniaturized size and flexibility also wearable sensors are more suitable for indoor and outdoor scenarios, particularly HAR [13]. Furthermore, simultaneously multiple wearable sensors can be worn on various parts of the human body including the chest, waist, arm, neck, leg, and wrist to collect human activities for HAR systems [13]. Among these wearable devices, smartphones as an economic and efficient approach have been widely exploited to perform HAR systems. Undoubtedly, the above attributes of wearable sensors to record human movements further increase the strength and flexibility of wearable sensor-based approaches for recognising human activities [13].

Smart home environments use ambient sensors to capture human movements and the interactions between the resident and the smart home. Smart home sensors such as pressure and motion sensors are recording the presence or movements of the resident for HAR systems. Smart home environments as an AAL aim to minimize healthcare costs and enable elderly people who are in need of care to stay independently in their homes through technology. Smart homes comprise different sensors and network technology to support independent older adults through monitoring their ADLs and assisting them to stay safe and healthy in their own homes. HAR-based smart home settings equipped with ubiquitous sensors in the field of AAL have gained increased attention for monitoring the ADLs of elderly people and informing caregivers in urgent cases. Smart homes with unobtrusive sensing technology such as reed switch sensor, Radio Frequency Identification (RFID), and passive infrared (PIR) for HAR have been used as a suitable solution for enhancing independent living when privacy is concerned

The rapid development in sensor technology and wireless communication networks has increased attention to HAR-based sensor data resulting in recording a vast amount of data and

improving the performance of HAR systems [12, 14, 15]. While the data supply is advanced using various sensors technology, the demand for techniques to process and capture useful information about human activities to get insight into the recorded data is also increased [16–18]. To meet this demand, many data-driven studies have been conducted on the recorded data to address the challenges of recognizing human activities. Machine learning methods have been used to extract informative features from the collect sensor collected to recognize human activities[14, 19].

1.2 Challenges of Activity Recognition in Sensor Data

HAR aims to identify human activities through a sequence of sensor observations. The rapid development of HAR systems has addressed several research gaps and various methods have been proposed to minimize the level of the challenges and improve the performance of HAR systems. Despite the progress in previous works, several challenges remain and require further investigation to be addressed adequately [20–22]. This PhD project addresses the following three main challenges: i) accuracy and robustness in HAR, ii) imbalanced class problems, and iii) need for less supervision data. Each of these challenges is elaborated in the following sub-sections.

1.2.1 Accuracy and Robustness in HAR

Accurately recognising human activities is difficult due to the diversity and similarity of human activities. Typically, human activity is composed of a series of actions. For example, a breakfast activity consists of a series of actions such as switching on the toaster and coffee maker and then getting cheese out of the fridge. The diversity in human activities regarding duration and the differences in the order of the actions will make the problem of HAR even more complicated [23, 22]. These actions can be performed in a different order at different times or the action may disappear. Another factor that may affect the robustness of HAR systems is that the activity could be performed by a different user or multiple users. Hence, the robustness of the HAR system to recognise the performed human activity by different users becomes a challenge. Furthermore, often HAR systems have been particularly developed to recognise human activities based on either wearable sensor data or smart home sensor data which prevents the models to be reproducible and generalizable well on both types of sensor data for HAR. Hence most of the developed systems are not sufficiently

robust to accurately recognise human activities from both wearable and smart home sensors data [21].

Besides, highly similar activities due to overlapping their features such as snacks with breakfast or lunch with dinner add extra difficulties to developing an accurate model for HAR to distinguish between these activities. Moreover, recognising and discriminating against a large number of ADLs is still a challenging task compared to identifying a small number of human activities. This is related to the fact that when the number of ADLs increases, the trained model has to discriminate many activities which makes the discriminating process harder.

In summary, an explicit insight of what human activity is being executed for HAR systems by an individual regarding accuracy and robustness is still highly challenging.

1.2.2 Imbalanced Class Problems in HAR

ADLs are inherently imbalanced since some activities require longer time compare to other activities for example snack and sleeping activities. The large differences among the samples of the activities will make the machine learning techniques focus on the classes with the majority of samples and ignore the classes with minority samples [24]. To address this problem recording additional training data on human activities cannot solve this problem since most of the activities often occur infrequently and few activities such as sleep, watching TV, or showering and eating occur for a long time. Thus, this problem requires further exploration and being able to exploit the classes with infrequent samples will have a great impact on rendering a robust and accurate model for HAR systems [25].

1.2.3 Lack of Labelled Data in HAR

Most of the proposed methods for HAR are supervised learning and rely on a large amount of labelled data of human activities to train a model. Acquiring a considerable portion of annotated data to train a model is time-consuming, erroneous, and could even be impossible for some scenarios since labelled data requires a domain knowledge expert to manually annotate sensor recording observations of human activities. Addressing this challenge by exploiting and learning from non or sparsely-labelled data which could be available much easier, will significantly facilitate multiple issues in the task of HAR.

Besides, the majority of the proposed methods for HAR are supervised learning and following the assumptions of which the datasets of the training and testing must lie in the identical feature space, have the same label space and have the same underlying distribution.

As a consequence, the performance of the supervised methods often decreases when the training and testing data have different feature spaces and underlying distributions. Hence, the above assumption lead to the major challenges of machine learning methods. In HAR, the same set of binary sensors is used to generate training and testing data in order to have the same feature space. Moreover, the participants should have similar preferences or habits in both training and testing data to render the same underlying distribution. Finally, the set of activities in the training and testing have to be identical to follow the same label space. However, the assumptions of supervised learning are not often valid in real-world HAR applications. Therefore, developing a method to minimize the need for large annotated data by firstly sharing knowledge across different but related domains of human activities to further increase the performance of HAR systems is demanding and will be an important contribution to addressing this challenge and relaxing the assumption. Secondly, developing a model to be trained on a small amount of labelled data to reduce the need for large annotated data and improve the performance of HAR systems will also be a useful solution to overcome this challenge.

1.3 Aims and Objectives

The aim of this PhD thesis is firstly to develop sequential deep learning networks for recognising human activities from sensor data. Secondly, we propose deep learning networks for sharing representation and transferring knowledge across different but related domains of human activities to boost the performance of HAR systems, decrease the learning time and reduce the number of learned models and minimise the need for large labelled data. Considering the challenges introduced in Section 1.2, this thesis will achieve the following research objectives:

i. Developing robust deep sequential neural networks to further enhance the performance of HAR systems compared to the new state-of-the-art methods. This is specifically designed to address the challenge that is introduced in Section 1.2.1.

ii. One of the objectives of this thesis is to handle imbalanced class problems of HAR systems from sensor data. This addresses the challenge that is reported in Section 1.2.2.

iii. Finally, another important objective of this thesis is minimizing the need for large annotated data for HAR systems. This objective addresses the challenge introduced in Section 1.2.3.

1.4 Summary of Contributions

The main contributions of this PhD thesis are proposing novel sequential deep learning networks to accurately recognise human activities based on sensor observations. To do that, a comprehensive review of recognising human activities is conducted. The contributions comprise three main elements, each presented in a separate chapter.

i. This thesis develops robust and sequential deep learning networks to efficiently recognise human activities and enhance the performance of HAR systems compared to the current state-of-the-art methods based on the public human activity datasets from the wearable sensor and smart home sensors data. The proposed networks integrate a multi-head self-attention mechanism and performer attention mechanism with the ConvNet to efficiently process sensor data and focus on the important time steps for recognising human activities.

ii. Novel methods are developed to handle imbalanced class problems for HAR systems. Joint learning of sequential deep learning algorithms, i.e., LSTM and 1D ConvNet are developed to improve human activity recognition, particularly for less represented activities. Moreover, An improved version of the synthetic minority over-sampling technique (SMOTE) is developed which we named (iSMOTE) to handle imbalanced class problems for HAR systems. These methods have outperformed the current state-of-the-art methods.

iii. This thesis develops sequential deep learning networks to reduce the need for a rich quantity of well-curated human activity data by transfer learning techniques. Firstly, a multi-domain learning network is developed to transfer knowledge across different but related domains of human activities and alleviate isolated learning paradigms using a shared representation. The benefit of our developed method is firstly to reduce the need and effort for labelled data of the target domain. The proposed network uses the training data of the target domain with restricted size and the full training data of the source domain, yet provided better performance than using the full training data in a single domain setting. Secondly, the proposed multi-domain learning network reduces the training time by rendering a generic model for related domains compared to fitting a model for each domain separately.

   This thesis also developed a self-supervised network for representation learning from generated sensor data themselves without relying on pre-defined semantic annotations, i.e., activity classes. The self-supervised pre-training model is then fine-tuned with a

small amount of labelled data for supervised learning. The developed self-supervised pre-training network renders human activity representations that are semantically meaningful and provides a good initialization for supervised fine-tuning. The developed network improves supervised activity recognition and minimizes the need for a considerable amount of labelled data.

1.5 Thesis Structure

The thesis is organised as follows: Chapter 2 explains the background and related works of HAR. This chapter presents a comprehensive review of HAR that includes the most well-known methods, applications and public datasets. The review also describes the pipeline of HAR in detail and evaluation metrics. Particularly the first objective that is introduced in Chapter 1.3 is achieved. Chapter 3presents the deep learning methods to improve recognising of human activities from sensor data. The developed networks integrate 1D ConvNet with attention mechanisms to entirely dispense recurrent architectures to make efficient computations and maintain the ordering of the time steps. This Chapter achieves the second objective that is introduced in Chapter 1.3. Chapter 4 presents the developed methods to handle imbalanced class problems. This chapter achieves the third objective of this thesis that is introduced in Chapter 1.3. Chapter 5 presents our methods to reduce the need for a large amount of annotated data. This chapter achieves the last objective of this thesis as described in Chapter 1.3. Chapter 6 concludes the thesis and presents a summary, limitations and future works of this thesis.

1.6 Research Outputs

Publications arising from this work.

1. Rebeen Ali Hamad, Wai Lok Woo, Bo Wei, and Longzhi Yang. "Overview of Human Activity Recognition Using Sensors Data." In UK Workshop on Computational Intelligence, Springer, Cham, 2022. Just accepted. This paper represents part of the works of Chapter 2.

2. Rebeen Ali Hamad, Masashi Kimura, Longzhi Yang, Wai Lok Woo, and Bo Wei. "Dilated causal convolution with multi-head self attention for sensor human activity recognition." Neural Computing and Applications 33, no. 20 (2021): 13705-13722. This paper represents part of the works of Chapter 3.

3. Rebeen Ali Hamad, Longzhi Yang, Wai Lok Woo, and Bo Wei. "ConvNet-based performers attention and supervised contrastive learning for activity recognition." Applied Intelligence (2022): 1-17. This paper represents part of the works of Chapter 3.

4. Rebeen Ali Hamad, Longzhi Yang, Wai Lok Woo, and Bo Wei. "Joint learning of temporal models to handle imbalanced data for human activity recognition." Applied Sciences 10, no. 15 (2020): 5293. This paper represents part of the works of Chapter 4.

5. Rebeen Ali Hamad, Longzhi Yang, Wai Lok Woo, and Bo Wei. "Cross-Domain Activity Recognition Using Shared Representation in Sensor Data." IEEE Sensors Journal (2022). This paper represents part of the works of Chapter 5.

6. Rebeen Ali Hamad, Wai Lok Woo, Bo Wei, and Longzhi Yang. "Self-Supervised Learning for Human Activity Recognition in Sensor Data." IEEE Transactions on Circuits and Systems for Video Technology. Under review. This paper represents part of the works of Chapter 4 and Chapter 5.

Chapter 2

Background and Related Work

Sensor-based HAR refers to automatically recognizing human activities from collected data generated by different sensing devices including wearable and ambient sensors (smart home environment sensors). HAR based on sensors data has been utilized in different fields of study such as healthcare system [26, 27], behaviour analysis [28], ambient assisted living (AAL) [29, 30] patient monitoring systems [31, 32]. HAR is often classified into two main categories as shown in Figure 2.1: sensor-based recognition and vision-based recognition. Vision-based HAR methods utilize one or several cameras for recording video examples of human activities. Moreover, multiple views of visual human activities are used to detect human movements [33]. However, people are generally reluctant to use cameras for recording daily activity data due to privacy concerns [34, 35]. Another limitation is that processing visual data for HAR using cameras could be computationally expensive. Unlike vision-based HAR, sensor-based HAR has gained more outstanding acceptability in the users and research communities due to low cost and privacy protection [36, 37]. Besides, rapid evolutions in sensor technologies and ubiquitous computing have enabled sensors-based HAR with a satisfactory performance at a lower computational cost [12, 36].

HAR using machine learning methods from a series of data captured by sensors provides insights into what people are performing such as walking, showering, eating, or sleeping [38, 39]. Conventional machine learning methods have shown great progress and reached satisfying performance on HAR. Such methods include NB, SVM, HMM, k-NN, DT, and RF [40]. However, conventional methods rely heavily on handcrafted heuristic feature extraction, which is mostly domain-dependent, expensive, and usually needs domain experts [36, 23]. Moreover, handcrafted features are mostly specific to a domain and often less generalizable for application domains. Handcrafted features cannot develop an adequate amount of features from raw sensor data and are time-consuming [23]. Due to the aforementioned issues,

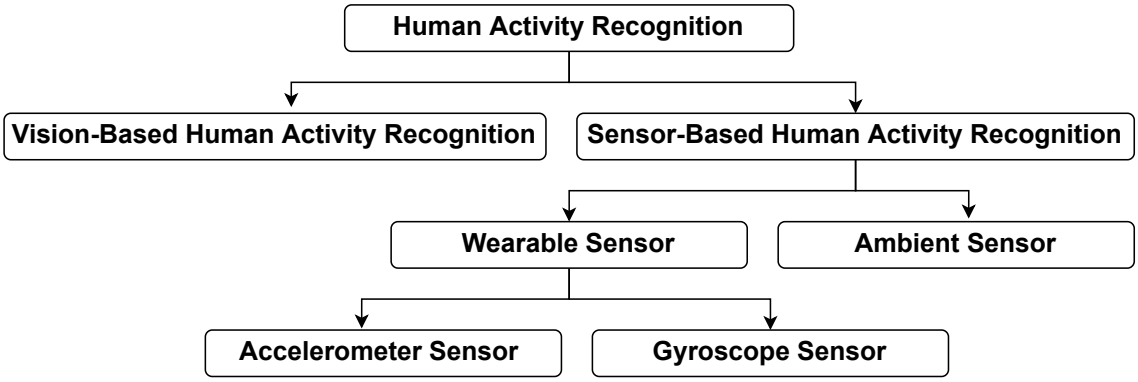


Fig. 2.1 Taxonomy of Human Activity Recognition

conventional machine learning algorithms have become less popular. Thereby, the power of automated feature extraction methods has received increasingly remarkable attention. Hence, more effective, capable and efficient deep learning algorithms have been developed for HAR systems [19]. The most popular methods of deep learning include RNN, LSTM and ConvNet [12]. The sensors deployed in HAR systems could be broadly categorized into two modalities: wearable sensor-based HAR and ambient sensors-based HAR [35, 36] as shown in Figure 2.1.

2.1 HAR Based on Wearable Sensor Data

Wearable sensors are unobtrusive, portable and inexpensive devices which can be conveniently worn by individuals. Wearable sensors are designed to meet some typical essentials including low battery consumption, small size, and high measurement accuracy. Furthermore, wearable sensors have recently received considerable attention and gained the potential to provide assisted living in healthcare applications and particularly to monitor human activities [41]. Wearable sensors are one of the most prevalent modalities in HAR systems. The wearable sensors generate a continuous stream of information based on the changes in the acceleration and angular velocity of human body movements. Thus the generated informative data are used to recognise human activities using machine learning methods [36]. The wearable sensors can be embedded in the human body and distributed from head to foot. The wearable sensors can work in a moderately wide area while the wearer is performing activities such as walking and running [12]. Wearable sensors such as accelerometers and gyroscopes are often embedded in smartphones, smartwatches, smart bands, clothes, belts, glasses, helmets, or shoes [12, 42]. Table 2.1 shows a review of HAR based on wearable sensors.

2.1.1 Accelerometer Sensors

Wearable sensors mainly consist of accelerometers and gyroscopes that have been broadly used to capture and extract diverse information related to the human motion for HAR [43, 44]. Accelerometers have been deployed for different applications such as fall detection [45], body motion analysis and movement [46], individual's postural orientation [47], and assessment of people with Parkinson's disease [48]. Human daily activities such as standing, cycling, walking, running, sitting, and walking downstairs and upstairs using accelerometer data can be effectively recognized [49, 50]. In 2004, Bao and Intille [51] employed numerous accelerometers to collect data from 20 subjects and to recognise 20 activities such as walking, brushing teeth, Reading, climbing stairs and vacuuming in a realistic setting. Since then, considerable systems of human activity recognition based on accelerometer sensor data have followed [52]. Accelerometers are the most commonly employed body-worn sensors in HAR systems due to their ability to render data based on the motion of the wearer [52]. In addition to HAR, accelerometers have been employed in different applications such as driving pattern analysis [53] breathing disease rehabilitation [54], and skill assessment [50, 55].

Smartphones as wearable sensors have been advancing rapidly and attracting great attention for HAR due to having various built-in sensing units including gyroscopes, ac-

Table 2.1 Review of human activity recognition based on wearable sensors

References	No.Sensors	Sensor placements	Activities
Zheng et al. [56]	1	Waist, Wrist, Hip Pocket	Standing, Lying, Walking, Dancing, Running, Upstairs, Downstairs, Jogging, Skipping, Sitting
Gjoreski and Gams [57]	7	Right ankle, Left thigh, Chest	Sitting, Going down, Standing up, Lying, Sitting on the ground, Standing
Jiang et al. [58]	4	Right forearm, Right shank, Left forearm, Left shank	Standing, Walking on an elliptical machine, Sitting, Running on an elliptical machine, Lying on a bed, Cycling, Jogging, Rowing, Walking and weight lifting
Jennifer et al. [59]	1	Smartphone	Sitting, Jogging, Upstairs, Walking, Downstairs, Standing
Chun and Weihua [60]	1	Right thigh	Stand-to-sit, Sitting, Lie-to-sit, Standing, Lying, Sit-to-stand, Walking, Sit-to-lie
Siirtola and Roning, 2012 [61]	1	Smartphone placed in pocket	Cycling, Walking, Driving a car, Standing, Sitting
Sweetlin [62]	1	Chest	Fall, Walking, Standing, Lying , Sitting
Mannini et al. [63]	1	Wrist, Ankle	26 daily activities
Lei et al.[64]	4	Chest, Left under-arm, Waist and thigh	Standing, Stand-to-sit, Walking, Sitting, Lie-to-stand, Lying, Sit-to-stand, Downstairs Stand-to-lie
Ronao, C.A. and Cho, S.B. [65]	1	Kept in the pocket	Walking, Laying, Sitting, Walking upstairs, Down upstairs, Standing
Davila JC et al. [66]	19	Hands, Left Foot, Back, upright knee, right foot, low right knee, Hip,	Standing, Sitting, Walking, Lie
Hassan MM et al. [67]	1	Smartphone: kept in the pocket	Standing, Stand-to-Lie, Walking, Stand-to-Sit, Lying down, Lie-to-Sit, Sitting, Lie-to-Stand, Walking-upstairs, Sit-to-Stand, Walking-downstairs, Sit-to-Lie
Wan S et al. [68]	1	Smartphone: kept in the pocket	Housecleaning, Sitting, Folding, Running, Laundry, Lying, Playing soccer, Walking, Computer work, Standing
Mekruksavanich S et al. [69]	1	Smartphone: kept in the pocket	Sitting, Walking downstairs, Laying, Walking upstairs, Standing
Han C et al. [70]	1	Smartphone: kept in the pocket	Going upstairs, Going downstairs, Jogging, Jumping and Walking

celerometers, cameras, and Global Positioning System (GPS) sensors [65]. Smartphones can be easily exploited to collect data instead of using different wearable sensors since Smartphones could be easily placed on different parts of an individual's body ranging from the upper such as the arm or wrist to the leg or ankle the lower [71, 72]. Moreover, smartphones can be utilized in both indoor and outdoor settings to record daily physical activities for HAR systems. The aforementioned features of smartphones for collecting human daily activities have increased the capability of the HAR system[73]. The accelerometer sensors data are used for HAR systems more than other sensors data from smartphone sensors[74]. Furthermore, the features of smartphones over other wearable devices have made smartphones a key ubiquitous platform for HAR systems [75]. The features are first: a smartphone is a low-cost machine that provides different software and hardware sensors in a single device. Second, smartphones can capture and process data since they are programmable devices. Moreover, smartphones are able to transmit and receive data as well as connect with other devices that have made smartphones an ideal platform for HAR systems in the research community [76].

2.1.2 Gyroscope Sensors

Gyroscope sensors can be used to measure angular velocity and preserve the orientation of an object. The difference in the angle could be detected over a period of time by comparing the initial know value with the change of the angle. The limitations of gyroscope sensors are output drift over time and the sensitivity of the gyroscope to a certain range of angular velocities. Often accelerometer sensors have been used in human activity recognition or the combination of gyroscope and accelerometer sensors. Narayanan et al. [77] performed estimating the fall risk of the movements of 68 older adults to measure the timed sit-to-stand, up-and-go test. Besides five more repetitions and alternate step tests are conducted using a triaxial accelerometer attached to the waist. Greene et al. [78] used both gyroscopes and accelerometers attached to two legs of a user to measure the timed up-and-go test to differentiate non-fallers from fallers. Varkey et al.[79] used both gyroscopes and accelerometers and attached these sensors to the right foot and right arm wrist of the user to obtain linear and angular accelerations. Regarding activity monitoring, inertial sensors can be useful due to these features low power requirements, low cost, compact size, non-intrusiveness, and the ability to provide data directly associated with the movement of the user. However, inertial sensors have these limitations. Firstly the placement of the inertial sensor on various parts of the human body causes an uncomfortable feeling for older adults which may lead to low acceptance by people. Secondly, due to collecting data continuously by the inertial sensors,

battery life could be reduced. inertial sensors cannot provide adequate information for monitoring complex movements and activities that involve many human-object interactions [79].

2.2 HAR Based on Smart Homes Sensor

The steady increase in the elderly people's population has been identified as a prominent social problem and financial challenge for the future decade including increased elderly healthcare demands, financial stress and unbalanced supply-demand. Consequently, the demand for caregivers to provide care for elderly people, expenditure on healthcare costs, and the desire of older adults to live unassisted in their residences have increased [19]. Smart home environments provide one of the most promising solutions to deliver ubiquitous and context-aware services and monitor activities of daily life. Smart homes are equipped with diverse types of sensors to help older individuals who wish to remain in their residences. Various sensors have been deployed in smart homes to capture various human activities. For instance, passive infrared (PIR) sensors to monitor the motions of the resident, reed switches on cupboards and doors to detect open or close status, and pressure sensors on beds, chairs and sofas to detect the presence of residents. Moreover, float sensors in the washroom detect whether the flush is performed or not [19]. Remarkably, with the rapid development of new low-cost sensors, cloud computing and the emerging Internet of Things (IoT) technologies, AAL systems deployed in smart homes have intelligently been used to monitor, record and act on physical surroundings and potentially support different applications such as health and wellness evaluation, short and long-term activity pattern analyses, consistent rehabilitation instruction, fall detection at home, chronic disease management, and timely medication reminder [80–86]. For senior people who hope to remain independently and functionally in their own home, they should have enough ability to perform Activities of Daily Living (ADL) such as preparing breakfast and lunch, drinking, cooking, eating, and washing grooming to be used by the caregivers to know a resident's functional status [82]. Smart home environments have been employed to unobtrusively observe the interactions of a user with the physical surrounding objects in the environments and the user's activities [30]. The interactions are the ADL and conducted by an individual in a particular place with the specific object in the environment which have been processed using machine learning methods to detect activities [87]. For example, cooking activity takes place in the kitchen, preparing dinner or breakfast usually takes place in the kitchen, shower related activities such as brushing teeth takes place in the bathroom. Thus, environmental sensors are deployed to capture an individual's

existence and interaction with the physical surrounding objects and furniture (e.g., coffee table, dining table, chairs, cupboard, bed, sofa) [88]. For instance, if the sensor readings indicate a stove, toaster, or microwave is working, and the door of the refrigerator or the cupboard is opened, then an activity is taking place in the kitchen. Hence based on the objects or furniture usage the activities will be distinguished and detected. Thus, correctly recognizing in-home human daily activities plays a significant role in comprehending and analysing the association between residents and their physical surroundings to help residents stay safe and healthy in their own independently residence and reduce the costs of healthcare [89]. Multiple smart homes have been developed for ambient assisted living particularly human activity recognition which is documented in Section 2.2.2.

2.2.1 Smart Environment Embedded Sensors

The most typical sensors that have been deployed in smart homes to record human activities will be described. Binary sensors are often used in smart environments that generate 0 or 1 to capture changes such as "door opened" or "door closed" [90]. Binary sensors such as pressure sensors, motion sensors, contact switch sensors, and state-change sensors have been deployed on different objects to monitor and capture interactions of a resident with the physical surroundings in a particular place of the environment [91].

- Motion sensors such as Passive InfraRed (PIR) [2, 92–95] have been widely used to track and record the presence of a resident in a particular place of the smart environment, for example, bedroom, kitchen, or bathroom. Besides, PIR sensors were employed for different applications such as stress monitoring and support to clinical decisions [96], and for moving targets detection and localization [90], and security [97].

- A simple state-change sensor deployed to detect human object interactions by recording the changes in the state of the objects [98–102]. For instance, the state-change sensor can be bound to the handset of a home telephone to capture the interactions between a resident and the home telephone when the resident lifts the handset from the telephone base station in the smart environment.

- Pressure sensors are utilised to detect sitting or lying on chairs or beds to unobtrusively monitor and record the presence and absence of a resident [75, 103–105].

- Contact switch sensors can be connected to different objects including the doors of beds, living or kitchen rooms. Further, this type of sensor has been linked with the

fridge door, or cabinets to detect interactions between residents with the physical surrounding objects [106–109].

In real-world scenarios, activity monitoring systems often employ multiple different sensors such as pressure sensors, motion detectors, break-beam sensors, and contact switches to record sufficient information associated with residents' activities around the smart home. The sensors detect a nearby motion, pressure or physical contact against objects [110]. Ordonez et al.,2013 [111] performed a study for HAR in a smart home environment and observes multiple activities such as Leaving Home, Toileting, Showering, Sleeping, Breakfast, Dinner, and Drinking. The study deployed various sensors to track these activities. PIR sensors were utilised to record motions around the smart environment, for instance, the living room area, kitchen area, or corridor. Reed switches sensors to record opening or closing the doors or cupboards. Float sensors are deployed to detect toilet flushing. Binary sensors are mostly deployed in smart homes due to their distinguished features such as low-cost, long-live and easy installation and replacement [7]. In addition to privacy-preserving of residents and unobtrusively recording human-object interaction, deployed binary sensors in smart homes require minimal computation resources to collect data.

Both active and passive Radio-Frequency Identification (RFID) have been used in smart homes to recognize physical human activities. Active RFID tag uses a battery and is essentially used by a resident to perform a personal identification in the smart home. On the other hand, passive RFID is not using a battery since the power is supplied by the reader and normally it will be attached to physical objects to detect human-object interactions in the smart home [112–115].

2.2.2 Smart Home Environment Projects

In this section, we review the most recognized smart home environments that are particularly used for human activities in addition to ambient assisted living purposes. These projects have been used to support senior individuals to stay independent and to monitor the healthy status of the residents remotely to reduce the burden on caregivers.

i. The GatorTech smart home that contains various smart devices equipped with sensors such as a smart mailbox, smart front door, and smart refrigerator for research on ambient assisted living was built by the University of Florida [116]. These smart homes have been used to provide services such as activity recognition and tracking as well as voice recognition. The smart home can sense and record the interaction of the residents with their physical surroundings. Besides, several similar projects developed

such as Adaptive Versatile home (MavHome) [117] from the University of Texas at Arlington, PlaceLab [118] from MIT, and Intelligent System Lab (ISL) [119] from the University of Amsterdam.

ii. Massachusetts Institute of Technology developed a smart home that was equipped with different sensors such as reed switches and piezoelectric switches on drawers sinks, doors, stoves, windows, cabinets, microwave ovens, lamps light switches refrigerators, toilets, and showers. These sensor-based objects in the smart home are used to detect and collect more than 20 activities [120]. The recorded data was labelled by using a naive Bayes classifier.

iii. In [121], van Kasteren et al., used deployed a wireless sensor network (WSN) based system in a real home using 14 sensors to record human activities for 28 days. A Bluetooth headset with voice recognition software is used by the subject to automatically label the collected data. The classification of seven activities was targeted by the system. A hidden Markov model and conditional random field were used to process the data and reported an accuracy of 79.4%. Kasteren smart home datasets were expanded to include three different real houses for human activity recognition research [122]

iv. The Centre for Advanced Studies in Adaptive Systems (CASAS) is a multi-disciplinary research project that was developed at Washington State University in 2007. This project mainly aimed at creating many smart home environments by deploying unobtrusive sensors and actuators [123, 124]. The smart homes in this project can record residents' daily activities which can be used by machine learning techniques to track and recognize human activities. The smart homes can monitor 15 different activities using analogue, temperature and motion sensors. CASAS consist of 66 independent event datasets recorded from seven physical smart homes. The datasets are used to perform a comparative study between several classifiers including HMM, CRF and naive Bayes (NB) classifiers. The reported results using three folds cross-validation over the dataset are 75.05%, 74.87%, and 72.16% for the HMM, NBC, and CRF, respectively [124].

v. Qing and Mohan proposed a smarter and safer home at CSIRO to improve senior individuals' quality of life [125]. To perform this, multiple environmental sensors are equipped in different places to monitor human activities. Based on this project, a Smart Assistive Living (SAL) platform was proposed to support senior adults in staying in their own homes independently. The sensors equipped in the smart home are expected

to render a continuous data stream to indicate the individual's activities. The recorded data from this smart home can be processed using machine learning techniques to help health caregivers to provide assistance and make a decision.

These smart home projects have generated numerous datasets some of which are publicly accessible and can be utilised for further investigation.

2.3 Applications of Human Activity Recognition

In this section, applications that can rely on HAR are highlighted. Particularly, we review healthcare, security and surveillance, entertainment and games, and home automation applications. For each of these applications, we present the benefits of automatically recognizing human activities that can deliver useful services.

- Healthcare Applications: HAR plays a crucial role in assisting the physical and mental well-being of the population. HAR can be used to conduct robust recognition of unsafe situations and detect deviations of behaviour to improve older adults' care are alert systems [126]. Furthermore, HAR could be potentially used to reduce or prevent the risk of various chronic diseases such as diabetes, obesity, neurological conditions, and cardiovascular [127]. People with these diseases in addition to their treatment are usually following an effective physical activity scheme or routines such as cycling, walking, running jogging. Accurate HAR can help caregivers to identify whether the patient or observee has any difficulties following the routines to perform activities. Besides, adequate information regarding the duration of activities is useful for individuals to compliance their ADL according to prescription and for practitioners to monitor and assess their health status. HAR in a healthcare application is a proactive process to adopt a healthy lifestyle for people who follow a daily routine, for example, everyday activities including exercising, sleeping, and social relationships [128].

- Security and Safety: HAR has been broadly employed for surveillance systems based on vision and sensor data. Multiple solutions for HAR-based visual data have been proposed to detect various suspicious activities in a public place [129, 130]. Besides sensor data streams are also used for security purposes. Sun et al. [131] proposed a HAR surveillance system to detect various physical assaults and criminal offences such as gunshots, abuse, and kidnapping. HAR systems have been employed to control appliances within the home and offices [132]. HAR used to turn on devices in a room

when a resident entered the room automatically and also the devices are turned off when the resident left the room to save energy and increase safety [133, 134].

2.4 Human Activity Recognition Pipeline

HAR aims to recognise human physical activities from a series of observations that are captured by sensors [1]. The availability of various sensors such as smart home sensors and wearable sensors enables the recording of human activities to monitor and analyse human behaviours for different applications such as healthcare and assisted living [2]. To develop HAR systems, multiple phases need to be carried out as shown in Figure 2.2. These phases for HAR systems include data collection from sensors, preprocessing of raw sensor data, segmentation of sensor data, data splitting, feature extraction and classification of human activities [1]. The details of each step are provided below.

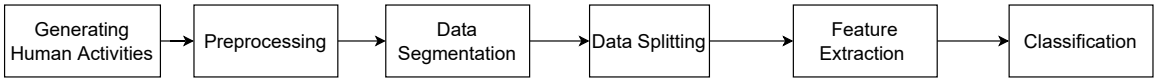


Fig. 2.2 Sequential phases of building a HAR system

2.4.1 Generating Human Activities

Human activities data are the foundation of HAR systems. HAR techniques exploited both smart home environments and wearable sensors. Commonly smart home environment sensors include motion sensors, PIR sensors, state-change sensors, pressure sensors and contact switch sensors [2, 135]. Often, these are non-intrusive binary sensors and equipped in smart environments to detect and read various human activities such as sleeping, showering, leaving, cooking, sitting and watching TV [38, 19, 136, 137]. Hence smart home sensors are used to collect data from human activities. Different from smart environments, wearable sensors such as accelerometers and gyroscopes must be worn to collect human activity data [74, 138, 12, 139]. In this thesis project, 12 public and benchmark collected datasets of human activities from the smart environment and wearable sensors are used to evaluate our proposed methods for improving HAR systems and to address the challenges of HAR. Each of these datasets is described in the following sub-sections.

Ordonez Smart Home A and B Datasets

Five human Activities of Daily Living (ADLs) recorded using binary sensors in real smart homes from public datasets are used for the evaluation. Among them, two datasets contain the sensor data from residents' daily routine, referred as to Ordonez Home A and B [111]. These two smart homes are typically equipped with different binary sensors that can capture the daily physical activities of the residents. The binary sensors in these datasets are passive infrared (PIR) motion detectors to detect physical activities and interactions in a limited area, pressure sensors on beds and couches in order to detect the user's presence, and reed switches on cupboards and doors to measure open or close status, and float sensors in the bathroom to measure toilet being flushed or not. Table 2.2 provides details of the two Ordonez smart homes A and B with information about the inhabitants, and the number of activities and sensors. In Ordonez Home A, nine physical activities were carried out in fourteen days over a period of 20,358 min, where data were recorded by twelve sensors in the home. In Ordonez Home B, ten physical activities were carried out in twenty-two days over a period of 30,469 min, where data were recorded by twelve binary sensors. The timeline of the physical human activities for all the smart homes data is segmented in time slots using the window size Δt = 1 mi. The activities of common activities from Ordonez Homes A and B are *Breakfast, Lunch, Sleeping, Grooming, Leaving, Idle, Snack, Showering, Spare Time/TV, and Toileting*, respectively. In addition to these activities, Ordonez Home B has the activity *Dinner*. T

Kasteren Smart Home A, B and C Datasets

The Intelligent System Laboratory (ISL) [140, 141] collected human activities from three other environments equipped with binary sensors as well, refer to as Kasteren home A, B, and C. The details of these datasets are shown in Table 2.3. In Kasteren home A, ten daily human activities were carried out in twenty-five days over a period of 40,005 min, which were recorded from fourteen sensors in smart home A. In Kasteren home B, thirteen daily physical human activities were carried out in fourteen days over a period of 38,900 min., which were recorded from twenty-three binary sensors. In Kasteren home C, sixteen daily human activities were carried out in nineteen days over a period of 25,486 min., which were carried out from 21 binary sensors. The timeline of the daily human activities for all Kasteren smart homes is segmented in time slots using the window size Δt = 1 min. as well.

Table 2.2 Details of recorded Ordonez datasets

	Home A	Home B
Setting	Home	Home
Number of days	14 days	21 days
Rooms	4	5
Sensors	12	12
Sensors	PIR(Shower,Basin,Cooktop), Magnetic (Fridge, Main door, Cabinet, Cupboard), Pressure (Bed, Seat), Flush (Toilet), Electric(Toaster, Microwave)	PIR (Basin, Shower, Door Bedroom, Door Bathroom, Door Kitchen) Magnetic (Cupboard, Fridge, Main door) Flush (Toilet) Pressure (Bed, Seat) Electric (Microwave)
Number of Activities	9	10
Activities	Showering, Toileting, Dinner, Leaving, Breakfast, Lunch, Sleeping, Grooming, Snack, SpareTime/TV	SpareTime/TV, Grooming, Leaving, Lunch, Showering, Sleeping, Dinner, Breakfast, Snack, Toileting

UCI-HAR Wearable Dataset

Dataset for human activity recognition was build by recording activities of daily living (ADL) of 30 study participants while carrying a waist-mounted smartphone with embedded inertial sensors [142, 143]. The participants within an age bracket of 19-48 years performed six daily activities in which three activities are static postures (standing, sitting, lying) and three activities are dynamic activities (walking, walking downstairs, and walking upstairs). The participants wore a smartphone (Samsung Galaxy S II) on the waist to record the activities. Embedded accelerometer and gyroscope were used to capture 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The activities were video-recorded to manually annotate the dataset.

Wearable Wireless Identification and Sensing Data *Roomset1* and *Roomset2*

Fourteen elderly volunteers from 78 to 78 $\pm$4.9 years old wore Wearable Wireless Identification and Sensing Platform (W^2ISP) tag [144–146]. The W^2ISP placed on top of their garment at the sternum level to capture trunk movements and recognize activities: i) sit on bed; ii) sit on chair; iii) lying; iv) ambulating. The activities were performed in two clinical room configurations (*Roomset1* and *Roomset2*) for ambulatory monitoring of older patients. Table 2.4 shows that these two datasets are collected to record four activities from 14 participants.

Table 2.3 Details of recorded datasets of the Kasteren Smart Home A, B and C

	House A	House B	House C
Activities	10	13	16
Duration	25 days	14 days	19 days
Gender	Male	Male	Male
Rooms	3	2	6
Setting	Apartment	Apartment	House
Sensors	14	23	21
Age	26	28	57
Activities A	Drink, Brush-Teeth, Go-to-Bed, Snack, Leave-house, Prepare-Breakfast,Use-Toilet, Shower, Prepare-Dinner		
Activities B	Eat-Brunch, Brush-Teeth, Drink, Eat-Dinner, Dressed, Prepare-Dinner, Go-to-Bed, Prepare-Brunch, Leaving-house, Use-Toilet, take-shower, Wash-Dishes		
Activities C	Get-Drink, Eating, Brush-Teeth, Get-Dressed, Get-Snack, Leave-House, Go-to-Bed, Prepare-Dinner, Prepare-Breakfast, Take-Shower, Prepare-Lunch, Take-Medication, Shave, Use-Toilet-Upstairs, Use-Toilet-Downstairs		

HHAR Dataset

TThe Heterogeneity Activity Recognition dataset [147] is recorded using two different smartphone sensors i.e. gyroscope and accelerometer as well as smartwatch sensors for six human daily activities. The human activities that are performed by participants are standing, biking, walking, sitting, stairs-up and stairs down. The data of the activities are recorded from nine participants (aged between 25 and 30) for five minutes to render similar class distribution. The participants wore eight smartphones attached to the participants' waists and four smartwatches (two on each arm). In this study, due to the need for the evaluation process and compatibility among datasets, only the triaxial accelerometer data is used. Various devices were used in recording the human activities , i.e. LG smartphone Nexus 4 (200 Hz), Samsung smartphone Galaxy S Plus (50 Hz), Samsung smartphone Galaxy S3 mini (100 Hz), Samsung smartphone Galaxy Wear (100 Hz), Samsung smartphone Galaxy S3 (150 Hz), and LG G smartphone (200 Hz). Table 2.4 shows that this dataset is collected from six activities of nine participants.

UniMiB SHAR Dataset

The UniMiB SHAR dataset [148] were recorded for HAR and fall detection from 30 healthy participants of different ages, ranging from 18 to 60 years old, heights ranging from 160 to 190 cm, and body mass ranging from 50 to 82 kg. Participants were requested to complete nine different activities and eight types of falls based on their popularity in other public datasets. A smartphone Samsung Galaxy Nexus I9250 was placed in the front trouser pockets of the participants to record their activities at 50 Hz. Moreover, audio recordings were also collected to support data annotation. The participants follow four different designed seqnences to perform their activities with ease. To record the activities, the smartphones were placed in trouser pocket during the experiments. The participants signed the consent form in accordance with the World Medical Association (WMA) Declaration of Helsinki. The collected data contains different nine activities including standing up from laying, lying down to standing, standing up from sitting, running, sitting down, going downstairs, going upstairs, walking, and jumping. Table 2.4 shows that this dataset is collected from nine activities of 30 participants.

MotionSense Dataset

The MotionSense dataset [149] was collected from 24 participants 14 men and 10 women using the accelerometer and gyroscope sensors of an iPhone 6s which was placed in the user's front pocket. The participants performed six activities which are walking downstairs, walking upstairs, walking, jogging, sitting, and standing. The participants had different ages, gender, weight, and height, i.e. their body mass ranged from 48 kg to 102 kg, their age ranged from 18 and 46 years old, and their height ranged from 161 cm to 190 cm. 15 trails in the same environment and condition to perform the activities by the participants were conducted. Each trail lasted between 30 seconds and three minutes. In this study, only the accelerometer sensor data are used. Table 2.4 shows this datasets consists of six activities from 24 individuals.

WISDM Dataset

The WISDM (Wireless Sensor Data Mining) project aimed to explore the problems of receiving sensor data from smartphone devices. The data was recorded from 29 volunteers in a controlled study for 6 different activities, i.e. standing, walking, sitting, jogging, walking upstairs and walking downstairs. The volunteers carried a smartphone (Nexus One, HTC Hero, or Motorola Backflip) to record their activities via an app developed for an Android

phone in their trousers' front pocket and performed each activity at different times. Table 2.4 shows that this dataset is collected from 36 participants for six activities.

Table 2.4 Number of Activities and participants of Wearable Sensor datasets

Datasets	Number of activities	number of users
Roomset1	4	14
UCI HAR	6	30
HHAR	6	9
UniMiB	9	30
WISDM	6	36
MotionSense	6	24
Roomset2	4	14

2.5 Sensor Data Processing

Raw recorded sensor data from smart home environments and wearable devices typically are not directly utilised for modelling. Explicitly raw data sensors need to be passed through processing and transformed into a readable format to be prepared and used by machine learning models[7, 74]. The processing stages are the most fundamental phase of machine learning projects and play a significant role in gaining insight from the data. Data processing consists of different phases including preprocessing, feature extraction and segmentation as shown in Figure 2.3. Each of these phases comprises different steps to further transform the data into an understandable format. The details of the phases and steps are presented below.

2.5.1 Sensor Data Preprocessing

Preprocessing data is an essential phase to prepare a well-readable format of the data. Collected raw data from various sensors of intelligent environments or wearable devices are inherently noise and contain missing values, or require to be transformed to a proper attributes format based on the systems requirements. Furthermore, The steps of data preprocessing practically include the data cleaning step to remove signals that carry irrelevant information. Data interpolation to manage the missing values in the raw sensor readings. Finally, data transformation to produce an understandable format of the raw collected sensor data [7, 74].

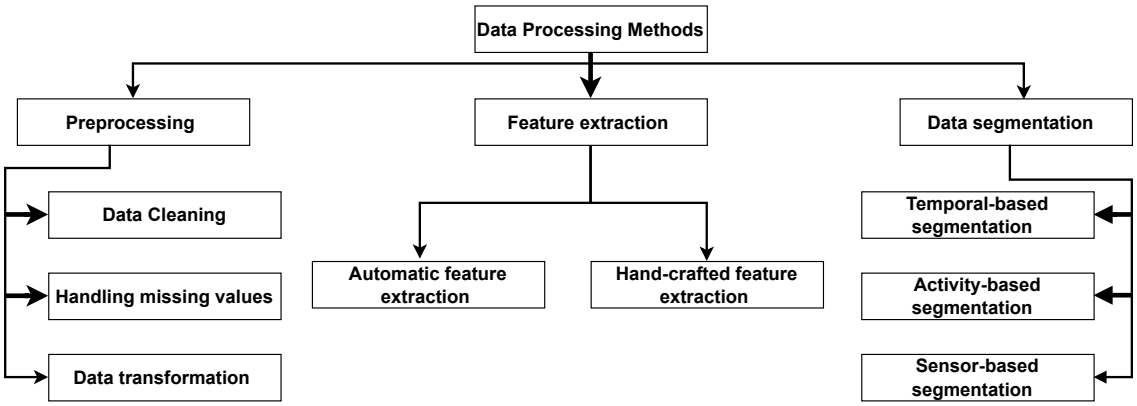


Fig. 2.3 Sensor Data Processing

i. Data Cleaning:
Recorded data from multiple sensors of smart home environments or wearable devices can contain inaccurate information including noise, redundant sensor signals, and sensor failure readings. Hence, data cleaning is an effective procedure to tackle these problems in the raw data. Different techniques are used to address these issues in the raw collected data to properly build the machine learning input datasets. For instance, Wilson and Atkeson [110] applied two preprocessing techniques to clean the data on four binary sensors and an RFID sensor. The cleaning techniques are Bayes and particle filters. After preprocessing the outcomes reveal that the Bayes filter performs well in a noisy setting on tracking a single person. On the other hand, the particle filter performs better compared to the Bays filter in scenarios with multiple people. Noury and Hadidi [150] used preprocess techniques to remove nonlinear artefacts and redundant information by employing a median filter and the first-order-hold filter respectively. Moreover, Guattari et al. [151] also employed a median filter to prevent abnormal measurements within a smart home environment from passive presence sensors to record human presence.

ii. Handling Missing Values:
Recorded sensor data from wearable devices or smart home environments may contain missing values. For example, recorded data using RFID sensors may contain up to 30% of missing values [7]. To manage these missing values, data interpolation can be utilised. To fill in these missing values in the raw recorded data mostly linear interpolation technique is conducted. Muaaz and Mayrhofer [152] used a smartphone to record data and measure parameters related to users' walking. To record the data an Android API is used and the accelerometer sensor could not provide data at equal

intervals. Hence, they used linear interpolation to reshape the data generated by the accelerometer sensor into equal intervals.

iii. Data Transformation:
Data transformation is used on the collected raw sensor data to prepare the data in an appropriate format based on the system requirements [7]. Rodner and Litz [153] conducted associated rule mining to model users' behaviour for ambient assisted living systems from a smart home environment. The raw sensor data were recorded using a motion sensor integrated with a lux meter. The format of the motion sensors data is a timestamp with a numeric, motion is binominal and lux is numeric. To prepare a suitable format for the rule mining the numeric values from the raw sensors data are converted to nominal values.

2.5.2 Data Segmentation

Once the preprocessing steps on sensor raw data including cleaning, handling missing values and transformation are completed for human activity recognition, the data should be passed through segmentation to be prepared as input datasets for machine learning models [7, 76, 81, 154, 155]. Multiple data segmentation techniques are used to build the input datasets for machine learning models. Five data segmentation techniques are described below.

i. Time based segmentation:
Time base segmentation is the most common sliding window with a fixed time window, for example, 60 seconds to segment the data into chunks of equal time duration [76, 81, 154, 155]. Ordonez et al. [101] used time-based segmentation with a one-minute time interval to build a dataset from sensor data. The time was designed to properly distinguish human activities. However, finding the best value for the time interval of the segmentation process is a challenging task to properly build the datasets [156–159]. For example, a short time interval for data segmentation may render duplicate human activities, particularly for activities with a long duration such as sleeping, which results in producing an imbalanced dataset. On the other hand, a long interval of time may overlap human activities by merging some activities into the same window which leads to losing information about some activities [160, 161]. Hence, to design a proper segmentation technique with a suitable time interval, effective heuristics are required.

ii. Activity-based segmentation

Segmentation of sensor data based on human activities includes each activity in a window by considering the start and end time of the activity. This approach often fails in accurately segmenting data because human activities are not well distinct and the boundaries of the activities are not precise. This approach rather can be used for annotating human activities. However, this activity-based data segmentation is not applicable for online HAR because the decision will be delayed due to waiting for future data [7, 154]. Yoshizawa et al. [162] applied this approach to separate static activities such as sleeping, sitting, and watching TV, from dynamic activities such as walking, and leaving home.

iii. Sensor-based segmentation

This method segments raw sensors based on sensor events to generate input datasets for HAR systems. Each window has a sequence of actions or movements and the raw sensor data are split into windows of an equal number of sensor events. Hence the duration of windows is different from one window to another since some sensor events could be very short while some other events could be long [120, 163, 164]. The labels of the windows are the labels of the last event in each window. The sensor events from the sensor data that precede the last event in the window during the segmentation process define the context for the last event. The drawback of this method is that in each window the last event may belong to activity "A" while all previous events in the same window belong to activity "B".

iv. Fuzzy temporal window segmentation

In addition to the aforementioned segmentation methods, dynamic sliding windows are also utilised in multiple studies to segment raw sensor data and generate input datasets based on activities or sensors ID [81, 158, 159]. Banos et al.,2014 [157] compared multiple window sizes with a non-overlap sliding window to segment sensor data and build datasets for HAR. The results of the study uncovered that a small window size typically renders better recognition performance. However, considering the human interpretation, these methods of data segmentation insufficiency capture a longer temporal representation which is essential for HAR and has been recognized as an important element of the performance of sliding window techniques. Therefore multiple incremental fuzz temporal windows (FTWs) [38] are proposed and extended by [19] to segment the timeline of human activities from sensor data and capture long and short-term human activities. FSWs are compared with other approaches such

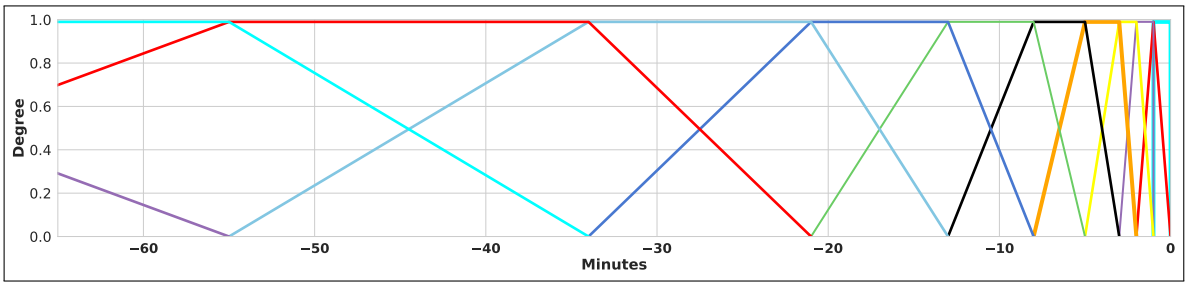


Fig. 2.4 Example of multiple incremental fuzzy temporal windows to segment raw sensors data [23]

as time base windows, sensor-based segmentation and activity-based segmentation [38, 137].

Fuzzy temporal windows (FTWs) are used to generate the input datasets from the raw sensor data for training temporal models. A fuzzy set that introduces each FTW T_k is specified by a membership function. The shape of the fuzzy set corresponds to a trapezoidal function $T_k[l_1, l_2, l_3, l_4]$ is shown in Equation (2.1) [23]. Four values that define the trapezoidal membership functions are a lower limit l_1, an upper limit l_4, a lower support limit l_2, and an upper support limit l_3. The four values of these l_1, l_2, l_3, l_4 are provided by the Fibonacci sequence which has recently been successfully used for introducing FTWs. The Fibonacci sequence can easily be used for FTWs to build training datasets without involving a knowledge expert definition[39, 38, 19]. Figure 2.4 shows 12 multiple and incremental FTWs are designed based on the Fibonacci sequence. To generate a training input dataset, the FTWs are slided over raw sensors data x in every minute according to Equation 2.1: For example, in Ordonez smart homes A, the training input dataset is generated by applying 15 FTWs on the raw sensor activations from all 12 binary sensors in every minute. The datasets of Ordonez smart home A and B have 20358 and 30469 examples respectively, where each example represents one minute of data with $12 \times 15 = 180$ features. Algorithm 1 shows the procedure of using FTWs to generate input training datasets.

$$T_k(x)[l_1, l_2, l_3, l_4] = \begin{cases} 0 & x \leq l_1 \\ (x - l_1)/(l_2 - l_1) & l_1 < x < l_2 \\ 1 & l_2 \leq x \leq l_3 \\ (l_4 - x)/(l_4 - l_3) & l_3 < x < l_4 \\ 0 & l_4 \leq x \end{cases} \tag{2.1}$$

Algorithm 1: Fuzzy temporal windows to generate input Datasets [23]

input: $Sensor\_data$ Sensor data from smart homes
$FTWs \leftarrow Fibonacci_sequence$ Fibonacci Sequence to define values of FTWs
$Intervals\_sensor \leftarrow Sensor\_data$ Raw sensor data
for $ftw \leftarrow FTWs$ **do**
  for $interval\_sensor \leftarrow Intervals\_sensor$ **do**
    apply ftw on $interval\_sensor$
  end for
  $extracted\_features \leftarrow maximum(ftw)$
end for
$datasets \leftarrow extracted\_features$
output: $datasets$

2.5.3 Data Splitting Techniques

Data splitting is the process of dividing the dataset into three parts training set, validation set and testing set. Usually, for machine learning algorithms, a set of data is required to be used by a model for training. Besides, a set of data for validating the model performance is also required during the training process to tune the hyperparameters. Also, a set of data which has not been used in the training process is required for the testing set to evaluate the trained model [165].

Different types of data-splitting techniques have been presented for HAR systems. The most common techniques are leave-one-day-out cross-validation, leave-one-subject-out cross-validation, k-fold cross-validation and percent splitting technique. The leave-one-day-out cross-validation is mostly used when data is collected for many days of a subject and then one-day data could be held-out for evaluation of the trained model and the rest of the days' data for the training and validation. Often a subject performs the same daily activities differently at different times, days, and places which are known as the case of *intra-subject variation* [13, 165]. To handle this case, the leave-one-day-out cross-validation technique has been mostly used to focus on a robust model which can adapt to the activities of a subject at different times, days and places.

Besides, one-subject-out-cross validation has been mostly used for developing a robust model which can adapt to the heterogeneity on the subject level. This technique is mostly used

when data is collected from multiple individuals. Therefore collected data of some subjects are used for training a model and the hold-out subjects are used for evaluating the model. Collecting the same set of activities from different subjects brings *intra-class variations* since each individual performs the activities based on their preferences and posture. Hence this technique is used with the aim to learn a model which is robust and renders satisfactory performance despite the heterogeneity at the subject level [27, 166–170].

Another wildly popular technique is k-fold-cross validation (KFC) in which the input dataset is split into k numbers of equal folds. These folds can be created by the number of instances, subjects or days in the input dataset [171–174]. The folds are used for training and evaluation of a model in which a fold is utilised as a hold-out for validating the model and the other folds are used for training the model. This technique uses all the folds in training and evaluation. Furthermore. this technique is mostly recommended for the small datasets to use each instance in the datasets for training and evaluating the model [175, 174].

The percent splitting technique is the process of dividing the input datasets based on a percentage such as 70%-30%, 75%-25% and 80 %-20%. Then the larger portion of the dataset, i.e. 70%, 75% or 80 % is fed into the model for training and the smaller portion is sued for evaluating the trained model. This splitting technique is mainly utilised in general problems of machine learning and has been utilised in many HAR systems [176–181].

2.5.4 Extracting and Selecting Features

After sensor data segmentation, feature extraction is performed to extract relevant information from the segmented data. The extracted features are fed to classification methods to assign the correct activity class. Extracting informative features from the sensor data is one of the key steps toward developing a successful machine learning model [36, 76, 139, 154, 165, 182, 183]. Thus, feature extraction is the procedure of determining the essential and relevant information about human activities from the sensors stream to distinguish one activity from other human activities. Besides, feature selection is conducted after feature extraction to select only certain features. Feature selection reduces dimensions, storage requirements, and the training and testing time to improve classification performance [184]. Manual feature selection is a challenging task, time-consuming and requires domain knowledge experts in HAR. Several techniques for automatically ranking and selecting features are available that can be grouped into the wrapper, filter and hybrid methods [185].

The feature extraction process could be broadly classified into two types: hand-crafted features by a domain knowledge expert or automatic feature extraction by deep learning models [186] which are described in detail below.

Hand-crafted Features

Hand-crafted features are extracted manually designed or engineered by a domain knowledge expert to compute measures from the segmented data to capture an informative representation of the data and distinguish activities in HAR systems. Wide range of studies have explored hand-crafted features to improve performance HAR systems [187–191]. Even though hand-crafted features are time-consuming and require domain knowledge experts, hand-crafted features are viably used for HAR since the features extracted are computationally lightweight to be processed particularly in ubiquitous devices [192]. The hand-crafted features include two techniques which are time domain features (TDM) and frequency domain features (FDF) [7, 165].

i. Time domain features (TDM)

   TDF is about extracting features of human activities from sensor data sequences by statistical measures. TDF features have been intensively studied in various applications and demonstrated to be useful and effective for HAR systems. statistical measures include meaning, median, standard deviation, average, variance, minimum value, maximum value, range, signal vector magnitude (SVM), root mean square (RMS), zero-crossing, zero-Correlation, difference, correlation, cross-correlation, integration, differences, velocity, signal magnitude area (SMA). Among the statistical measures of TDF, the mean value is one of the most common features for HAR systems due to its low computational cost compared to other measures and minimum memory requirements. The standard deviation that denotes the stability of a signal in the sensor data around its mean value is often utilised as a fundamental classifier metric and threshold-based method. [193] studied standard deviations for HAR to distinguish similar human activities such as walking, standing and stairs. Multiple other well-applied and studied TDF are median [194], zero crossing rate [195], variance [196], skewness [67, 197], autoregressive coefficient [67].

ii. Frequency domain features (FDM)

   Frequency domain features are the features that describe the periodic structure of sensor data. A set of FDF is used in HAR systems to extract useful features including wavelet

transformation, and Fourier transform. For instance, the wavelet transformation is often utilised to detect the transition between different human activities since it can capture sudden changes properly in signals of the sensor data [7, 165]. Moreover, extracted features by Fourier transform are reported to be useful in improving the performance of HAR systems [67].

Automatically Learned Features (Deep Learned Features)

The second technique to learn features from raw sensor data is using deep learning for HAR systems. With the satisfactory improvement in deep learning-based computation, applying explicit feature learning is gradually decreased. Deep learning techniques can be used for both operations learning features and the classification of human activities together in an end-to-end manner[198]. Deep learning techniques have accelerated the development of HAR systems by removing hand-crafted feature learning. Deep learning also lifted the burden from the shoulders of investigators to acquire a domain knowledge expert for the feature engineering process.

Convolutional Neural Network (ConvNet) is particularly known for its ability to automatically learn features and extract the local pattern using convolution layers. ConvNet uses multiple filters with different sizes of kernels to extract high-level interpretations from the segmented sensor data to represent generic features for downstream tasks i.e. classification. A wide range of studies has explored ConvNet for extracting features and reported the state-of-the-art performance [173, 179, 199–202]. The details of ConvNet are provided in Section 2.8.4.

LSTM has also been commonly employed in HAR systems to learn the temporal sequence dependencies relationships from raw sensor data. The performance of HAR using LSTM has been explored to specify the role of temporal sequence relationships in the raw sensor data particularly long-term dependencies in human activities [169, 171, 179, 180, 203]. The details of LSTM are presented in Section 2.8.3.

The hybrid of both ConvNet and LSTM has been developed to learn and classify human activities. Often ConvNet is used to extract local patterns from the sensor data whereas LSTM focuses on learning temporal sequence dependencies from features maps through LSTM cells. Various hybrid models are proposed to improve HAR systems and their results have proved to be successful in feature learning [170, 181, 203–205]. The details of the hybrid of ConvNet and LSTM are presented in Section 2.8.5.

Feature learning based on both the hand-crafted features and deep-learned features is explored to further improve the performance of HAR systems [67, 167, 171, 174, 206–208].

Ignatov [168] used both the hand-crafted features and deep-learned features to enhance HAR systems on two datasets. Statistical features are used as the hand-crafted features and ConvNet for deep learned features. Similarly, [170, 209] these studies performed feature learning using the hand-crafted features and deep learned features through ConvNet. They explored and evaluated the performance of deep models for HAR systems with hand-crafted features and without hand-crafted features.

2.5.5 Activity Classification

The last phase of the HAR workflow is activity classification. Activity classification for HAR systems denotes the procedure of uncovering the hidden relationship between the extracted features from the input dataset with specific activity classes according to the employed classification principle [21]. Hence, supervised learning algorithms perform classification by learning a mapping function from the input sensor data to human activities. To do that, the classification algorithms minimise a loss function of the pairs of input data and the corresponding output data, i.e. human activity class [2]. Supervised learning algorithms are trained on a training dataset to perform classification for detecting patterns between the feature maps and the classes of human activities. Then the training model is validated on validation datasets with different samples to optimize the model and finally, the model is evaluated on the testing datasets. Comparing the classification outcomes by the trained model with the true activity classes enables assessing the accuracy of the model.

There are two classification groups for HAR systems. The shallow or traditional machine learning algorithms include Naïve Bayes (NB), Support Vector Machine (SVM), Hidden Markov Model (HMM), k-Nearest Neighbour (k-NN), Logistic regression (LR), Decision Tree (DT), and Random Forest (RF). The other group is deep learning algorithms such as ConvNet and LSTM [138]. These classification groups are machine learning algorithms which are described in detail in Sections 2.7 and 2.8.

2.5.6 Sequence Modeling

Before describing the details of the machine learning models, we show the sequence modelling task for human activities. Input human activity sequences $x_0, ..., x_T$ are fed into a model to predict corresponding activity outputs $y_0, ..., y_T$ at each time. Predicting the activity output y_t for particular time t should be derived only by considering the observed times steps before time t: $x_0, ..., x_t$ [1, 2, 210, 211]. Hence, sequence modeling is a function

$f : x_0, ...x_T \rightarrow y_0, ...y_T$ (where x and y are the input and output respectively) that renders the mapping as shown in Equation 2.2.

$$\hat{y}_0, ..., \hat{y}_T = f(x_0, ..., x_T) \tag{2.2}$$

The model f is expected to minimize a loss L between, $L(\hat{y}_0, ..., \hat{y}_T, f(x_0, ..., x_T))$, the actual label and the predicted outputs where the input sequential data and the outputs are rendered based on some distribution. This formalism could not directly be used for domains such as sequence-to-sequence prediction or machine translation since these domains require the entire sequence input (past and future states) [210]. However, the setting can be extended for these domains.

2.6 Machine Learning

Machine learning algorithms have been broadly developed, improved and adapted for HAR. In this section, we describe different types of machine learning methods that have been mainly used for HAR. Machine learning algorithms are generally categorized into two basic approaches: supervised learning and unsupervised learning.

- Supervised machine learning algorithms learn a function that maps instances of the input data to an output variable[212]. The algorithms learn the relationships between the input and output variables by minimizing a loss function on the input and output pairs. The learned relationship between input and output variables is known as a model. Supervised learning models are used for predicting the output of unseen instances given the values of the input data [2]. The supervised learning models can be classified into two main categories: classification and regression. The main difference between classification and regression is the output variable. In classification, the output variables are categorical values while the output variable of the regression is real values. For example, classification in HAR maps features that are extracted from the sensor's raw data to their corresponding human activity labels. On the other hand, regression in HAR maps features of the extracted sensors' raw data to predicted time points in turn in the future when the human activity will next happen [2].

- Unsupervised machine learning is used to learn from unlabeled input datasets [213]. Different from supervised learning which learns on annotated data, unsupervised learning is mostly an approach to modelling the probability density of the input datasets and learning the inherent structure of the unlabeled input data. Unsupervised learning

is used for different tasks such as clustering, dimensionality reduction of features, and representation learning [213]. Unsupervised machine learning approaches can be used to uncover certain similarities or distinctive structures within the input datasets. The most common unsupervised techniques are k-means clustering, and principal component analysis (PCA).

2.7 Shallow Machine Learning

Conventional machine learning methods such as Conventional machine learning methods have shown great progress and reached satisfying performance on HAR. Such methods include NB, SVM, LR, k-NN, DT, and RF [40]. These methods have made remarkable progress on HAR and rendered reasonable outcomes. SVM is widely employed for classification problems and aimed at finding the hyper-planes also known as the decision boundaries to separate the data into classes [214]. SVM is employed for HAR and results in either it surpassing or is on part with several prior methods [215]. SVM was conducted for detecting abnormal activities by identifying the normal activities [216]. Abnormal activities are commonly unexpected events that happen in random forms. SVM has been applied in considerable studies of HAR and obtained reasonable accuracy [40, 217–224]. A brief introduction about each of the aforementioned conventional machine learning algorithms are provided.

2.8 Deep Learning

Traditional machine learning algorithms have shown reasonable performance for HAR. However, due to the intrinsic complexity of physical human activities, traditional machine learning models may not successfully learn non-linear relationships among sensor-generated data. Besides, the traditional machine learning algorithms rely on hand-craft features which need domain knowledge experts which is the most expensive and time-consuming. Furthermore, feature extraction become a pre-step for classification thus leads to sub-optimization [12, 225]. In contrast to traditional machine learning methods, deep learning models, i.e. are able to automatically learn complex features from the sensors generated data and jointly optimize both feature extraction and classifier learning [76].

Deep learning is essentially based on artificial neural networks as a part of machine learning and is a subfield of artificial intelligence that focuses on developing large neural network models that can make accurate data-driven decisions [226, 227]. Similar to ma-

chine learning, the deep learning approaches can become supervised, semi-supervised or unsupervised learning [227]. Deep learning has been used in many applications such as computer vision, natural language processing, and robotics. Deep learning models enable automatically learning numerous levels of representations of the underlying distribution of the input modelled data. Deep learning models have shown superior learning and promising results in different applications such as bioinformatics, medical image analysis, speech recognition, audio recognition, social network filtering, and machine translation [228–230].

This thesis focuses on deep learning methods and their performance on the raw sensor data for HAR systems. A deep overview of the deep learning models, i.e. deep neural network (fully connected layers), convolutional neural network and long short-term memory that have been used in this thesis are provided below.

2.8.1 Multi-Layer Perceptron (MLP)

Deep neural network (DNN) has been developed as computational systems that emulate the abilities of biological neural network and process information via interconnected computational neurons (nodes or units) [212]. Neurons in deep learning are associated with a particular weight. DNN consists of multiple layers of interconnected neurons, each layer is built upon the previous layer to distil and optimize the classification. In DNN, the progression of computations of the neurons through the network is named forward propagation. Moreover, DNN can also servers as a dense layer of other deep models. For instance, several dense layers are often added to the convolutional neural network layers or long short-term memory layers for prediction or categorization. Different from shallow learning models, DNN models usually have deep architecture by including multiple hidden layers in between input and output layers which are called visible layers. The input layer of the network is used to ingest the input data for processing and the output layer of the network is used to render the classification. Furthermore, DNN has more capable of learning from large amount data compared to shallow learning models. The aim of the DNN is to learn some function f for instance $y = f(x;\theta)$ that maps the input data x to output by learning optimal parameters θ [212]. Figure 2.5 shows computation of a neuron where inputs $x_1 ... x_n$ with their corresponding weights $w_1 ... x_n$ and a bias (b) are fed into an activation function f to learn non-linear patterns of the features. The output layer uses the softmax activation function to transform the output data to a probability distribution over predicted classes. Furthermore, in regression problems, the activation function is not required to be applied to the raw numerical

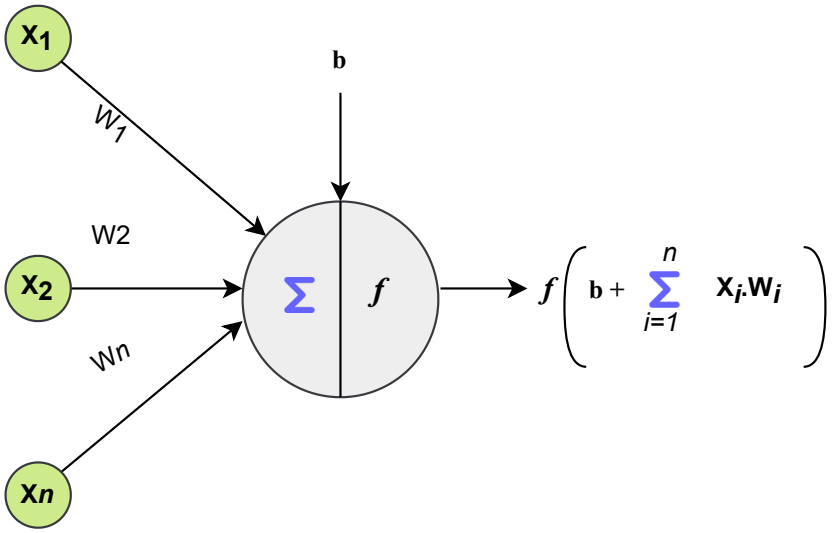


Fig. 2.5 An example of a neuron showing input ($x_1 ... x_n$) with their corresponding weights ($w_1 ... x_n$) with a bias (b) and the activation function f applied to the weighted sum of the inputs

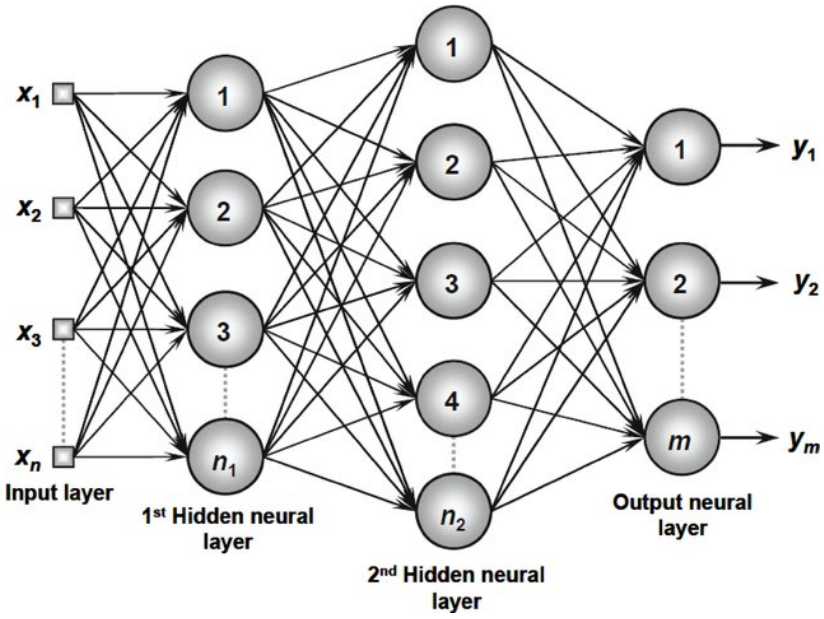


Fig. 2.6 An example of fully connected multi-layer neural network[231]

output. The activation functions are non-linear functions such as Sigmoid, hyperbolic tangent, rectified linear unit, or softmax, etc.

The fully connected multi-layer neural network is a typical structure where neurons are arranged in layers and each neuron from a particular layer ingests the outcomes of all the

neurons from the previous layer as its inputs. Such a structure with multiple layers allows us to uncover complex decision boundaries for the classification tasks[212]. Figure 2.6 shows the general architecture of the fully connected multi-layer neural network.

Backpropagation is another process in DNN which uses algorithms like gradient descent or stochastic gradient descent to minimize the cost function by computing the error in predictions and adjusting the weights. The process is used to train the DNN by moving backwards through the layers of the model. Both processes forward propagation and backpropagation enable DNN to produce predictions and correct any errors accordingly. In this thesis project, fully connected layers are only used and stacked on the layers of the convolutional neural networks and long short-term memory.

2.8.2 Recurrent Neural Network

Fully connected layers, i.e., dense layers can process two-dimensional inputs $x = (x_1, ...x_N) \in R^{N \times F}$ where N are the instances and F is the features, however, often complicated systems such as smart homes generate high dimensional data with internal dependencies. For instance, sequential sensors data is typically represented as three-dimensional tensors $x = (x_1, ...x_N) \in R^{N \times T \times F}$ where T is the time steps in the sequential temporal data. Hence, each instance in the sequential data contains both time steps and their associated features. In the implementation aspect, fully connected feedforward layers can still process such sequential temporal data with three-dimensional however, it will not take the ordering patterns of the sequential data into account because the three-dimensional input data should be flattened ($T \times F$) before the dot products of the features with the weights. To address this limitation, new approaches to neural networks are proposed that share the same weights across multiple time steps and process sequential data. In this thesis, mainly variants of Recurrent Neural Networks(RNN) and convolutional Neural Networks(ConvNet) are used.

The Recurrent Neural Network (RNN) was initially proposed in 1991 [232] for finger alphabet recognition from 42 symbols. Latter in 1995 [233], RNN was used for hand shape recognition from 66 different shapes. RNN archived promising results with about 98% accuracy. Since then, the RNN has been used broadly for temporal sequential data including human activities or estimated hand gestures [234, 235]. Different from a feedforward neural network, the RNN models learn on sequential temporal data. RNN is distinguished by its memory as a unique feature to keep information from previous inputs to affect the current input and output. Unlike conventional machine learning algorithms which assume that the inputs and outputs of sequential data are independent of each other, RNN utilises information

from the prior sequences to learn the current sequence. Due to this reason, RNN is generally employed to learn the sequential temporal data.

Furthermore, extensive studies have been performed on the performance of RNN methods in the context of HAR and multiple methods based on RNN have been proposed including Continuous Time RNN (CTRNN) [236], Independently RNN (IndRNN) [237], Colliding BodiesOptimization RNN (CBO-RNN) [238], and Personalized RNN (PerRNN) [239]. Different from the previous methods with one-dimensional time series input, a hybrid ConvNet+RNN model is proposed for HAR from multi-model sensor data. This hybrid method uses hierarchical ConvNet to exploit the relationships among similar sensors and merge relationships of different sensor modalities and then apply RNN to model the temporal relationships of the sequential data [240]. RNN model is also utilised to address the domain adaptation task caused by intra-session, sensor position, and intra-subject variances [241].

RNN has been used in multiple studies of HAR and often RNN models report state-of-art performance. However, RNN suffers from the problems of vanishing and exploding gradient. To overcome this problem, Long Short-Term Memory Networks (LSTM) are proposed [242] to handle long dependencies of sequential data. In this thesis, LSTM is used to accurately apply HAR.

2.8.3 Temporal modeling via Long Short-Term Memory Networks

LSTM is an artificial RNN and has been used to learn from temporal sequential data. LSTM can handle and learn from long-term dependencies which alleviate vanishing and exploding gradient problems [242]. LSTM as a temporal model has been used to recognize ADLs from sensor data [19, 38]. LSTM processes temporal data using forget gate, input gate, and output gate to append or delete information to the cell state throughout the processing of the sequence data. The cell state is the main part of LSTMs that carry and transfer relevant information from earlier time steps to later time steps. Figure 2.7 shows the connection of the gates with the cell state in a single LSTM cell. The gates learn to keep relevant information and forget irrelevant information during training to update the information on the cell state. Hence each LSTM cell works as a memory to remove, read, and write information that is controlled by the forget, output, and input gates, respectively. Forget gate process both inputs the previous output h_{t-1} and new time step X_t using sigmoid activation function to indicate relevant or irrelevant information. The forget gate keeps the information if the outcome of the sigmoid function is 1 while deletes the information if the outcome of the sigmoid function is 0. Equation 2.3 shows how the forget gates within a single LSTM cell are computed.

The next step consists of two parts to determine new information kept in the cell state. The first part is the input gate that indicates new information from the current input (X_t, h_{t-1}) is appended to the cell state. The tanh activation function is the second part that renders $\tilde{C}_t$ a vector of new candidate values and can be added to the cell state. Equations 2.4, 2.5 show how the input gate and the new candidate values are computed, respectively. A new cell state C_t is generated based on the summation of the multiplication of these two parts and the multiplication of the forget gate with the previous cell state C_{t-1}. Equation 2.6 shows how the new cell state is computed. The multiplication of the previous cell state with the forget gate deletes part of the information which was decided to be forgotten earlier. Then the new candidate values are scaled by how much the cell state is updated using $it \times \tilde{C}_t$. Finally, the sigmoid activation function processes both the previous hidden state h_{t-1} and the current input timestep x_t to produce the output gate.

Finally, the output gate is computed based on filtered information using two different activation functions and also specifies the next hidden state. Then the tanh activation function processes the newly updated cell state. The output of the tanh functions multiplies by the output of the sigmoid function to render the next hidden state. The updated cell state and the newly generated hidden state pass information to the next time step. Equations 2.7 and 2.8 show how the calculation of the output gate and hidden state.

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{2.3}$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{2.4}$$

$$\tilde{C}_t = \tanh(W_C.[h_{t-1}, x_t] + b_c) \tag{2.5}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{2.6}$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{2.7}$$

$$h_t = o_t \times \tanh C_t \tag{2.8}$$

where x is the input data, σ is the sigmoid activation function, tanh is the hyperbolic tangent activation function, W is the weight matrix.

LSTM has been proposed for HAR systems and achieved promising results [138, 19, 23, 38]. Table 2.5 shows a list of the proposed LSTM methods with their performance for HAR systems from multiple HAR datasets. Chung et al. [245] performed a study on 15 participants and proposed a model-based LSTM. The study investigated simple activities using eight wearable sensors. Their proposed model is evaluated in a real-world scenario and

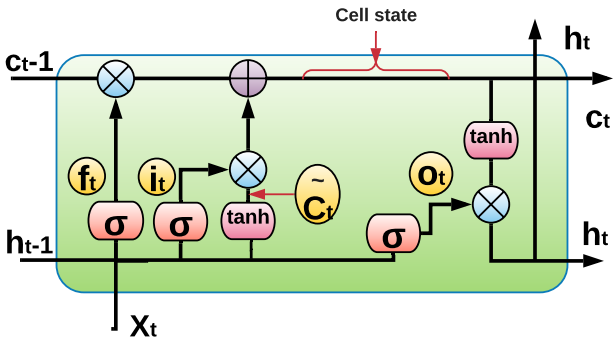


Fig. 2.7 Single LSTM cell

a controlled environment. The proposed LSTM-based model is integrated with an ensemble model to perform HAR on multi-sensor modalities. The challenge of this model is that the model is applied to recognize a few activities which were collected on a test bed and thus the proposed model is not suitable for large-scale applications. To develop a model for HAR systems with less computation power and reduced latency, a lightweight LSTM-based model was proposed by [249]. The proposed method was designed to be used on edge devices that decrease communication latency traffic and can be utilised for real case scenarios. However, this proposed method was only evaluated on a few activities before being taken into the real-time implementation of senior patients. Rashid et al.[251] extended the method that is proposed by [249] and proposed an energy-efficient and memory-efficient method for low-power edge devices. The method is validated by performing experiments on many activities using opportunity and w-HARdataset. Moreover, Zhao et al., [250] proposed a residual bidirectional LSTM method to utilise additional backward state information (negative time direction) in addition to the forward state information (positive time direction) to improve memory capability. The residual connection employed between the stacked LSTM cells in the proposed method avoids vanishing and exploding gradients. The methods proposed by [249, 251, 250] have relatively low performance on dynamic activities such as running, walking, or going upstairs, and postural activities such as sitting or standing. To address this problem, Wang and Liu [248] proposed a hierarchical deep LSTM model to improve the system's performance. Furthermore, to process the raw sensor data Ashry et al.,[252] proposed an online and offline model based on LSTM to process large labelled datasets using hand-crafted features as explained in [253, 254]. Although their proposed method produced a satisfactory performance, utilising hand-crafted features is not an optimal method to be used for the HAR system.

Table 2.5 List of LSTM models for HAR systems

References	Datasets	Performance
Chen et ali., 2016[243]	WISDM	92.1 %
Singh at el., 2017 [244]	Kasteren-Home A	89.8 % on raw sensor, 95.3 % on last-fired sensor
	Kasteren-Home B	85.7 % on raw sensor, 88.5 % on last-fired sensor
	Kasteren-Home C	64.22 % on raw sensor, 85.9 % on last-fired sensor %
Chung et al., 2018 [245]	Multimodal Sensor Test Bed	94.47 %
Sun et Al., 2018[181]	Opportunity	95.78 %
Zhao et al., 2018 [170]	UCI-HAR Opportunity	93.6 % 90.5 %
Tahmina at et., 2018 [169]	UCI-HAR	92%
Yu and Qin, 2018 [246]	UCI-HAR	93.79
Ullah et al., 2019 [247]	UCI-HAR	93%
Wang and Liu 2020 [248]	UCI-HAR HHAR	99.65 % 91.15 %
Agarwal et al., 2020 [249]	WISDM	95.78 %
Zhou et al., 2020 [250]	UniMiB SHAR	79 % F1-score
	Position aware HAR with wearable device	97 % F1-score on positioning the wearable device on chest and shine
Rashid et al., 2021 [251]	Opportunity w-HAR	91.57 % F1-score 97.64 % F1-score

2.8.4 Temporal modeling via Convolutional Neural Network

ConvNet consists of hierarchical structures that integrate learnable filters for convolutional operations and activation functions to introduce non-linearity, downsampling operations and classifiers. Models based on ConvNet map their input data into a more compact representation for downstream tasks based on their objective function. ConvNet layers can capture distinct features at different locations and could be used to extract more complex and abstract features from their input data [255]. ConvNet consists of several essential elements including sparse interactions since the size of the filter of the ConvNet is smaller than the input and each output value depends only on a small number of inputs. Each filter contains several kernels. In addition, ConvNet shares parameters which reduced the number of the parameters and reduces the computational cost. ConvNet models are equivariant to translation which indicates that if the input is translated to a convolutional layer, the output will translate accordingly [256]. Shared kernels in ConvNet allow the learning process of space invariant features. Each filter in the ConvNet has a defined receptive field that enables the ConvNet to capture local dependency in input datasets better than the fully connected neural network. The hierarchical structure of ConvNet models by stacking several layers contributes to its ability to collect low-level local features into high-level semantic meanings. This enables ConvNet models to learn distinct features as shown in [190] that compares the extracted features from ConvNet to heuristic hand-crafted time and frequency domain features.

ConvNet-based methods often utilise a pooling layer after each convolutional layer as shown in Figure 2.8. The pooling layer squeezes the learned representation and reduces the dimension of the convolutional map to support the methods against noise by dropping a portion of the outcome to a convolutional layer. Typically, a few fully connected neural network layers, i.e., dense layers, are added to the convolutional and pooling layers that reduce the dimensionality of the feature map being fed into the output layer. The final output layer of the ConvNet-based models is usually a softmax layer. However, as an exception, conventional classifiers are used in some studies as the output layer in ConvNet models [181, 257–260].

In 1998, LeNet5 model based on ConvNet was developed for hand-written digit recognition in documents and face recognition [261]. This method have shown excellent performance in visual classification tasks. Moreover, the convNet-based model was developed and proposed for autonomous systems slowly until its breakthrough in image classification [229] in 2012. In recent years, ConvNet has significantly improved image classification and object detection. Different ConvNet models are developed such as ZFNet [262] ,VGG [263], GoogleNet [264], BN-Inception [265], and ResNets [266] that have achieved promising

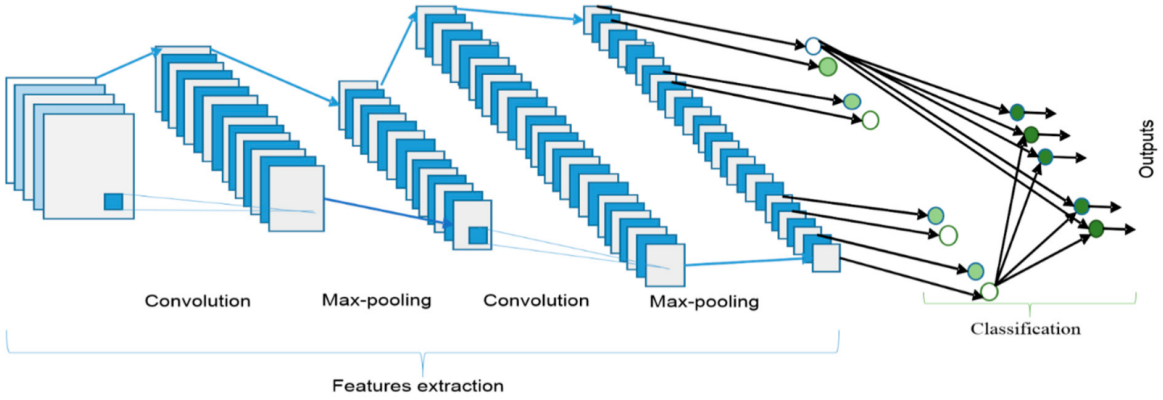


Fig. 2.8 An example of CovNet Based architecture for image domain [267]

results in image classification. Inspired by the success of ConvNet models for various tasks in the image domain, researchers have also utilised ConvNet for temporal sequential data.

ConvNet layers could be developed in different dimensions 1 D, 2D or 3D. In Image or video domains typically ConvNet models are developed with 2D or 3D, whereas 1D ConvNet models are more appropriate for the temporal sequential data. 1D ConvNet for human activity recognition has particularly two advantages which are local dependency and scale invariance. Local dependency refers to local movement patterns and signals correlation within human activities from the sequence input data. Besides, scale invariance refers to the ConvNet's power in detecting activity patterns even when the activity motion changes in some way, for instance, an individual may run with various movement intensities [19, 268]. Multiple ConvNet-based models have been proposed for HAR systems in different studies as listed in Table 2.6 with their performance from various HAR datasets.

Until recently RNN based models have been the dominating option for sequence modelling [269]. However, empirical results indicate that methods based on 1D ConvNet are competitive or even better than the RNN-based methods in a set of diverse tasks [210]. Furthermore, generally, recurrent-based models are severely suffering from two problems. Firstly RNN based models excluding LSTM are impotence in capturing very long-term dependencies in temporal sequential data. Secondly, recurrent-based models including LSTM are incapable to parallelize the sequential computation procedure. In contrast, ConvNet models are faster than LSTM models to train due to the inbuilt ConvNet parallelization strategy in processing input data. More formally, given an input dataset $x = (x_1, ...x_N) \in R^{N \times T \times F}$ where N are the instances and F is the features, and T is the time steps in the sequential temporal data. Stacked 1D ConvNet layers ingest the input datasets and scan over the sequence with 1D windows and learn filters $f : \{0, ..., k-1\} \in R$. The convolutional procedure C of a

sequential temporal data s is defined as

$$C(s) = (x * f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-1} \tag{2.9}$$

where $*$ is the convolutional operator between the input datasets and 1D convolutional windows, k is the filter size, and $s-1$ records the convolution step. the 1D window learns patterns from the sequential temporal data. Often after the convolutional operation, pooling is applied to reduce the dimensionality of the feature map.

Chen and Xue [270] have proposed a ConvNet-based to process and classify human activities from smartphone sensor data. Their proposed method adapts convolutional kernels according to signals of the wearable sensors' data. Moreover, a ConvNet model was proposed to automatically extract features from raw signals by Ranao and Chao [65]. The proposed method consists of several ConvNet layers and a pooling layer that cancels the impact of small translations in the input data [212]. Ignatov [168] proposed a ConvNet-based method to capture local dependency in the feature map in parallel with statistical features including mean, median, and variance of the temporal input data to maintain information about global features of raw sensors data. This study performed an investigation to uncover the impact of sliding windows in capturing the signal length on performance. Furthermore, Avilés-Cruz et al., [201] proposed a new method to extract local features using three parallel and different ConvNet subnetworks fine-ConvNet, medium-ConvNet, and coarse-ConvNet. The proposed method fuses these subnetworks into a single classification layer to classify and analyze human activities. This method is evaluated only on wearable sensor data and only on a few human activities. Besides, this model requires a high computing resource due to many parallel ConvNet layers with many filters to extract features and mappings.

To reduce the computation cost and energy consumption, a ConvNet model is proposed to extract high-efficiency local features from wearable sensor data. The proposed model is compared with multiple methods including LSTM, Bi-LSTM, MLP, and SVM. Besides, to reduce the computation cost and memory of the ConvNet, lego filters are utilised instead of the convolutional filters to develop a lightweight model in [271]. The proposed method performs better and reduces memory and computation costs compared to the ConvNet models. Also, the method is smaller, faster and more accurate compared to ConvNet-based models. The method can be employed without using any special network structure and computational resources. Cheng et al. [272] proposed a ConvNet model to process and classify human activities in real time on smartphones and wearable devices. The method is developed using

conditionally parameterized convolution to decrease the computation cost and reduce the size of the model.

To further minimize the computation cost and generate a more efficient ConvNet model multiple methods based on ConvNet were proposed for real-time HAR systems. For example, the research work [273] evaluated multiple ConvNet pre-trained models on different real-time human activities. Besides, different hyperparameters were used to assess the ConvNet pre-trained models to identify the best ConvNet models for HAR. The identified ConvNet Model was employed as a feature extractor model to evaluate a large-scale public dataset. Nutteret al. [274] proposed a deep learning method to recognize human activities in real-time from smartphones. To feed the input data into the proposed model, the conventional hand-crafted features of IMU data are extracted and fed into Principal Component Analysis (PCA) to reduce the dimension of the data. This minimizes the size of their proposed model on the learning and inference phases and hence reduced the computation cost to run on an embedded processor and preserves the battery life of the wearable device. To minimize the accuracy rate fluctuations and enhance the accuracy of HAR, Zhu et al. [275] proposed an ensemble of ConvNet models with different layers and filters to recognize activities with less training data. Khan et al. [200] presented a heterogeneous deep ConvNet model to transfer knowledge from a source domain to a target domain using wearable sensor data. Besides, multiple studies presented state-of-the-art ConvNet-based models to enhance HAR systems [275, 207, 276]. However, these studies are limited to wearable sensor data which are often less imbalanced and evaluated on a small number of activities.

1D ConvNet compared to LSTM has obtained competitive results in several applications such as activity recognition, machine translation, and audio generation with much faster learning time. However, 1D ConvNet models are not sensitive to the order of the sensors' sequential data which is crucial for HAR due to processing sensor's sequential temporal data in parallel [12]. Hence 1D ConvNet alone is not an optimal solution instead of LSTM.

2.8.5 Temporal modeling via Hybrid of 1D ConvNet and LSTM

The hybrid model based on stacking 1D ConvNet and LSTM sequentially has been used to improve the performance of the HAR system [19, 225]. 1D ConvNet layers in the hybrid model are often applied when recurrent-based models cannot realistically handle and process long-term dependencies from input sequence data. In such cases, 1D ConvNet in the hybrid model can make the long-term dependencies shorter through down-sampling by extracting higher-level features. Then the extracted features generated by 1D ConvNet could be better

Table 2.6 List of ConvNet models for HAR systems

References	Datasets	Performance (Average Accuracy)
Chen and Xue, 2015 [270]	Self-recorded data	93.8 %
Ranao and Chao, 2016 [65]	UCI-HAR	94.79% on raw data, 95.75% on temporal FFT
Lee et all., 2017 [277]	UCI-HAR	92.7%
Ignatov, 2018 [168]	WISDM	90.42%
Khan et al., 2018 [200]	Self-recorded data	76% on transfer from smartphone to smart-watch, 80% on transfer from smartwatch to smartphone
	Heterogeneity Activity Recognition (HHAR)	78.75% on phone to watch, 76.74 watch to phone
	position aware activity recognition	72.24% on phone to watch, 74.71% on watch to phone
Avilés-Cruz et al., 2019 [201]	UCI-HAR	100%
Wan et al., 2019 [68]	UCI-HAR PAMAP2	92.71% 91.00 %
Tang et al., 2019 [271]	UCI-HAR PAMAP2 UNIMIB-SHAR Opportunity WISDM	96.27 % 91.40% 74.46 % 86.10 97.51 %
Zhu et al., 2019 [275]	Self-recorded data	96.11 %
Cheng et al., 2020 [272]	PAMAP2 UNIMIB-SHAR Opportunity WISDM	94.01 % 77.31% 81.18 % 99.60 %
Cruciani et al., 2020 [273]	UCI-HAR DCASE(Audio-base HAR)	91.98 % 92.30 %
Nutter et al., 2018 [274]	UCI UCF	99 % F1-score 97 % F1-score
Xiao et al., 2020 [278]	AMASS dataset DIP dataset AMASSDI	87.46 % 89.08 % 91.15%
Shojaedini and Beirami, 2020 [276]	WISDM	Improves accuracy to 5% than traditional methods
Almaslukh et al., 2018 [207]	Position aware HAR	84 to 88%

processed by the recurrent-based models [204]. Table 2.7 shows a list of hybrid of ConvNet and LSTM models with their performance that has been proposed for HAR systems from various HAR datasets. Xia et al. [279] proposed a hybrid of ConvNet and LSTM models to accurately recognize human activities from wearable sensor data. The hybrid model is integrated with global average pooling (GAP) [280] to compress the feature map and minimize overfitting by reducing the total number of parameters in the proposed model. GAP is performed instead of conventional fully connected layers i.e., dense layers. Compared to fully connected layers, the GAP layer is faster in processing data due to the absence of parameter optimization in the GAP. Moreover, batch normalization [265] is also appended to the GAP layer to speed up the convergence of the model.

Most of the proposed HAR systems are designed to recognise a few human activities based on wearable sensor data and have certain challenges in the recognition of multiple dynamic activities such as walking and going upstairs. To track and recognize dynamic human activities, Qi et al. [281] proposed an adaptive recognition method to classify 12 human activities with an accuracy of 95.15% and 92.20% for the mobile phones kept in the waist and pocket. Wang et al. [282] proposed a hybrid model of ConvNet and LSTM to extract features using ConvNet layers and capture the dependencies between the actions using LSTM to enhance the HAR identification rate. However, the presented methods in [281, 282] have some limitations in distinguishing similar and dynamic human activities such as walking and going upstairs and downstairs with less inter-class and high intra-class scatter. To address this limitation Lv et al. [283] proposed a margin mechanism to capture discriminative features for improving HAR. The proposed network integrated four modified neural networks and outperforms the unmodified and multiple conventional models on three wearable sensor data. Furthermore, an ensemble deep learning model of ConvNet and LSTM is proposed with traditional shallow learning and statistical features to improve HAR [284]. Besides, bidirectional LSTM is combined with ConvNet to process long sequences of human activities and capture local features that are fed to the fully connected layers with a softmax activation function to classify human activities [285]. The above proposed methods in [285, 284] have two main limitations. Firstly the computation cost of these two methods is very high due to the many parameters to optimize. Secondly, these two methods are only applied to wearable sensor data that are less imbalanced. Ordonez and Roggen. [204] proposed a generic hybrid of ConvNet and LSTM to model human activities. In addition to improving HAR systems, Zhu et al. [286] proposed a semi-supervised learning method to reduce the need for large labelled data for the learning process.

Attention mechanisms are also used in modelling temporal activities to further focus on the important time steps and effectively expose deep semantic correlations from action sequences involving human activities besides the hybrid of ConvNet and LSTM to further improve HAR systems [14, 178, 287, 288].

2.9 Imbalanced Class Problems in HAR

Supervised machine learning algorithms require labelled data to train a model in which each instance in the labelled datasets belongs to a known class[289–292]. Often the instances in labelled datasets are unequally distributed over the classes. This will bring the class imbalance problems which occur when one class has a large number of samples compared to other classes. Typically the minority classes that have significantly fewer examples than the majority classes are the most of interest [293–295]. A comprehensive understanding of the class imbalance problem and the existing techniques for handling it is crucial, s such skewed data exists in various applications, particularly in HAR. During training a model on a dataset with class imbalance problems the model will mostly misclassify the instance of the minority classes and over-classify the majority classes due to the increased prior probability of the majority classes. Machine learning methods that address class imbalance problems can be divided into three groups: data-level techniques, algorithm-level methods, and the hybrid of data-level and algorithm-level approaches [296, 297].

Data-level techniques attempt to mitigate the level of imbalance problems using different sampling data approaches. Algorithm-level methods for addressing imbalanced class problems are often performed by cost-sensitive schema to focus more on the samples of the minority classes. Lastly, a hybrid of the data level and algorithm-level methods[296, 297].

2.9.1 Data Level Solution

Data-level methods include over-sampling and under-sampling techniques to handle imbalance class problems. These two techniques have been used to change the distribution of the training dataset to reduce the impact of the skewed class proportions on the learning process [298]. In their simplest forms, the random over-sampling technique increases the number of samples of imbalanced training sets by duplicating the samples of the minority classes. It has been shown that over-sampling can cause over-fittin that represents a high variance and appears when a model fits very well on the training datasets and is then incapable to generalize to unseen data. On the contrary, the random under-sampling technique removes random

Table 2.7 List of Hybrid ConvNet and LSTM Models for HAR

References	Datasets	Performance
Ordonez and Roggen., 2016[204]	Opportunity	93.0 %
Zhu et al., 2018 [286]	UCI-HAR	97.2 %
He et al., 2018 [287]	UCI-HAR	95.35 %
Wang et al., 2019 [178]	UCI-HAR SWLM	93.41 % 93.83 %
Xia et al., 2020 [279]	UCI-HAR WISDM Opportunity	95.78 % F1-score 95.85 % F1-score 92.63 % F1-score
Qi et al., 2020 [281]	Smartphone-based adaptive HAR dataset	Waist 92.93% Pocket 88.37 %
Wang et al., 2020 [282]	HAPT dataset	95.87 %
Lv et al., [283]	Opportunity UniMiB-SHAR PAMAP2	92.30 % 77.88 % 93.52 %
Su et al., 2019 [285]	HCI-HAR	97.95 %
Mukherjee et al., 2020 [284]	WISDM UniMiB-SHAR MobiAct	97.1 % 98.7 % 95.1 %
Singh et al., 2020 [14]	MHEALTH USC-HAD UTDMHAD1 UTDMHAD2 WHARF WISDM	94.86 % 90.88 % 58.02 % 89.84 % 82.39 % 90.41 %
Gao et al., 2021[288]	WISDM UNiMiB-SHAR PAMAP2 Opportunity SWLA	98.85 % 79.03 % 93.16 % 82.75 % 94.86 %

samples of the majority classes in the training sets, hence, the under-sampling approach decreases the total amount of information that has to be used to train a model. Different sampling methods have been developed to handle the imbalance class problems in an attempt to avoid over-fitting and preserve valuable information.

Zhang and Mani [299] presented k-NN to select several samples from the majority classes based on their distance from minority classes. Kubat and Matwin [300] proposed a technique to remove noisy and redundant examples of the datasets only for the majority of classes. Furthermore, Barandela et al. [301] proposed a method to remove the misclassified example on the class boundaries from the training dataset

Many over-sampling techniques have also been designed to avoid over-fitting, strengthen class boundaries, and enhance discrimination. Chawla et al. [302] proposed a new method for over-samples which is called Synthetic Minority Over-sampling Technique (SMOTE). This technique generates new samples among the minority samples and their nearest minority neighbours. However, SMOTE in generating new instances fails to consider neighbouring instances of different classes which can append further noise and expand the overlapping of the classes. To overcome this problem, different techniques based on SMOTE have been developed, e.g. Borderline-SMOTE [303], BLL-SMOTE [304] and Safe-Level-SMOTE [305] to consider the majority class neighbours. For example, border limited link SMOTE (BLL-SMOTE) is proposed, to avoid misgenerated new samples [304]. BLL-SMOTE focuses on the distances between the newly generated samples with their k-nearest neighbours and the nearest sample in the dataset. To mitigate the distance calculations of the BLL-SMOTE and improve the oversampling process, we propose improved SMOTE (iSMOTE) that computes k-nearest neighbours of each generated instance to make sure the new instances are accurately annotated. Each new instance of the minority classes with its k-nearest neighbours must have the same class. For example, generated new instances of *Shower* activity must have k *Shower* activity as nearest neighbors. The details of our proposed iSMOTE will be presented in Chapter 4.

2.9.2 Algorithm Level Solution

The algorithm-level methods attempt to reduce the level of imbalanced class problems in the training process mainly through cost-sensitive learning and new loss functions. Wang et al. [306] proposed new loss functions that enable samples of the minority classes to contribute more to the loss function. Lin et al. [307] proposed the focal loss function to address the imbalanced class problem between background and foreground classes during training in one

stage object detection scenario. The focal loss is designed to down-weight well-classified examples and focuses on hard-classified examples. The loss value of hard-classified examples is much higher compared to the loss values of the well-classified examples by a classifier using the focal loss function. Since the focal loss focuses more on a sparse set of hard-classified samples, hence the focal loss is used in our proposed network to improve the learning of minority classes in HAR systems. The focal loss function is shown in Equation 2.10.

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{2.10}$$

Wang et al. [308], Khan et al. [309] and Zhang et al. [310] proposed DNN methods based on cost-sensitive learning by providing more weights to the minority classes to address the imbalanced class problems. In addition to that, the approaches presented by Khan et al. and Zhang et al. have an adaptable cost function that learns the cost matrices during the training process of the model.

To address imbalanced class problems without modifying the training samples and without assigning different weights using cost-sensitive learning as well as improving the performance of HAR systems, we propose a joint learning method of two different temporal models, i.e., LSTM and 1D ConvNet. The proposed joint learning of temporal models learns from the same form of input features for HAR. Different from ensemble learning which often combines the outputs of many learners while using a specific aggregation function to handle imbalanced data [311], the proposed method combines the learning processes of two temporal models in a single joint training mechanism to improve the accuracy on minority classes in addition to maintain the accuracy on majority classes. Therefore, joining the learning processes of two different temporal models in the proposed method is expected to obtain a better combined model compared to simply aggregating the outputs of multiple learners. It is also expected to obtain more accurate and reliable estimates or decisions than single models. The two temporal learners of the jointly proposed methods can exploit different features from the input data to render a strong mutual complementary model. Complementarity in joint learning based on different models can greatly boost the performance as compared to simply combining the same learners (e.g., LSTM with LSTM in this work) in a joint learning model [312]. This is because each base learner brings different features into the joint learner to enrich the joint learning process and each learner improves the earlier layers of the other learner, but at the same time, the weaknesses of each individual learner are avoided. The proposed method jointly trains the two base learner i.e., LSTM and 1D ConvNet, and combines the based learners by a fully connected layer, which is followed by the output

layer. The joint optimization that leads to increasing the functionality of the proposed joint temporal model to gain more insight into the input data and features reduces the recognition error rate. Thereby, the proposed model increases the performance of activity recognition, particularly for minority classes. We also adopt the incremental multiple fuzzy temporal windows approach in order to compute informative features to enhance recognition accuracy, particularly for minority classes as well. The details of our proposed method are presented in Chapter 4

2.10 HAR Using Transfer learning

A large number of studies are conducted for HAR and most of them are reporting the state-of-the-art performance and addressing different problems of HAR systems. However, there are still some challenges that require further investigation to be appropriately addressed. One of the main challenges is that HAR systems require a considerable amount of labelled data which is not always available and expensive to acquire.

Several transfer learning approaches are presented to reduce the need for labelled data with maintaining reasonable accuracy. Transfer learning is defined as the capability to expand what has been learned in the training process in one context to new and related contexts. Transfer learning in machine learning refers to various strategies [313] which include multi-domain learning [314–316], self-supervised learning [317–319], domain adaptation [320], multi-task learning [321, 322], sharing of knowledge and representations [323].

In this thesis project, transfer learning is performed through two different techniques: cross-domain learning and self-supervised learning using shared representation to make HAR systems more robust, and adaptable and effectively reduce the need for annotated data for HAR systems. Cross-domain learning is about transferring knowledge from a domain to other related domains to reduce the need for labelled data, reduce the training time and improve the accuracy of the target task [313]. Besides, a self-supervised learning network is proposed to learn a good representation of human activities from unlabeled data, enhance accuracy and substantially minimize the labelled data required for a downstream task, i.e. activity classification.

2.10.1 Cross-domain Learning

In deep learning, a domain refers to a dataset where its examples are drawn from the same distribution [324]. Multiple datasets with different data distributions can be used to target

the same problem by fusing multi-domain features. Multi-domain learning (MDL) aims to perform a task (e.g., activity recognition) across different but related domains simultaneously. MDL has become an interesting problem since the success of deep learning models has been based on large-scale training data. MDL as a transfer learning technique addresses the problem of learning a task from different but related domains that share some commonalities in their input-output mapping functions [315]. MDL learning optimizes multiple objectives jointly and simultaneously to improve the performance of multiple domains by exploiting commonalities and differences across domains using a shared representation [325]. Moreover, multi-domain learning reduces the sampling burden required to improve generalization by learning multiple similar domains as opposed to learning each domain in isolation [315, 316]. Similar domains using a shared representation transfer knowledge to each other during the learning process to mitigate the negative effect of data scarcity [326]. In contrast to an SDL that only uses domain-specific data to independently learn a model for each domain, MDL leverages data of all domains and shares knowledge between domains causing better prediction performance and model generalization. Figure 2.9 shows the architecture of SDL and MDL. There is limited research for activity recognition using MDL particularly based on sensor data. MDL proposed between activities of two domains using Web search [314]. The limitation of these works yet firstly relied on classical statistical features engineering. Second Web search is used to obtain relevant Web pages for the activities, and then techniques of information retrieval are employed to further process the extracted Web pages. MDL methods are proposed for human activity recognition by [327, 328], however, the methods are only validated using wearable sensors data.

2.10.2 Self-supervised Learning

Deep learning approaches have been broadly used and led to significant improvements in different applications of healthcare, ubiquitous computing, and pervasive intelligence. Since deep learning techniques can eliminate the need to manually extract features from input data [70], deep learning methods have become dominant in building intelligent systems [36]. Recurrent and convolutional neural networks have demonstrated satisfying performance in various recognition systems from sequential data such as human activity recognition [204, 138]. The majority of the deep learning systems for human activity recognition demonstrate state-of-the-art performances [36]. LSTM as a recurrent method is commonly used in addressing temporal recognition problems due to taking the ordering of the time steps into account which is crucial for accurate temporal recognition tasks [36]. Besides, Con-

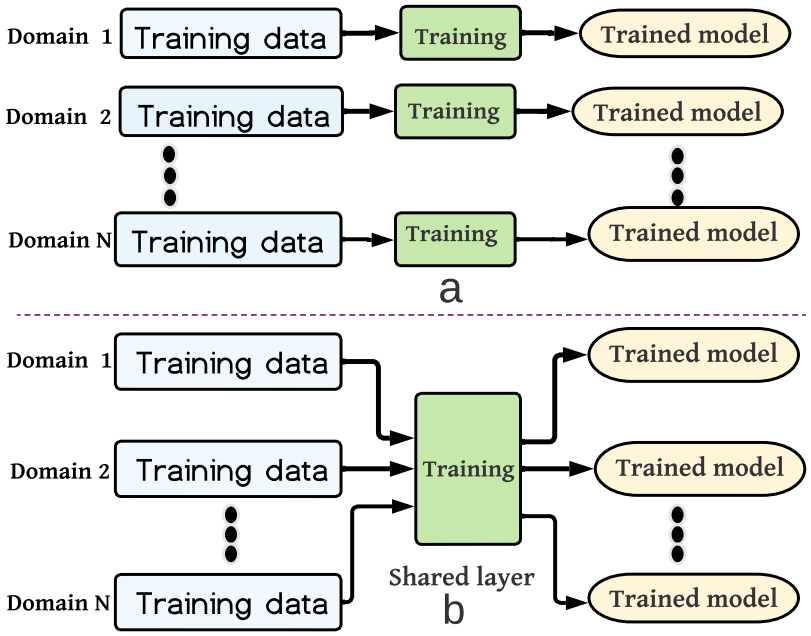


Fig. 2.9 Comparison of two systematic Architecture : (a) Single Domain learning (SDL) and (b) Multi-domain learning (MDL), denotes model learning in either SDL or MDL architecture

vNet methods have become increasingly prevalent in temporal modelling due to processing temporal data in parallel, weight sharing and also is translation invariance [329]. These methods are mostly employed in supervised manners and require large labelled samples which are infeasible in many domains and require knowledge experts to manually label samples. Therefore, these major inherent limitations of the supervised learning methods emphasize the significance of using unsupervised learning to leverage abundant unannotated data for representation learning [330] which can be obtained in a real-world setting.

Self-supervised learning as unsupervised learning which we study in this paper can exploit the inherent structure of the feature map in the human activity datasets to obtain a supervisory signal from unlabeled data for the downstream task, e.g. HAR. Self-supervised learning has drawn considerable attention in the vision domain and numerous state-of-the-art models have been proposed to learn useful visual features from static images and videos [317]. Moreover, self-supervised learning methods have been actively studied in language modelling due to their excellent data efficiency and generalization ability [331] and graph learning [332].

Even though self-supervision has been well explored for vision and language modelling, it remains challenging for HAR systems from sensor data of smart home environments and wearable devices. Few studies using self-supervision are focused on HAR from sensor

data [329, 333, 334]. Haresamudram, et al.[333] proposed a masked reconstruction model as a variant of the BERT model [331] to randomly replace sensory signals to zero for certain time steps. Their model was trained using the mean squared error loss function to reconstruct the masked sensor signals. Saeed, et al. [329] proposed a multi-tasking method for self-supervised sensor representation learning. Their multi-tasking method was trained on temporal data to identify possible transformations (augmentation) that may have been used on the raw sensor data. Haresamudram, et al. [335] proposed a self-supervised method based on contrastive predictive coding (CPC). Their proposed method compared to CPC network [336] that is the encoder only trains on short time windows (1s) with a 50% overlap between subsequent windows. A lightweight model of BERT is proposed for HAR based on inertial measurement unit data to learn representations of human activities from unlabeled data [337]. SelfHAR [338] is another self-supervision model proposed for HAR systems from sensor data of wearable devices. These methods have not addressed handling imbalanced datasets in representation learning for HAR and also only applied to the wearable sensor data which are less imbalanced compared to smart home environment data.

Chapter 3

Sequential Learning for Human Activity Recognition

3.1 Introduction

Human activities are highly diverse due to different sensor readings and even the same subject tends to perform an activity in different ways. Also, the intrinsic characteristic of categories denoting daily human activities is inherently imbalanced, and hence building a robust machine learning model for HAR is challenging. Moreover, occasionally generated data by sensors could be noisy which adds extra challenges and ambiguity to the interpretation of human activities [38, 127, 339, 340].

Deep learning models are widely employed in different applications of computer vision, audio recognition, and natural language processing. Furthermore, deep learning approaches have improved HAR systems based on sensors generated data and show promising results. Since mostly HAR problems are formed as a sequential learning [341], RNN as a type of sequential learning and its variations particularly LSTM have demonstrated satisfying and state-of-the-art performance [38]. LSTM integrated models are commonly used and increase the performance of HAR systems, but LSTM requires a large amount of memory and high computational capacity for its memory cells and gating mechanism in learning to process temporal sequential contextual information [19]. Further, LSTM models process times steps of sensor temporal data sequentially because processing any timestep requires the outcomes of the previous time steps [342, 210]. ConvNet is employed to extract the temporal contextual information for HAR systems from sensor data [268]. Even though training one-dimensional (1D) ConvNet models is remarkably faster than LSTM due to the nonexistence of recurrent settings, LSTM models show better performance than 1D ConvNet for HAR

systems. Furthermore, 1D ConvNet models are not sensitive to the order of the sensors' sequential data which is crucial for HAR due to processing sensors' sequential temporal data in parallel. To address this problem and efficiently process sensors' data for HAR, we have developed two ConvNet models integrated with different attention mechanisms. The models are described below.

1. Dilated causal convolution with the self-attention mechanism (ConvNet+Self) is proposed to entirely forgoes recurrent settings and improve the performance of HAR. Dilated convolutions within the proposed method are used to maximize the receptive field by orders of magnitude and aggregate multi-scale contextual information without considerably increasing computational cost. The self-attention technique is used to focus more on important time steps of the feature maps by computing similarity scores for all time steps [342].

   Even though the proposed dilated causal convolution with the self-attention outperformed the state-of-the-art methods as shown in the results, computational and memory requirements of the self-attention technique are quadratic with the length of the input sensor sequential data which leads to slow learning and occupy more memory. To address this limitation, this thesis proposes a new method for HAR to further accelerate the training time and enhance the performance of HAR. The new method introduces a lightweight attention mechanism which is called performers-attention and integrates with causal ConvNet and supervised contrastive learning. The new method is further delineated in the next section.

2. Causal supervised contrastive ConvNet based on performers-attention (ConvNet+Performers) is also developed. This proposed network improves the results of the HAR systems in sensors generated data. In addition, the proposed method also accelerates the learning process compared to the existing methods. Causal convolution [210, 343] is adopted to avoid violating the ordering time steps of the input datasets, which is crucial in HAR systems. Performers-attention [344] which scales linearly with the input sequence length is proposed to reduce the computation and memory cost compared to the self-attention mechanism for HAR systems. Moreover, supervised contrastive learning is adopted to learn a good representation from the input sensors data that supports classifiers to gain useful information [330, 345]. Due to integrating supervised contrastive learning, the proposed network has two learning stages. The network learns a good representation of human activities in the first stage to learn a more accurate classifier in the second stage. Further, in the first stage, the supervised contrastive loss function

is applied to learn the representation of human activities which is further propagated through a projection network. In the second stage, a linear classifier is trained on top of the frozen representations while the projection network is discarded. The two stages of learning prepare a discriminative representation that renders a more accurate classifier [345].

Moreover, due to the diversity of human activity recognition which leads to generating long-tailed datasets with skewed class distributions, often classifiers tend to be more biased towards majority classes and misclassify minority classes. To address this limitation, the focal loss function [307] based upon the effective number of samples [346] is proposed by assigning higher weights to hard-classified examples to sufficiently learn minority classes. The focal loss function is conducted in the second stage to learn a linear classifier for HAR.

These methods are proposed to entirely forgoes recurrent settings and to boost the performance of HAR systems since parallelization is inhibited in recurrent networks due to sequential operation and computation that lead to slow training, occupying more memory and hard convergence. The developed networks are evaluated on eight benchmark HAR datasets and compared with the existing state-of-the-art methods. Both proposed networks ConvNet+self and ConvNet+Performers that are developed to accelerate learning time and improve the performance of HAR systems are delineated in the following sections.

3.2 Dilated Causal ConvNet with Self-attention (ConvNet+Self)

In this section, we describe the proposed network ConvNet+Self which is developed using dilated causal convolution based on the multi-head self-attention mechanism for HAR from sensors data. We aim to design and propose a more efficient convolutional network model better than recurrent-based architecture models in terms of recognition score and training time. The distinguishing characteristics of our proposed method are: 1) the proposed model stops information leakage from future to past using causal convolution; 2) the proposed model can handle temporal sequential data of any length and map it to a series output of the same length; 3) the model can simultaneously focus on different important time steps of the sequence input using the multi-head self-attention mechanism. The proposed method consists of two layers of 1D dilated causal convolution followed by the multi-head self-attention mechanism to learn feature maps of human activities from the sensors data. Then a fully connected layer followed by a softmax layer is applied on the feature maps to recognise and

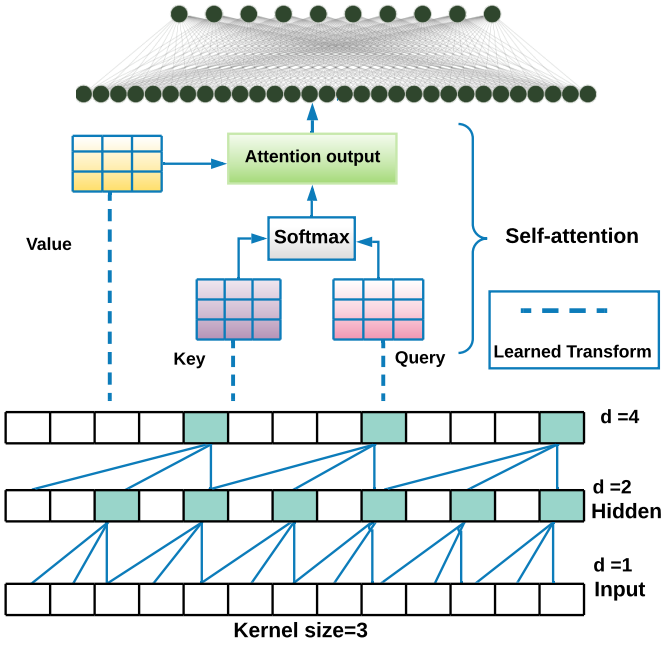


Fig. 3.1 Dilated causal convolution and self-attention model

distinguish human activities. Figure 3.1 provides more information about how the layers of the proposed method are stacked. The details of the proposed model are described in the following subsections.

3.2.1 Dilated Causal Convolutions

Causal convolutions used in the proposed method to control the model and predict output at time t based on only the convolutions of the sequence inputs from time t and earlier in the previous layers [210]. Causal convolutions also preserve the ordering of sequential input patterns. However, causal convolutions require very large filters or many hidden layers to expand the receptive field [343]. To maximize the receptive field and aggregate multi-scale contextual information without considerably increasing computational cost, dilated convolutions are integrated into the proposed method. Dilated convolutions enable the model to increase the receptive field exponentially using a few layers and keeping the computational efficiency [347]. The dilated causal convolution DCC for one dimensional input sequence $x \in R^n$ with a filter $f : \{0, 1..., k\text{-}1\} \rightarrow R$ on element s of the sequence is defined as:

$$DCC(x \star_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \tag{3.1}$$

where d is the dilation factor, k is the filter size, and $s - d \cdot i$ shows the past direction. The dilation factor d is exponentially increased when the depth of the model is increased i.e., $d = 2^l$ at layer l of the model. Formally, we increase the dilation factors d exponentially by a factor of 2 in each layer $l = 1, ...L$ where L is the number of layers of the dilated causal convolutions in the proposed model. Equation 3.2 shows the dilation factor in this study.

$$d \in [2^0, 2^1, 2^2, ..., 2^{L-1}] \tag{3.2}$$

In addition, the dilation convolution renders the standard convolution when d=1. Figure 3.1 shows the dilation causal convolutions in the proposed model for dilations 1, 2, and 4. Dilated convolution with different dilation factors can be integrated with a filter at different ranges. The filters convolve input values over an area larger than its length using dilated convolutions by skipping input values with a certain step which is the dilation factor. Hence, dilation convolution is equivalent to a standard convolution with one dilating, but importantly more efficient. Dilation convolution effectively enables the model to aggregate multi-scale contextual information with fewer layers and the same receptive field compared to a standard convolution [347]. Therefore, the number of learnable parameters is reduced by using stacked dilated causal convolutions that lead to yielding more efficient training and lightweight model.

3.2.2 Self-Attention Network

The self-attention mechanism is a robust technique to compute correlation and the weighted combination between all the time steps in the input sequence [342]. After applying dilated causal convolution to render aggregated multi-scale contextual information, multi-headed self-attention is used to enable the model to focus on important and relevant time steps more than the insignificant time steps from the sequential feature maps during recognition. Hence, the attention mechanism aims to learn the most important time steps from the sequence feature maps that aid in determining more accurate recognition. Moreover, self-attention identifies relative weights for each time step in the sequence feature map by considering its similarity to all the other time steps within the sequence. Then, the representation of each time step with relevant and important information from other time steps is transformed by the relative weights according to their importance. The self-attention mechanism has three learned matrices: queries $Q \in R^{N \times D_k}$, keys $K \in R^{M \times D_k}$, and values $V \in R^{M \times D_v}$, where N and M are the lengths of the queries and keys (or values), D_k and D_v are the dimensions of keys (or queries) and values [342]. To obtain attention scores, dot product attention is applied between each query as considered to the transformed matrix of a specific time step

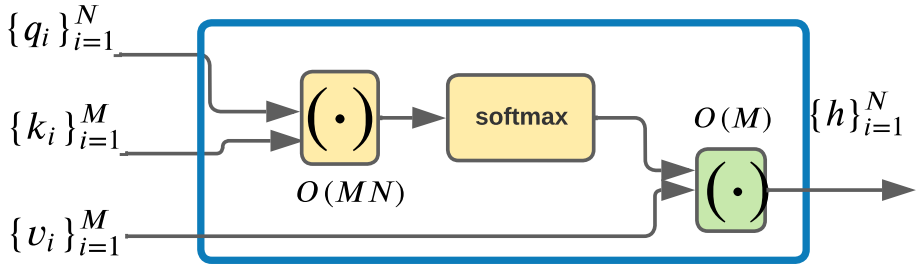


Fig. 3.2 Self-attention

and the key matrix of every other time step. The dot-products of Q and K are scaled by $\sqrt{D_k}$ to mitigate the softmax gradient vanishing problem. Often the softmax $(\frac{Q \cdot K^T}{\sqrt{D_k}})$ is called the attention matrix. Then the softmax function is applied on the scaled dot product value of the queries and keys to generate the attention scores. Lastly, the attention scores are used to produce a weighted representation of the value matrix for each of the time steps in the sequence as shown in Figure 3.2. Equation 3.3 shows the multi-head self-attention is entirely implemented as a matrix multiplication operation.

$$f_{sa}^{(hj)}(Q,K,V) = softmax(\frac{Q \cdot K^T}{\sqrt{D_k}})V \tag{3.3}$$

The model computes the attention numerous times in parallel (multi-head) to capture distinct correlation information of the input sequence. Hence, hj in Equation 3.3 shows the output from attention head j and sa refers to self-attention. Distinct parameters are used in Equation 3.3 for computing each the key, query, and value of the n attention heads. The outputs from the distinct multi-head attention are concatenated and transformed to the dimension of the input sequence using the learned parameter W_o as defined in Equation 3.4. The outputs of the multi-head self-attention (M_{ha}) are fed into fully-connected layers, i.e., a dense layer with a ReLU activation function and a soft-max layer.

$$M_{ha} = W_o \cdot concat(f_{sa}^{(h1)}, ..., f_{sa}^{(hn-1)}, f_{sa}^{(hn)}) \tag{3.4}$$

The proposed method based on dilated causal convolution foregoes recurrent architectures to accelerate the training and inference time. Causal convolution maintains the ordering of data which is crucial for HAR systems. Dilated convolution increases the receptive field and produces feature maps with multi-scale receptive fields using the different dilated rates in the convolution layers. Dilated convolution preserves the resolution of the data since the layers are dilated instead of pooling. The multi-head self-attention mechanism is employed

in the proposed method to capture informative time steps in the feature map to improve HAR systems. Dilated causal convolution with a self-attention mechanism is used to make the proposed method computationally efficient and improve the result scores.

The operation of the self-attention mechanism scales quadratically with the input sequence length which can increase training time because it appends more weight parameters to the model. To address this limitation, in the next section, we proposed a new method to recognise human activities to further accelerate the training time and enhance the performance of HAR by introducing a lightweight attention mechanism.

3.3 Causal Supervised Contrastive ConvNet-based Performers Attention (ConvNet+Performers)

The proposed network ConvNet+Performers is developed using causal 1D ConvNet with the performers-attention based on supervised contrastive learning. The proposed ConvNet+performers network has four main components which are: causal convolution, performers-attention, supervised contrastive learning for two stages of learning (representation learning and classifier learning), and focal loss. The causal convolutions component in the proposed network is used to avoid information flow from future to the past by processing results at time t based on solely the convolutions of the time steps of the temporal data from time t and earlier in the previous layer. Therefore, predicting time steps at time t cannot rely on any of the future time steps from the sensor sequential data. This helps the proposed network to maintain the ordering of the temporal data [343] which is significant for HAR systems [19]. The performers-attention is used in the proposed network to focus more on the important time steps to improve the recognition process. Moreover, the details of the performers-attention are provided in Sections 3.3.2. Supervised contrastive learning is used to prepare a discriminative representation and further reduce the classification error compared with several existing methods for HAR. Further, the focal loss function is used to address imbalanced activities problems and improve the less presented human activities.

The proposed network consists of two layers of 1D causal convolution followed by a fully connected layer and the performers-attention mechanism to build an encoder. Then the encoder is projected to render a representation of human activities in the first stage of learning. Next, in the second stage of learning the trained encoder is frozen and followed by a fully connected layer and a softmax layer to recognise human activities. Figures 3.3 and 3.4 presents the structure of the proposed network and the two stages of learning in which the

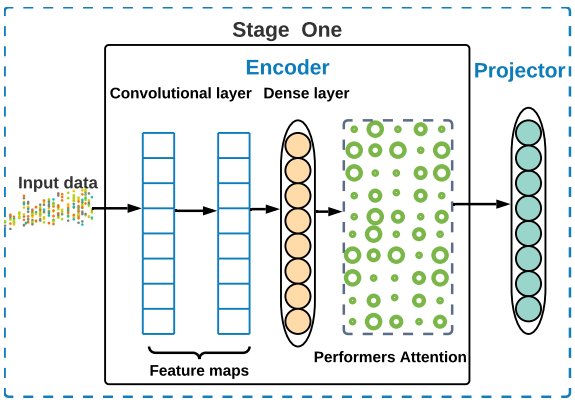


Fig. 3.3 Representation learning

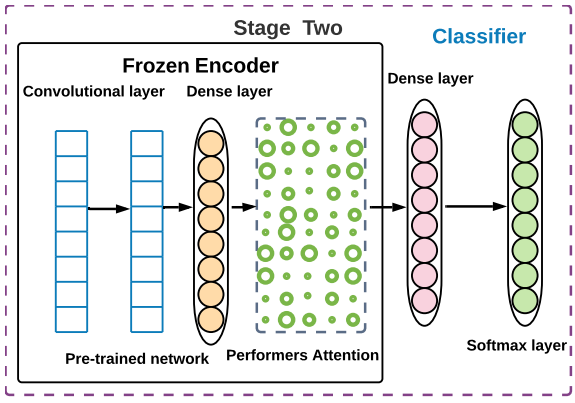


Fig. 3.4 Classifier learning

representation learning uses supervised contrastive loss function and the classifier learning uses the focal loss function. More details about supervised contrastive learning and both learning stages are provided in Section 3.3.3.

3.3.1 Focal Loss

The focal loss [307] is introduced to address the imbalanced class problem between background and foreground classes during training in one stage object detection scenario. The focal loss is designed to down-weight well-classified examples and focuses on hard-classified examples. The loss value of hard-classified examples is much higher compared to the loss values of the well-classified examples by a classifier using the focal loss function. Since the focal loss focuses more on a sparse set of hard-classified samples, hence the focal loss is used in our proposed network to improve the learning of minority classes in HAR systems. The focal loss function is shown in Equation 3.8.

The focal loss function is an updated version of the cross-entropy loss function to solve the imbalanced class problem. The cross-entropy loss function is shown on Equation 3.5.

$$CE(p,y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1-p) & \text{otherwise} \end{cases} \tag{3.5}$$

where $y \in \{\pm 1\}$ is the actual class ground-truth and $p \in [0,1]$ denotes the estimated probability of the model for the class with label $y = 1$. To make convenient notation, the above equation can be written as:

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{otherwise} \end{cases} \tag{3.6}$$

The cross-entropy loss function could also be illustrated as:

$$CE_(p,y) = CE_(p_t) = -\log(p_t) \tag{3.7}$$

The focal loss as shown in Equation 3.8 is produced by appending a modulating factor $(1-p_t)^\gamma$ with a tunable focusing parameter $\gamma \geq 0$.

$$FL(p_t) = -\alpha_t(1-p_t)^\gamma \log(p_t) \tag{3.8}$$

The focal loss function focuses on a sparse set of hard-to-classify samples to prevent the majority of easy-classified samples to overwhelm the training. Precisely, α_t in the focal loss is responsible to handle the imbalanced class problem by providing high weights to the minority classes and small weights to the dominating classes. The class weights for α is calculated by effective number of samples [346] E_{n_c} based on Equation 3.9.

$$E_{n_c} = \frac{1-\beta}{1-\beta^{n_c}} \tag{3.9}$$

where $\beta \in [0,1)$ is a hyperparameter, n_c is the number of samples from class c. Besides, the γ is automatically down-weighting the contribution of well-classified samples in training to pay more attention on the hard-classified examples. The γ is a float scalar modulating loss from hard-classified and well-classified samples.

3.3.2 Generalized Kernelizable Attention

The complexity of the self attention mechanism with the length of the input temporal sequence scales quadratically which increases model learning time and requires more memory. This is the limitation of the self-attention mechanism. To address this limitation, we adopt performers-attention [344] as an efficient attention mechanism whose complexity scales linearly with the size of an input sequence L. The performers uses a Fast Attention Via Positive Orthogonal Random Features (FAVOR+) algorithm and substitutes Transformer self-attention by generalized kernelizable attention. The FAVOR+ algorithm is used to estimate the regular softmax attention by random feature map decompositions. Hence the core idea of the performers is to decompose the attention matrix into a matrix product. This algorithm

leverages positive orthogonal random features to approximate softmax attention kernels with provable accuracy and $O(N)$ for both computational and space complexity [344]. Previous attention mechanisms such as sparsity and low-rankness relied on structural assumptions for the attention matrix without approximating the original softmax function. Generalized kernelizable attention can make the model process longer input sequences and train faster compared to previous attention mechanisms. The aim of using generalized kernelizable attention and FAVOR+ is to approximate the softmax and choose the order of computation of the matrices of Equation 3.3.

Hence, the goal is to approximate $\exp(QK^T)$ by $Q'K'^T$ ($Q'K'^T \approx \exp(QK^T)$), decompose the softmax distribution and form a lower-dimensional attention matrix (A). FAVOR+ is used for attention blocks using matrices $A \in \mathbb{R}^{L\times L}$ of the from $A(i,j) = K(q_i^T, k_j^T)$ where q_i, k_j are the ith, jth query, key row-vector in QK, and Kernel K: $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$ is defined for randomized mapping: ϕ $\mathbb{R}^d \rightarrow \mathbb{R}_+^r$ (for some $r > 0$) as:

$$K(x,y) = \mathbb{E}[\phi(x)^T \phi(y)]. \tag{3.10}$$

The kernel function: ϕ: $\mathbb{R}^d \rightarrow \mathbb{R}_+^r$ that maps $Q \rightarrow Q'$, and $K \rightarrow K'$ is of the form

$$\phi(x) = \frac{h(x)}{\sqrt{r}} \left(f_1(w_1^T x), .., f_1(w_m^T x), .., f_l(w_1^T x), .., f_l(w_m^T z)\right) \tag{3.11}$$

where $h(x)$ and $f_1 \ldots f_l$ are deterministic functions, $w_1, \ldots, w_m \overset{iid}{\sim} \mathcal{D}$ are random vectors from a distribution $\mathcal{D} \in \mathcal{P}(\mathbb{R})^d$. Also, $r = l \cdot m$, with the best performance when $m = r$, $l = 1$, $w_1, \ldots, w_r$ are orthogonal random feature maps and the ReLU function is used for f_1. In Equation 3.12, the attention that is produced by generalized kernelizable attention and FAVOR+ for the performers to reduce the computation and memory cost.

$$A_{\text{Perf}} = \widehat{D}^{-1} Q'K'^T \ . \tag{3.12}$$

Consequently, the generalized kernelizable attention Z' reads the content of Equation 3.13.

$$Z'(Q,K,V) = \widehat{D}^{-1} \left(Q' \left((K')^T V\right)\right) \tag{3.13}$$

where the normalization $\widehat{D}^{-1} = \text{diag}\left(Q'((K')^T 1_L)\right)$, and the dimensions of the involved matrices are the following: $Q' \in \mathbb{R}^{L\times r}$, $K'^T \in \mathbb{R}^{r\times L}$, $V \in \mathbb{R}^{L\times d}$, $Z' \in \mathbb{R}^{L\times d}$. In Equation 3.13, brackets indicate the order of computations and Z' stands for the approximate attention. Regarding the complexity of the generalized kernelizable attention, the matrices product $K'^T V = W$ and $Q'W$ can be computed in linear time in L. Moreover, the diagonal matrix $\widehat{D}^{-1}$ can also be computed in linear time in L. Based on the Equation 3.13 and as shown in Figure 3.5, the space complexity of Z' is $O(Lr + Ld + rd)$ and time complexity is $O(Lrd)$ as opposed to the regular attention with space complexity $O(L^2 + Ld)$ and time complexity $O(L^2 d)$ of . Hence the complexity of Z' scales linearly with L.

3.3.3 Supervised Contrastive Learning

In this study, supervised contrastive learning (SCL) is used to build a model for HAR that outperforms the state-of-the-art HAR methods. The proposed method based on SCL consists of two stages of learning. In the first stage, two components are trained which are encoder and projection networks. The first stage learns representations used in the second learning stage to build a robust and accurate classifier for HAR systems. The details of the first stage are as follows:

1. Encoder network $E(\cdot)$ maps temporal input sequential data x to a representation vector $r = E(x) \in R^{D_E}$ where D_E = 512. The encoder network specifically consists of two 1D ConvNet layers followed by a fully connected layer. The performers-attention is then applied to effectively extract deep semantic correlations from action sequences involving human activities. After each layer, normalization and dropout regularization are applied to make the learning process faster and prevent the encoder from overfitting. 1D ConvNet-based networks have been proposed as fast and accurate models for HAR systems [19]. This is due to the ability of 1D ConvNet in extracting mostly correlated features by considering local dependency from temporal sequential input data.

2. Projection network $Proj(\cdot)$ maps the representation vector r to a projected vector $z = Proj(r) \in R^{D_E}$ where D_E = 512. The projector network is only a single fully connected layer appended to the encoder. The Encoder and projection networks are trained using contrastive loss function to make embeddings of similar classes are close together and dissimilar classes are far apart. The projection is discarded at the end of the contrastive training. Equation 3.14 shows the supervised contrastive loss function which is used in the first stage to learn the encoder.

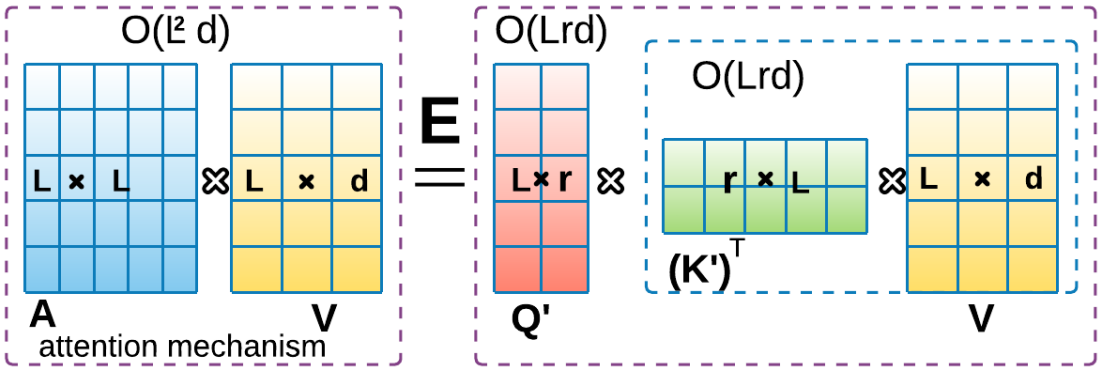


Fig. 3.5 Approximation of the regular attention mechanism AV via random feature maps. Dashed blocks show the order of computation with corresponding time complexities [344]

$$\mathcal{L} = \frac{-1}{2N_{\tilde{y}_i} - 1} \sum_{j=1}^{2N} \mathbb{1}_{i \neq j} \cdot \mathbb{1}_{\tilde{y}_i = \tilde{y}_j} \cdot \log \frac{\exp(z_i \cdot z_j / \mathcal{T})}{\sum_{k=1}^{2N} \mathbb{1}_{i \neq k} \cdot \exp(z_i \cdot z_k / \mathcal{T})} \tag{3.14}$$

where

- N is the number of random samples in a mini-batch;
- N_y is the total number of samples in the mini-batch with the same label y;
- $z_i = Proj(E(x_i))$ and $z_j = Proj(E(x_j))$ are the projected vectors of the samples belonging to the same class;
- while $z_k = Proj(E(x_k))$ is the projected vector of a different class;
- $\mathcal{T}$ is a positive scalar temperature parameter;
- $\mathbb{1}_{i \neq j}$ avoids inner product of the same vector;
- $\mathbb{1}_{\tilde{y}_i = \tilde{y}_j}$ ensures that the z_i and z_j are the projected vectors of the same class;
- $\mathbb{1}_{i \neq k}$ is used to ensure that the z_k does not belong to the class of z_i and z_j.

In the second stage, a classifier with a fully connected layer followed by a softmax layer is trained using the encoder network. However, the encoder network of the first stage is frozen and the projector network is discarded. The learned representation from the encoder network without the projector network is used to learn the classifier. In the second stage, the network uses the focal loss function to predict human activities. The proposed network

causal ConvNet based on supervised contrastive learning and Performers-attention forges recurrent settings to further accelerate the learning phase and improve recognition score for HAR systems. Causal convolution ensures that the model does not violate the ordering of the time steps of the temporal sensors data. The performers-attention supports the proposed network to pay extra attention to the discriminative features to accurately recognize human activities. Supervised contrastive learning is used to build the proposed network in two stages of learning, where the first stage is used to learn a good data representation for learning the classifier in the second stage. Two stages of learning are used to learn a better representation with more discriminative features that support the classifier to better distinguish human activities compared to a normal one stage learning. The focal loss function according to the effective number of examples is used to prevent skewed learning toward majority activities and improve the recognition scores of the minority activities.

3.4 Experimental Setup

In the section, we will show the details of the experiments of the proposed methods based on eight collected sensor datasets.

3.4.1 Datasets and Preprocessing

Eight public datasets are used to evaluate the above proposed networks for HAR systems. Five smart home sensors collected datasets and three wearable sensor collected datasets which described in Section 2.4.1 are used in the evaluation process. The five smart home datasets are collected from Ordonez home A and B [111] and Kasteren home A, B and C datasets. Table 3.1 shows details of these five smart home datasets with respect to the residents, sensors, and the number of activities. Besides, Tables 3.2 and 3.3 show the frequency of the human activities from Ordonez homes A and B as well as Kasteren homes A, B and C datasets after preprocessing and segmentation using FTWs which is delineated in Section 2.5.2. The wearable sensor datasets: UCI-HAR dataset, Roomset1 and Roomset2 datasets are also used in the evaluation process. The distribution of these wearable datasets are shown in Tables 3.4 and 3.5. The proposed networks are evaluated on the datasets with multiple human activities ranging from 4 to 13 from collected sensors data.

Table 3.1 Smart home environment datasets

	Ordonez smart home		Kasteren smart home		
	Home A	Home B	Home A	Home B	Home C
Setting	Home	Home	Apartment	Apartment	House
Gender	-	-	Male	Male	Male
Activities	10	11	10	13	16
Age	-	-	26	28	57
Rooms	4	5	3	2	6
Sensors	12	12	14	23	21
Duration	14 days	21 days	25 days	14 days	19 days

Table 3.2 Frequency of human activities in the Ordonez datasets

Activity	Home A	Home B
Leaving	1664	5268
Breakfast	120	309
Toileting	138	167
Spare Time/ TV	8555	8984
Dinner	-	120
Sleeping	7866	10763
Snack	6	408
Grooming	98	427
Showering	96	75
Idle	1598	3553
Lunch	315	395
Total	20,456	30,427

3.4.2 Performance Evaluation Metrics

The performance of HAR systems is evaluated using different metrics to determine how well a model is performing on a given dataset. Different types of performance metrics have been used based on the purpose or the type of problem in HAR systems. Two common metrics are mostly used in HAR systems accuracy and F1-score. Accuracy is the ratio of correctly classified instances over all classified instances. The accuracy provided the overall performance of the model and could be used to compare with other models [199, 200, 203, 209, 348]. Accuracy is often used to evaluate the performance of classifiers on balanced datasets. However, accuracy in the presence of imbalanced classes cannot be an appropriate measure for classification because less presented classes have very little impact on accuracy as compared to the prevalent classes hence it can be misleading. Hence, in this thesis F1-score is employed to measure and evaluate all the proposed temporal models since the F1-score is the weighted average of recall and precision that can provide more insight

Table 3.3 Frequency of activities in the Kasteren datasets

Home C Activities		Home B Activities		Home A Activities	
Eating	345	Brush_teeth	25	Idle	7888
Idle	5883	Eat_brunch	132	Brush_teeth	21
Brush_teeth	75	Eat_dinner	46	Get_drink	21
Get_dressed	70	Get_a_drink	6	Get_snack	24
Get_drink	20	Get_dressed	27	Go_to_bed	11599
Get_snack	8	Go_to_bed	6050	Leave_house	19693
Go_to_bed	7395	Idle	20049	Prepare_Breakfast	59
Leave_house	11915	Leaving_the_house	12223	Prepare_Dinner	325
Prepare_Breakfast	78	Prepare_brunch	82	Take_shower	221
prepare_Dinner	300	Prepare_dinner	87	Use_toilet	154
Prepare_Lunch	58	Take_shower	109	-	-
Shave	57	Use_toilet	39		
Take_medication	6	Wash_dishes	25	-	-
Take_shower	184		-	-	
Use_toilet_downstairs	57	-	-	-	-
Use_toilet_upstairs	35	-	-	-	-
Total	26,486	Total	38,900	Total	40,005

Table 3.4 Frequency distribution of activities in the UCI-HAR Wearable smartphone (inertial sensors) dataset

Activity	Training samples	Testing samples
Walking	1226	496
Walking_upstairs	1073	471
Walking_downstairs	986	420
Sitting	1286	491
Standing	1374	532
Laying	1407	537

Table 3.5 Frequency distribution of wearable wireless identification and sensing datasets

Activity	RoomSet1	RoomSet2
Sit on bed	15162	1244
Sit on chair	4381	530
Lying	30983	20537
Ambulating	1956	335

into the functionality of the temporal models than the accuracy metric [38, 165]. F1-score is used in other evaluations in Chapters 4 and 5. F1-score is calculated in Equations 3.15, and 3.16.

$$\textbf{F1-score} = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{3.15}$$

$$\textbf{recall} = \frac{TP}{TP+FN}, \textbf{precision} = \frac{TP}{TP+FP} \tag{3.16}$$

where TP, FP, and FN are the number of true positives, false positives and false negatives, respectively. Moreover, F1-score is widely used in activity recognition [349, 38].

3.4.3 Implementation Details of the Proposed Networks

The proposed networks use these hyper-parameters, 128, 0.001 and 20% for the batch size, learning rate, and dropout rate, respectively to converge at the minimum of the validation loss. Early stopping as one of the techniques of regularization is used to determine the number of epochs and to prevent overfitting by stopping the training when the validation error of the proposed network starts increasing. 20% dropout rate as another regularization technique after each learning layer is used to further avoid overfitting [350]. Batch normalization as a normalization technique is used to normalize the input data across the batches after each learning layer [265] to make deep learning models faster and more stable during training.

3.5 Results

The experimental results and findings of the proposed networks, ConvNet+Self and ConvNet+Performers, are shown and discussed. The results of the proposed networks ConvNet+Self and ConvNet+Performers for HAR systems are compared with several state-of-the-art methods: HAR+Attention [351], DeepConvLSTM+Attention [14], and many temporal models i.e. LSTM, 1D ConvNet, hybrid of 1D ConvNet and LSTM, Bi-LSTM and CuDNN LSTM. The architectures of the state-of-the-art methods and temporal models are shown and described in Appendix A. To evaluate the proposed methods against existing methods, eight benchmark human activity datasets are used. Tables 3.2 to 3.14 show the results of the experiments in which the proposed networks outperform the existing methods from all the datasets. The results of the proposed networks are shown in bold for all the datasets. Moreover, the proposed networks enhance the performance of the minority classes compared to the existing methods. The achieved results based on each of the datasets are separately discussed and evaluated in the following Sections.

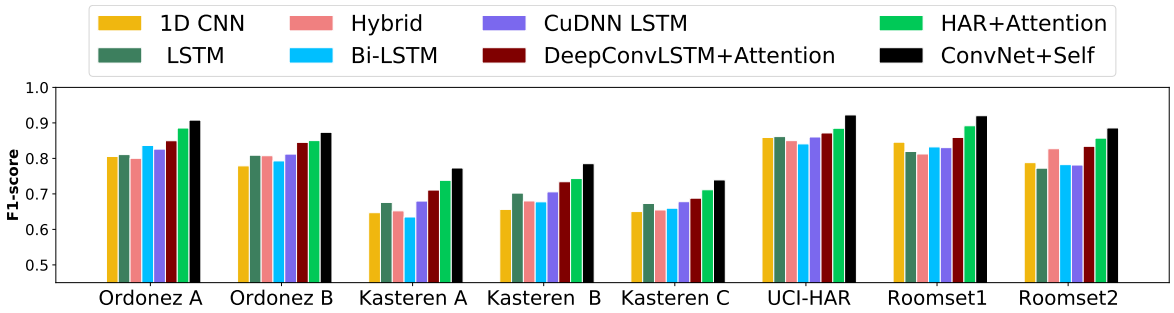


Fig. 3.6 F1-score results of the proposed ConvNet+Self network compared with the state-of-the-art methods and temporal models from eight the datasets

The results of the proposed ConvNet+Self model are compared with the current state-of-the-art methods and basic temporal models as shown in Figure 3.6. The results show that the proposed ConvNet+Self model outperforms the current state-of-the-art methods and basic temporal models from all the sensor collected datasets. Even though the proposed ConvNet+Self model renders better results compared to the current methods, the ConvNet+Self model has several limitations due to adopting the self-attention mechanism. The ConvNet method requires a large amount of memory and high computational capacity since the self-attention mechanism scales quadratically with the length of the input data which creates redundant features and delays the learning process [352]. The limitation of ConvNet+self model resulted in proposing ConvNet+Performers model that is developed using causal convolution and performers-attention based on supervised contrastive learning to accelerate learning process and further improve performance of HAR systems .

To evaluate the proposed methods, the leave-one-day-out cross-validation is used for the smart home datasets as it is commonly used for HAR [38, 19]. The human activity recorded data for a single day are used to inference the model and the recorded data for the rest of the days are used to train the model. This technique is commonly used in HAR. Besides, K-fold cross-validation technique is used to evaluate the wearable sensors data since information about recording dates is not provided in the wearable sensors' data. To show the results of the proposed models, the average F-score of the cross-validation is computed as done in the following research [19, 38, 39, 204].

3.5.1 Results of Ordonez Datasets

The outcomes of the experiments for the proposed networks ConvNet+Self and ConvNet+Performers against the existing state-of-the-art methods and temporal models based on the Ordonez smart environments A and B are shown in Tables 3.6 and 3.7. The results

demonstrate that our proposed networks obtained better results compared with the temporal models (LSTM, 1D ConvNet, hybrid, Bi-LSTM and CuDNN LSTM) in addition to several existing methods [14, 351] for HAR. The proposed networks improves the F1-scores of all the activities, particularly the minority classes. The minority classes such as *Snack, Grooming, Toileting, Showering, Dinner and Breakfast* as shown in Table 3.6 are well improved using our proposed network compared to the existing methods. The proposed network achieved better average results for all classes in addition to the results of each activity in both of the smart home datasets. The results of the proposed networks are shown in bold.

Table 3.6 F1-score results in Ordonez home A dataset

Activity	ConvNet	LSTM	Hybrid	Bi-LSTM	CuDNN LSTM	DeepConvLSTM + Attention	HAR+ Attention	**ConvNet+ Self**	**ConvNet+ Performers**
Breakfast	82.74	80.19	84.65	85.65	79.98	83.11	84.51	**85.71**	**86.89**
Grooming	46.66	57.14	51.28	74.21	62.19	75.32	80.00	**80.01**	**83.00**
Leaving	97.20	97.28	96.43	96.11	96.77	95.29	95.51	**99.75**	**99.81**
Lunch	95.65	96.92	94.87	95.42	95.34	95.44	94.39	**96.93**	**97.11**
Showering	78.94	75.45	77.94	79.42	78.12	80.65	86.89	**93.84**	**93.91**
Sleeping	96.77	96.34	96.63	95.57	94.89	97.53	97.11	**97.63**	**97.69**
Snack	64.66	67.22	55.83	67.02	69.99	70.74	82.63	**84.82**	**85.31**
Spare Time	98.50	98.04	97.84	96.66	98.81	96.83	97.21	**98.57**	**98.82**
Toileting	63.75	61.09	64.71	62.71	67.42	69.89	77.25	**79.76**	**81.26**
Average	80.54	81.07	80.02	83.64	82.61	84.97	88.55	**90.78**	**91.53**

Table 3.7 F1-score results in Ordonez home B dataset

Activity	ConvNet	LSTM	Hybrid	Bi-LSTM	CuDNN LSTM	DeepConvLSTM + Attention	HAR+ Attention	**ConvNet+ Self**	**ConvNet+ Performers**
Leaving	88.34	93.05	90.87	89.52	92.86	89.79	92.09	**93.33**	**94.98**
Sleeping	98.11	85.71	86.76	83.29	86.82	96.37	95.42	**98.30**	**98.45**
Grooming	65.75	85.55	85.36	86.12	81.62	85.33	87.91	**88.87**	**90.11**
Breakfast	75.12	64.44	69.38	68.93	66.83	74.87	75.39	**76.87**	**78.46**
Showering	78.94	79.91	78.56	77.69	80.78	79.43	79.12	**82.84**	**83.32**
Lunch	98.95	81.18	77.00	79.68	83.21	95.21	95.31	**99.63**	**99.65**
Snack	67.92	75.86	73.41	75.42	73.21	76.16	76.31	**78.59**	**79.59**
Toileting	48.91	80.00	83.62	76.47	83.32	83.56	83.24	**86.11**	**88.74**
Spare Time	74.63	78.51	77.24	73.32	77.98	78.21	79.32	**81.48**	**84.39**
Dinner	82.23	84.78	85.32	82.51	85.49	86.19	86.49	**89.45**	**90.22**
Average	77.89	80.89	80.75	79.29	81.21	84.51	85.06	**87.34**	**88.79**

3.5.2 Results of Kasteren Datasets

The results of the proposed networks convNet+Self and ConvNet+Performers based on the datasets A, B, and C from Kasteren smart homes against the temporal models (LSTM, 1D ConvNet, hybrid 1D ConvNet + LSTM, CudNN LSTM and Bidirectional LSTM) in addition to the existing methods are shown in Tables 3.8, 3.9 and 3.10. The proposed networks enhance the performances of each human activity and the average F1-score of all activities including the minority classes such as *Get_dressed, Get_snack* as shown in Table 3.3 compared with the existing methods. The results of the proposed networks are shown in bold.

Table 3.8 F1-score results of Kasteren smart home A dataset

Activity	ConvNet	LSTM	Hybrid	Bi-LSTM	CuDNN LSTM	DeepConvLSTM + Attention	HAR+ Attention	**ConvNet+ Self**	**ConvNet+ Performers**
Get_snack	50.00	53.21	51.23	55.71	56.42	57.22	58.71	**63.69**	**65.24**
Get_drink	51.76	56.84	48.87	42.81	57.21	59.33	59.54	**66.92**	**68.21**
Brush_teeth	20.22	24.08	37.86	21.56	31.46	43.59	52.22	**54.44**	**57.19**
Prepare_breakfast	76.66	74.51	72.41	74.95	75.57	76.97	79.54	**83.32**	**85.31**
Go_to_bed	79.72	74.63	73.21	78.8	73.31	80.16	81.76	**86.54**	**87.99**
Leave_house	79.80	81.58	80.28	76.37	78.89	80.02	82.19	**84.28**	**86.76**
Use_toilet	56.60	66.11	63.06	58.42	67.85	67.82	69.34	**71.97**	**73.21**
Take_shower	84.37	81.43	79.71	74.86	83.69	85.13	89.23	**89.11**	**90.87**
Prepare_dinner	83.20	85.24	80.39	87.94	86.48	89.56	91.87	**95.42**	**96.66**
Average	64.70	67.59	65.22	63.49	67.98	71.09	73.82	**77.29**	**79.04**

3.5.3 Results of Wearable Sensors Datasets

The results of the proposed networks ConvNet+Self and ConvNet+Performers for HAR from wearable sensors data are compared with the results of the existing methods. Tables 3.11, 3.12 and 3.13 show the detailed results of our proposed networks compared with the existing methods. The results of the proposed networks from smartphone sensors data are shown in Table 3.11. The results of the wearable sensors data from Roomset1 and Roomset2 are shown in Tables 3.12 and 3.13 and demonstrate that the proposed networks outperformed the state-of-the-art techniques. The proposed networks enhanced the performance of the individual activity and the average performance of all activities compared to the existing methods from all wearable sensor data. Moreover, the proposed networks improved the results of the minority class such *Sit on chair, Ambulating, and Walking_downstairs* compared with the existing methods. The results of the proposed networks are shown in bold.

Table 3.9 F1-score results in Kasteren smart home B datasets

Activity	ConvNet	LSTM	Hybrid	Bi-LSTM	CuDNN LSTM	DeepConvLSTM + Attention	HAR+ Attention	**ConvNet+ Self**	**ConvNet+ Performers**
Brush_teeth	23.10	37.62	33.25	32.57	39.55	42.89	47.82	**51.18**	**53.21**
Eat_brunch	88.42	90.14	87.53	91.72	89.87	90.93	91.11	**95.92**	**96.21**
Eat_dinner	83.19	85.23	86.68	86.01	86.31	86.79	86.29	**90.02**	**91.87**
Get_a_drink	17.84	31.18	22.34	25.61	33.03	44.15	44.75	**53.00**	**55.21**
Go_to_bed	95.11	99.01	99.21	98.91	97.94	96.32	94.48	**99.73**	**99.88**
Leaving_the_house	91.13	91.75	86.14	87.46	92.00	92.98	93.21	**96.39**	**97.78**
Prepare_brunch	77.48	80.19	83.11	85.92	79.96	85.62	84.29	**88.10**	**89.92**
Get_dressed	16.66	22.58	20.08	27.10	23.41	31.79	41.11	**42.63**	**45.89**
Prepare_dinner	93.11	97.29	94.90	97.00	96.87	96.21	95.31	**97.51**	**97.98**
Take_shower	76.82	79.12	82.71	75.91	78.91	81.95	82.13	**83.13**	**85.32**
Use_toilet	47.78	52.51	47.08	55.71	53.22	56.13	54.19	**62.18**	**64.86**
Wash_dishes	76.61	76.12	73.19	49.28	75.80	75.38	77.25	**82.36**	**84.29**
Average	65.63	70.22	68.01	67.76	70.57	73.42	74.32	**78.51**	**80.20**

Table 3.10 F1-score results in Kasteren home C datasets

Activity	CNN	LSTM	Hybrid	Bi-LSTM	CuDNN LSTM	DeepConvLSTM + Attention	HAR+ Attention	**ConvNet+ Self**	**ConvNet+ Performers**
Eating	76.71	81.32	79.69	80.18	80.36	80.98	84.22	**85.31**	**86.89**
Brush_teeth	51.27	61.56	62.82	60.73	62.59	63.55	67.76	**68.11**	**69.98**
Get_dressed	53.47	55.90	51.47	54.78	56.32	56.82	60.27	**61.17**	**64.19**
Get_drink	42.13	47.61	50.40	38.99	47.91	48.11	50.37	**51.71**	**55.25**
Get_snack	64.14	67.74	65.53	68.16	68.39	67.86	70.45	**72.23**	**74.22**
Go_to_bed	94.86	95.11	91.48	94.21	96.04	95.41	93.21	**96.12**	**97.59**
Leave_house	93.81	90.18	92.74	89.05	91.52	92.57	91.39	**94.17**	**95.76**
Prepare_Breakfast	76.35	75.74	73.15	77.78	76.81	78.45	81.24	**83.42**	**84.68**
Prepare_Dinner	77.01	79.74	68.32	71.53	78.49	79.68	80.14	**84.29**	**86.21**
prepare_Lunch	74.12	76.21	77.21	73.55	77.07	78.39	79.31	**85.73**	**87.63**
Use_Toilet_Downstairs	42.68	40.90	41.46	37.98	41.69	45.17	49.55	**59.29**	**62.59**
Use_toilet_upstairs	35.21	43.27	30.97	45.57	45.32	46.21	48.64	**51.19**	**53.44**
Shave	73.83	75.32	77.15	71.73	76.42	78.25	77.82	**81.01**	**83.41**
Take_medication	48.37	43.74	45.32	49.32	42.39	45.21	56.52	**57.09**	**61.36**
Take_shower	72.18	75.29	74.34	75.87	75.71	75.42	76.61	**78.16**	**80.99**
Average	65.02	67.30	65.47	65.96	67.80	68.79	71.16	**73.93**	**76.27**

3.5.4 Ablation study of the proposed network

Since the proposed network ConvNet+Performers demonstrated better results compared to the proposed network ConvNet+Self, we perform an ablation study to show the contribution

Table 3.11 F1-score results in smartphone dataset

Activity	ConvNet	LSTM	Hybrid	Bi-LSTM	CuDNN LSTM	DeepConvLSTM + Attention	HAR+ Attention	**ConvNet+ Self**	**ConvNet+ Performers**
Laying	89.03	87.14	86.76	85.35	87.51	89.67	89.91	**95.69**	**96.79**
Sitting	84.32	82.29	81.22	82.53	81.98	86.45	90.25	**95.94**	**95.99**
Standing	88.38	86.71	87.42	86.32	86.43	88.92	88.56	**92.77**	**94.32**
Walking	75.90	80.17	78.87	75.64	81.03	80.89	81.11	**84.89**	**86.89**
Walking_downstairs	76.31	80.01	79.11	78.76	80.21	80.11	83.91	**84.14**	**86.38**
Walking_upstairs	96.44	95.42	94.63	95.89	96.05	96.93	97.23	**100.00**	**100.00**
Average	85.89	86.14	85.00	84,08	86.03	87.16	88.49	**92.24**	**93.39**

Table 3.12 F1-score results in wearable dataset of RoomSet1

Activity	ConvNet	LSTM	Hybrid	Bi-LSTM	CuDNN LSTM	DeepConvLSTM + Attention	HAR+ Attention	**ConvNet+ Self**	**ConvNet+ Performers**
Ambulating	92.91	92.58	93.67	92.02	93.19	93.67	95.22	**97.63**	**98.01**
Lying	93.94	91.34	87.94	94.71	92.97	94.12	95.21	**97.70**	**97.92**
Sit_on_bed	94.91	94.74	95.89	94.64	95.94	95.31	96.25	**99.90**	**99.93**
Sit_on_chair	56.51	49.12	47.48	51.58	50.04	60.52	70.11	**72.84**	**74.41**
Average	84.56	81.94	81.24	83.23	83.03	85.90	89.19	**92.02**	**92.56**

Table 3.13 F1-score results in wearable dataset of RoomSet2

Activity	ConvNet	LSTM	Hybrid	Bi-LSTM	CuDNN LSTM	DeepConvLSTM + Attention	HAR+ Attention	**ConvNet+ Self**	**ConvNet+ Performers**
Ambulating	79.42	83.75	84.95	78.95	84.67	85.43	87.11	**89.79**	**91.93**
Lying	89.75	82.29	89.50	84.97	83.16	89.42	89.32	**94.85**	**95.32**
Sit_on_bed	94.74	94.66	96.75	95.49	95.21	96.21	97.52	**99.79**	**99.95**
Sit_on_chair	51.31	48.27	59.70	53.75	49.51	62.56	68.87	**69.87**	**72.31**
Average	78.80	77.24	82.72	78.24	78.13	83.40	85.70	**88.57**	**89.87**

of each component in the ConvNet+Performers network for HAR systems. The proposed network ConvNet+Performers without performers attention, two stages learning, causal convolution, and focal loss. Table 3.14 demonstrate the results of the proposed network without these four components and the proposed network from the experimental datasets. The results indicate the impact of each component in the proposed network. For instance, the proposed network obtained the F1-score of 91.53, while the proposed network without performers attention obtained the F1-score of 87.64, without two stages learning obtained the F1-score of 86.73, without causal convolution obtained the F1-score of 88.29 and without using the focal loss, the F1-score is 88.42. This example confirms the contribution of supervised contrastive learning. Moreover, the proposed network without using two stages of learning has gained the lowest results from the sensor datasets compared with other

components of the proposed network. Hence, the results show that the higher contribution is made by the proposed supervised contrastive learning with two stages of learning in the proposed network compared to the performers-attention, causal convolutions, and focal loss.

Table 3.14 Ablation study results of the proposed network

Datasets	Without Performers attention	Without two stage learning	Without Causal	Without focal loss	ConvNet+ Performers
Ordonez Home A	87.64	86.73	88.29	88.42	**91.53**
Ordonez Home B	85.36	85.03	86.14	85.32	**88.79**
Kastern Home A	75.26	75.00	77.84	75.96	**79.04**
Kastern Home B	77.43	76.21	78.67	77.43	**80.20**
Kastern Home C	74.35	73.46	75.03	74.39	**76.27**
Smartphone dataset	89.42	87.11	90.93	91.71	**93.39**
Wearable RoomSet1	88.19	88.23	89.24	87.18	**92.56**
Wearable RoomSet2	85.42	85.85	87.32	86.43	**89.87**

3.6 Discussions

In this chapter, two sequential deep learning networks for HAR systems from sensors data are proposed. The networks ConvNet+Self and ConvNet+Performers aim to further improve recognising of human activity in the datasets collected from smart home environments and wearable sensors.

First, ConvNet+Self network comprises dilated causal convolution with multi-head self-attention is proposed to entirely forgoes recurrent settings, accelerate training time and improve the performance of HAR systems from smart home and wearable sensor data. Thorough experiments are conducted on eight real-world smart home and wearable datasets to evaluate the proposed method against the temporal and recurrent-based architecture methods. The results of the experiments show that the proposed method significantly improved the accuracy of HAR systems compared with the state-of-the-art techniques. The proposed method improved the performance of HAR systems by up to 7% compared with LSTM, 1D CNN, hybrid 1D CNN + LSTM, CuDNNLSTM, and Bidirectional LSTM using wearable sensors and smart home sensors data. Despite the effectiveness of the proposed ConvNet+self network, this method requires high computation and memory cost due to adopting self-attention mechanism. The operation of the self-attention mechanism scales quadratically with the input sequence length which can increase training time and creates redundant features because it appends more weight parameters to the model. To address this limitation, we proposed ConvNet+ performers model to further accelerate the training time and enhance

the performance of HAR by adopting a lightweight attention mechanism and supervised contrastive learning.

The proposed ConvNet+performers network that comprises causal ConvNet-based performers-attention and supervised contrastive learning accelerates the learning time and improves the performance of HAR systems. This is because, firstly, the performers-attention mechanism linearly scales with the length of the sensor input data which makes the learning process faster. Secondly, supervised contrastive learning increases the performance of the proposed network by replacing one stage learning with two stages of learning where both the first stage is representation learning and the second stage is classifier learning.

Extensive experiments are performed on eight datasets to evaluate the proposed networks compared to the basic temporal models and existing state-of-the-art methods. The results of the thorough experiments reveal that the proposed networks outperform the current methods and reduce the learning time compared with the existing state-of-the-art methods. We further perform ablation studies to highlight the contribution of each component of the proposed ConvNet+Performers network. The results of the ablation studies show that the proposed supervised contrastive learning with two stages of learning provides a larger contribution in our proposed network compared with the performers-attention, causal convolutions, and focal loss.

3.7 Conclusion

To improve HAR systems, two deep learning models are proposed in this chapter. We proposed ConvNet+Self comprised dilated causal convolution with the self-attention mechanism to entirely forgoes recurrent settings and improve the performance of HAR.

Thorough experiments are performed on eight real-world smart home and wearable datasets to evaluate the proposed method against the temporal and recurrent-based architecture methods. The results of the experiments show that the proposed method significantly improved the accuracy of HAR systems compared with the state-of-the-art techniques. The proposed method improved the performance of HAR systems by up to 7% compared with LSTM, 1D CNN, hybrid 1D CNN + LSTM, CuDNNLSTM, and Bidirectional LSTM using wearable sensors and smart home sensors data. Even though the proposed method outperformed the state-of-the-art approaches, the proposed method adds more learning parameters to the model due to adopting self-attention mechanism which scales quadratically with the input sequence length and increases training time. To address this limitation, we proposed ConvNet+performers model to further accelerate the training time and enhance the perfor-

mance of HAR by adopting a lightweight attention mechanism and supervised contrastive learning.

The proposed ConvNet+Performers method comprised a causal supervised contrastive ConvNet based on performers-attention. This proposed network improves further the results of the HAR systems in sensors generated data. In addition, the proposed method also accelerates the learning process compared to the existing methods.

Chapter 4

Class Imbalanced Human Activity Recognition

4.1 Introduction

Deep learning techniques have been applied for ADLs recognition and reported satisfying results. However, it is still challenging and remains an open research issue to build an accurate HAR system due to the high diversity of human activities [353, 354, 22, 355]. Besides, the frequency variance of human activities is usually imbalanced leading to additional challenges. When building a machine learning model with an imbalanced dataset, it tends to partially or completely ignore the minority classes in order to achieve satisfying overall accuracy. For example, in HAR datasets, cooking, watching TV in the living room, and sleeping usually occur at a higher frequency than showering and snack eating. Moreover, the intrinsic property of physical human activities makes classes representing imbalanced, which leads to the importance of the topic of HAR learning algorithms for imbalanced class handling, especially with a large dataset for deep learning study.

To address the imbalance class problems For HAR systems, this thesis proposes two approaches for both algorithm-level and data-level solutions to handle skewed class proportions.

1. For the algorithm level, we propose joint learning of temporal LSTM and 1D ConvNet models in order to learn from the same form of input features for HAR. Different from ensemble learning that often combines the outputs of many learners while using a specific aggregation function to handle imbalanced data [311], the proposed method combines the learning processes of two temporal models in a single joint training

mechanism to improve the accuracy on minority classes in addition to maintain the accuracy on majority classes. Therefore, joining the learning processes of two different temporal models in the proposed method is expected to obtain a better combined model compared to simply aggregating the outputs of multiple learners. It is also expected to obtain more accurate and reliable estimates or decisions than single models. The two temporal learners of the jointly proposed methods can exploit different features from the input data to rendering a strong mutual complementary model. Complementarity in joint learning based on different models can greatly boost the performance as compared to simply combining the same learners (e.g., LSTM with LSTM in this work) in a joint learning model [312]. This is because each base learner brings different features into the joint learner to enrich the joint learning process and each learner improves the earlier layers of the other learner, but in the same time the weaknesses of each individual learners are avoided. The proposed method jointly trains the two base learner i.e., LSTM and 1D ConvNet , and combines the based learners by a fully connected layer, which is followed by the output layer. The joint optimization that leads to increasing the functionality of the proposed joint temporal model to gain more insight into the input data and features reduces the recognition error rate. Thereby, the proposed model increases the performance of activity recognition particularly for minority classes.

2. For the data-level, to balance human activity datasets, synthetic minority over-sampling technique (SMOTE) [302] as a common method is mostly used. However, SMOTE in generating new instances fails to consider neighbouring instances of different classes which can append further noise and expand overlapping of the classes. To overcome this problem, we propose an improved SMOTE (iSMOTE) version that computes *k*-nearest neighbours of each generated instances to make sure the new instances are accurately annotated. Each new instance of the minority classes with its *k*-nearest neighbours must have the same class. For example, generated new instances of *Snack* activity must have *k Snack* activity as nearest neighbors.

The proposed approaches to handle imbalanced class problems from the algorithm-level and data-level are further delineated in the following sections. Firstly, this thesis presents a deep learning model to handle skewed class proportions during the learning processing without changing the distribution of the input data. Secondly, a data-level method is presented which changes the distribution of the input datasets. These approaches are properly handled imbalanced class problems and improve human activities.

4.2 Joint ConvNet and LSTM Temporal Models

In this section, we propose joint learning of the temporal model from an algorithm-level perspective to handle the imbalanced class problem for HAR systems. Temporal models are jointly employed in order to learn and recognise human daily living activities from smart homes aiming at increased diversity between base learners which is crucial for joint learners. The aim of joining different learner models is to produce a mutual complementary network by contributing each network with different learning approaches to build strong joint learners with good performance. Therefore, robust learners LSTM and 1D ConvNet for temporal data are used, which have high variance and low bias due to their almost universal function approximation ability [349] for delivering the joint learning recognition. Furthermore, the different base learners in the proposed model can expose features of different aspects of the input data which can boost the recognition performance. Joint different temporal models in addition to using leave-one-out cross-validation are used to reduce high variance. Figure 4.1 shows the architecture of the proposed joint learning of temporal models. Different from the Hybrid 1D ConvNet + LSTM model, our proposed joint learning of the temporal model includes the two parallel sequences that include LSTM and 1D ConvNet. The proposed method is composed of the following layers: LSTM, 1D ConvNet, and fully connected layers. Here, we show the details of these layers:

- The raw temporal human activity data are used as the input of the model and segmented by fuzzy temporal windows for feature extraction before passing to the deep learning model.
- The proposed deep learning model has two parallel temporal models, i.e., LSTM and 1D ConvNet.
- The first part of the model consists of two LSTM layers.
- The second part of the model consists of two 1D ConvNet layers each with 64 filters. The kernel size is equal to 3 which specifies the length of the 1D convolution window, and the stride is equal to 1.
- Each LSTM and 1D ConvNet layer is followed by a dense layer to make the output shape of the LSTM and ConvNet layers compatible for the next shared fully-connected layer since the output shapes of LSTM and 1D ConvNet layers are different.
- Features from two separate dense layers are combined (fused) and fed to the next shared layer.

- One shared fully connected layer with ReLU activation function is followed.
- The final layer is the output layer with a softmax activation function to recognise and distinguish human activities.

Two layers in each of the LSTM and 1D ConvNet followed by a dense layer are used in order to build the model. The two multi-layer temporal models are jointly trained with a 0.001 learning rate. A new shared fully-connected layer is added and connected to the dense layers of both individual temporal models. The shared fully-connected layer aggregates different exposed features of the two different models to boost the recognition performance of the minority classes in addition to the majority classes. Moreover, aggregating different exposed features in the proposed learning method helps the classifier in detecting rare activities and avoids having models biased toward one class or the other when compared to an individual learner. During updating parameters i.e., model weights, both LSTM and 1D ConvNet models contribute to adjusting the weights through the shared layer to correctly map the input data to the output class activities. Hence, the new shared layer is used to further learn and share learned information across both joint models of the system to allow each temporal model to improve the earlier layer of the other temporal model. Thereby, the joint optimisation will maximise their capacity to enhance the recognition performance of all the classes, including the minority classes. Designing the parallel and joint learning of temporal models by combining the order-sensitivity of LSTM with the speed and lightness of 1D ConvNet renders an efficient model for human activity recognition. The shared fully connected layer is followed by an output layer to properly recognise human activities. When designing the deep learning structure, we consider joint robust learners with diversity, which can help to boost the recognition performance of minority classes.

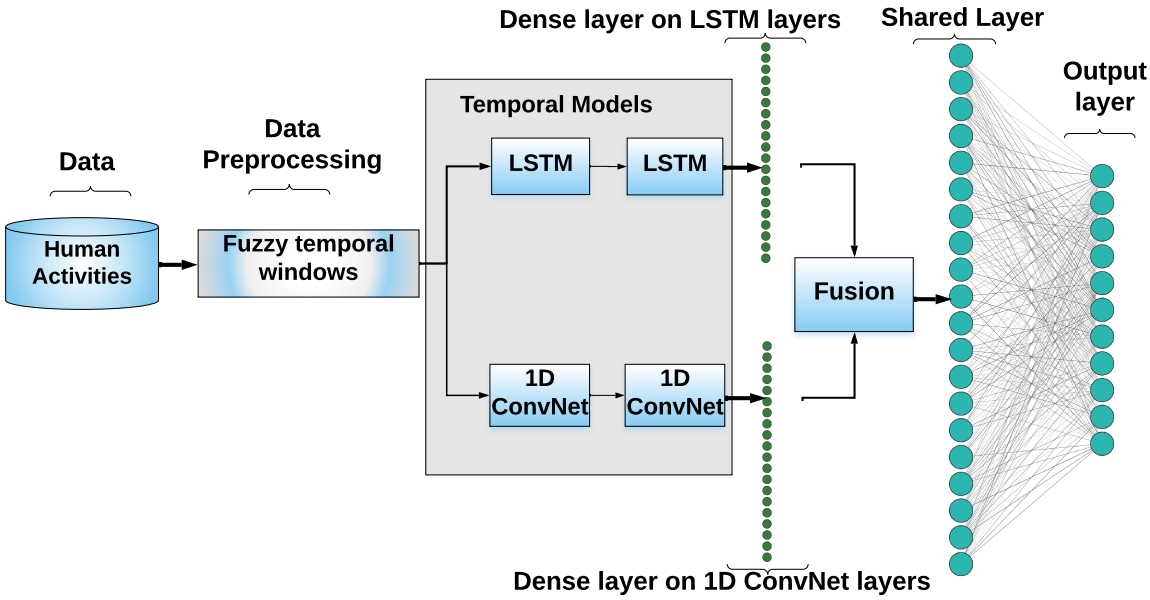


Fig. 4.1 Architecture of the proposed joint learning of temporal model for HAR

4.3 Improved Synthetic Minority Oversampling Technique (iSMOTE)

The SMOTE technique oversamples minority classes to create balanced classes in the training dataset [302]. The instances of the minority classes are only used by the SMOTE technique to handle the generation and distribution of synthetic instances to make a balanced dataset. SMOTE presumes that an instance generated between the nearby instances in the minority class is still an instance of the class. Particularly, the essential principles of the SMOTE technique are as follows:

i. For each instance in the minority class x_i, SMOTE selects k-nearest neighbours only within the same class.

ii. SMOTE randomly picks one sample $\bar{x}_i$ from the k-nearest neighbours and then generates a new synthetic sample x_{new} using Equation 4.1.

$$x_{new} = x_i + \lambda \times (\bar{x}_i - x_i) \tag{4.1}$$

where $\lambda \in [0,1]$ is the random number, which allows to randomly create the new instances x_{new} along the line between x_i and $\bar{x}_i$. However, SMOTE in oversampling minority classes fails to take neighbouring samples from other classes into account. This can maximise the

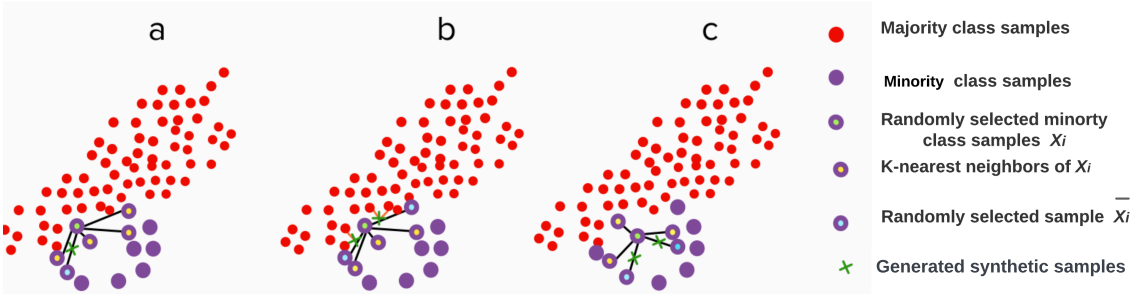


Fig. 4.2 Overview of the proposed iSMOTE technique

noise in the datasets and increases the overlapping of the classes. To address this problem, we propose iSMOTE which computes K nearest neighbours of each generated synthetic instance to avoid the misgenerated new sample. iSMOTE only accepts the newly generated instance for each of the minority classes that its class is the same class as the K nearest neighbours. For instance, the new generated synthetic instances of *Snack* activity must have K *Snack* activities as nearest neighbours to be accepted by the iSMOTE. Figure 4.2 (a) shows an example of how SMOTE oversamples minority classes, while Figures (b) and (c) illustrate the proposed iSMOTE method. Since generated synthetic samples in Figure 4.2 (b) are very close to the samples of other classes, iSMOTE rejects the generated samples whereas generated instances from Figure 4.2 (c) are accepted because they can have K nearest neighbours from their classes. iSMOTE generates new samples for the minority classes and mitigates the impact of the imbalanced class issues in learning. Hence, iSMOTE avoids misgenerating new samples for the minority classes during the oversampling process.

4.4 Experimental Setup

In this section, we will show the details of the experiments of the proposed joint temporal model from five collected smart home sensor datasets. The experiments and results of the iSMOTE are shown in the next chapter because iSMOTE is integrated with a proposed method to reduce the need for large annotated data.

4.4.1 Datasets and Prepossessing

Five public smart home datasets are used to evaluate the joint temporal models to handle imbalanced class problems for HAR systems. Only Smart home datasets are used since wearable sensor data are less imbalanced. The five smart home sensors collected datasets

are all described in Chapter 2.4.1 on Page 20. The smart home datasets are collected from Ordonez home A and B [111] and Kasteren home A, B and C datasets [140, 141]. Table 4.1 shows details of these five smart home datasets with respect to the residents, sensors, and the number of activities. Besides Tables 4.2 and 4.3 show the frequency of the human activities from Ordonez homes A and B as well as Kasteren homes A, B and C datasets after preprocessing and segmentation using FTWs which is delineated in Chapter 2.5.2 on Page 27. The frequency of human activities in these smart homes shows that the activities are highly imbalanced. The proposed joint temporal network is evaluated on the datasets with multiple human activities ranging from 9 to 13 from collected sensor data.

Table 4.1 Smart home environment datasets

	Ordonez smart home		Kasteren smart home		
	Home A	Home B	Home A	Home B	Home C
Setting	Home	Home	Apartment	Apartment	House
Gender	-	-	Male	Male	Male
Activities	10	11	10	13	16
Age	-	-	26	28	57
Rooms	4	5	3	2	6
Sensors	12	12	14	23	21
Duration	14 days	21 days	25 days	14 days	19 days

Table 4.2 Frequency of human activities in the Ordonez datasets

Activity	Home A	Home B
Leaving	1664	5268
Breakfast	120	309
Toileting	138	167
Spare Time/ TV	8555	8984
Dinner	-	120
Sleeping	7866	10763
Snack	6	408
Grooming	98	427
Showering	96	75
Idle	1598	3553
Lunch	315	395
Total	20456	30427

Table 4.3 Frequency of activities in the Kasteren datasets

Home C Activities		Home B Activities		Home A Activities	
Eating	345	Brush_teeth	25	Idle	7888
Idle	5883	Eat_brunch	132	Brush_teeth	21
Brush_teeth	75	Eat_dinner	46	Get_drink	21
Get_dressed	70	Get_a_drink	6	Get_snack	24
Get_drink	20	Get_dressed	27	Go_to_bed	11599
Get_snack	8	Go_to_bed	6050	Leave_house	19693
Go_to_bed	7395	Idle	20049	Prepare_Breakfast	59
Leave_house	11915	Leaving_the_house	12223	Prepare_Dinner	325
Prepare_Breakfast	78	Prepare_brunch	82	Take_shower	221
prepare_Dinner	300	Prepare_dinner	87	Use_toilet	154
Prepare_Lunch	58	Take_shower	109	-	-
Shave	57	Use_toilet	39		
Take_medication	6	Wash_dishes	25	-	-
Take_shower	184		-	-	
Use_toilet_downstairs	57	-	-	-	-
Use_toilet_upstairs	35	-	-	-	-
Total	26486	Total	38900	Total	40005

4.4.2 Implementation Details of the Proposed Joint Networks

This thesis uses a range of learning rates from 0.0001 to 0.001, a range of batch sizes from 32 to 256, a range of dropout rate from 20% to 40%, to converge at the minimum of the validation loss. Early stopping as one of the techniques of regularization is used to determine the number of epochs and to avoid overfitting by stopping the training when the validation error of the proposed joint temporal networks starts increasing. A series of trial and error experiments are conducted over these ranges. The experiments revealed a learning rate of 0.001 and a batch size of 64 with a 40% dropout rate are optimal for the models to converge. While a large batch size often can render rapid training, it needs more memory space and it delays the convergence of deep learning models. On the contrary, smaller batch sizes that need less memory space could make the training process slower but could make the convergence of deep learning models faster; therefore, it is mostly a trade-off problem [356]. To prevent the models from overfitting, the 40% dropout rate as a regularization technique in addition to early stopping is used [350]. The dropout technique ignores neurons that are randomly selected during the training process. The dropout technique temporally disconnects the ignored neurons on the forward pass; hence, in the backward pass their weights will not be updated.

4.5 Results

In this section, the experimental results of the experiments of the proposed joint learning of temporal models are presented and discussed. The joint temporal model is compared with LSTM, 1D ConvNet, and hybrid 1D ConvNet+LSTM. The architectures of the temporal models, i.e., LSTM, 1D ConvNet, and hybrid 1D ConvNet+LSTM are shown and described in Appendix A. The leave-one-day-out cross-validation is used in the evaluation for all of the models, specifically, the human activities on an individual day are used for testing of the models and the models are trained on the human activities of the rest of the days. This procedure is circulated until the human activities data from all the recorded days are involved in the testing set [121]. The average F1-score is calculated from the results of the cross-validation for all the models that have successfully been performed in [204, 268]. Because the classes of the human activity datasets collected from smart homes are highly imbalanced, the proposed joint learning of temporal models handles the imbalanced human activity classes and avoids having classifiers biased toward the majority classes.

4.5.1 Results of Ordonez Datasets

The proposed joint learning of temporal models are compared with the individual and hybrid learners from Ordonez Home A and B datasets. Table 4.4 shows the results of the proposed joint temporal models that outperforms LSTM, 1D ConvNet, and the hybrid ConvNet+LSTM. The proposed method specifically enhanced the results of the minority classes such as *Showering, Breakfast, Dinner, and Snack* in addition the majority classes such as *Sleeping and Spare_time/TV*. When compared with LSTM, 1D ConvNet, and the hybrid model, the use of our proposed model increases the F-scores by 4%, 6%, and 6% in Ordonez home A, and by 4%, 6%, and 10% in Ordonez home B respectively. Regarding to the infrequent activities *Breakfast, Grooming, Lunch, showering, toileting, snack, Dinner*, the proposed method improves F-scores by 4%, 8%, 3%, 7%, and 3% from Ordonez home A and by 4%, 1%, 2%, 2%, 11%, and 10% in Ordonez home B, respectively. This confirms the proposed model is capable of achieving better performance for the recognition of minority classes. When compared with LSTM, the proposed method improves F-scores by 4% from Ordonez home A, B, respectively. When compared with 1D ConvNet, the proposed method improves F-scores by 5% from Ordonez home A, B, respectively. When compared with the hybrid model, the proposed method improves F-score by 6% and 10% from Ordonez homes A, and B, respectively.

Table 4.4 F1-score results of Ordonez Smart home datasets.

Activities	Home A				Home B			
	LSTM	**1D ConvNet**	**1D ConvNet+ LSTM**	**Joint Learning**	**LSTM**	**1D ConvNet**	**1D ConvNet+ LSTM**	**Joint Learning**
Breakfast	82.27	78.43	86.79	**87.05**	78.65	77.10	73.91	**82.74**
Grooming	62.06	46.66	57.14	**70.96**	62.99	59.67	53.63	**64.08**
Leaving	89.90	88.60	88.39	**91.73**	96.43	97.31	96.48	**98.19**
Lunch	95.50	95.45	94.57	**98.31**	86.45	84.47	79.76	**88.13**
Showering	75.86	80.00	64.00	**82.35**	75.00	80.00	51.81	**82.71**
Sleeping	97.23	97.23	97.13	**99.66**	99.47	99.49	99.26	**99.62**
Snack	66.66	66.66	66.66	**73.32**	70.37	68.96	62.11	**73.19**
Spare_Time/TV	97.79	95.93	97.28	**98.97**	95.81	94.51	95.51	**96.60**
Toileting	72.21	69.23	69.84	**75.23**	18.51	0.07	18.46	**31.18**
Dinner	-	-	-	-	40.00	34.28	29.41	**50.27**
Total	82.16	79.79	80.20	**86.39**	72.36	69.59	66.03	**76.51**

4.5.2 Results of Kasteren Datasets

Regarding the Kasteren datasets A, B, and C, the F1-scores of the proposed joint learning of temporal models compared with the LSTM, 1D ConvNet, and hybrid model are shown in Tables 4.5 , 4.6 and 4.7. The results show that the joint learning of temporal models outperforms the individual temporal and hybrid learners by more than 4% in total from all the datasets. The results reveal that the minority classes from all datasets are improved by the proposed method when compared with the individual learners and the hybrid learner. For example, the proposed method improves F1-scores by 4% from Kasteren home A, as well as by 9% and 11% from Kasteren home B and C, respectively. When compared with 1D ConvNet, the proposed method improves F1-scores by 11%, 15%, and 14% from Kasteren home A, B, and C, respectively. When compared with the hybrid model, the proposed method improves F-score by 6%, and 10% from Ordonez home A, B respectively.

Furthermore, the results of the minority classes in addition to majority classes are increased using the proposed model when compared with the individual models and the hybrid model from Kasteren datasets A. The minority classes from home A are *Brush_teeth, Get_drink, Get_Snack, Prepare_Breakfast, Take_shower, Use_toilet*, those from home B are *Get_a_drink, Get_dressed, Use_toilet, Brush_teeth, Eat_dinner, Wash_dishes*, and those from home C are *Brush_teeth, Get_dressed, Get_snack, Get_drink, Go_to_bed, Prepare_Breakfast,*

Prepare_Dinner, Prepare_Lunch, Shave, Take_medication, and Use_toilet. The experimental results show that the proposed joint learning of temporal models can improve the performance of minority classes besides the majority classes for human activity recognition.

All of the models perform poorly for some minority activities, such as *Get_drink* with only 20 samples, and *take medication* with only six samples from Kasteren Home C. Firstly, a small number of samples are given high dimensional data with 315 features (21 sensors * 15 fuzzy temporal windows) from Kasteren Home C dataset, and one likely problem is the curse of dimensionality of the smart home datasets, since high dimension creates difficulties for the classifier to search. Secondly, the results indicate that there is not enough variation in the smart home data with respect to these two activities and the input features are non-informative and useless in the separation of these activities from the rest. To further improve these minority classes, imbalanced data could be handled from data level, i.e., oversampling minority class using iSMOTE [302] in addition to handling imbalanced data from algorithm level as we have performed by proposing joint learning, which could be considered in the future work of this study.

To further evaluate, the proposed method is compared with joint learning based on two LSTM models and joint learning based on two 1D ConvNet models with the same configuration of the proposed method. The results show that the proposed method achieves a higher F1-score as shown in Figure 4.3. This indicates that the diversity and complementarity of the proposed method are more important than training combined the same learners, i.e., LSTM with LSTM or 1D ConvNet with 1D ConvNet.

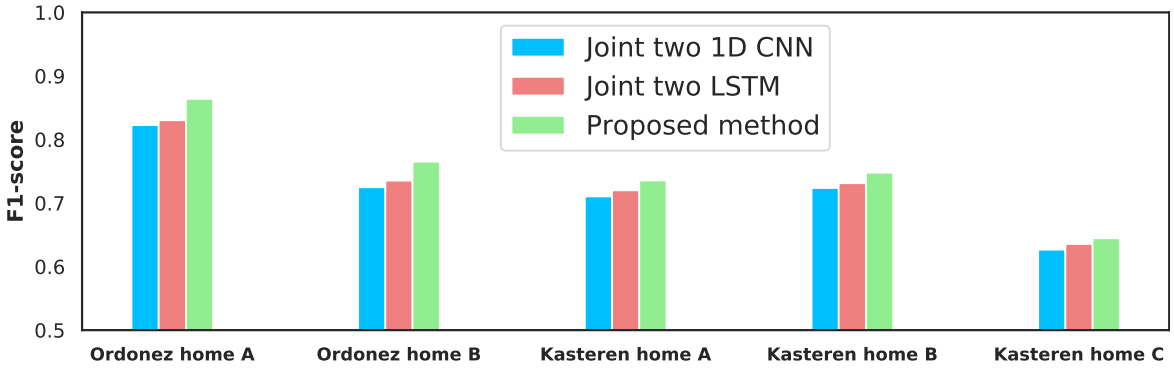


Fig. 4.3 Joint learning compared with joint LSTM+LSTM and Joint 1D ConvNet+1D ConvNet.

Table 4.5 F1-score results of Kasteren Smart home A datasets.

Activities	**LSTM**	**1D ConvNet**	**1D ConvNet+ LSTM**	**Joint Learning**
Brush_teeth	51.09	22.03	50.00	**56.32**
Get_drink	40.00	22.20	40.01	**47.11**
Get_Snack	30.14	22.22	28.57	**43.36**
Go_to_bed	88.20	88.18	87.96	**89.96**
Leave_house	99.53	99.75	99.45	**99.88**
Prepare_breakfast	78.00	72.00	75.00	**79.19**
Prepare_Dinner	88.88	94.01	96.55	**96.73**
Take_shower	85.24	79.45	80.00	**86.31**
Use_toilet	60.86	56.60	57.69	**63.33**
Total	69.10	61.82	68.07	**73.54**

Table 4.6 F1-score results of Kasteren Smart home B datasets.

Activities	**LSTM**	**1D ConvNet**	**1D ConvNet+ LSTM**	**Joint learning**
Brush_teeth	0.00	0.00	16.66	**28.57**
Eat_brunch	91.42	89.28	91.22	**92.12**
Eat_dinner	88.00	88.88	83.33	**88.91**
Get_a_drink	0.00	0.00	00.00	**40.00**
Go_to_bed	99.08	99.20	99.28	**99.66**
Leaving_the_house	95.7	90.89	88.09	**98.72**
Prepare_brunch	84.65	78.57	82.75	**87.80**
Get_dressed	26.66	0.00	15.38	**49.63**
Prepare_dinner	96.36	96.96	90.90	**97.11**
Take_shower	83.63	74.50	80.00	**84.93**
Use_toilet	40.00	25.00	16.66	**51.33**
Wash_dishes	74.33	72.72	66.66	**77.72**
Total	65.40	59.66	61.74	**74.70**

Table 4.7 F1-score results of Kasteren home C datasets.

Activities	**LSTM**	**1D ConvNet**	**1D ConvNet+LSTM**	**Joint Learning**
Eating	74.28	80.00	80.00	**82.70**
Brush_teeth	47.61	58.33	50.00	**62.50**
Get_dressed	48.48	53.33	32.87	**54.67**
Get_drink	00.00	0.00	0.00	**32.85**
Get_snack	66.66	30.00	66.66	**68.00**
Go_to_bed	93.65	94.68	91.48	**94.86**
Leave_house	92.86	91.81	91.64	**98.96**
Prepare_Breakfast	75.00	72.22	36.36	**76.75**
Prepare_Dinner	75.55	80.70	54.44	**83.69**
prepare_Lunch	70.00	72.72	72.72	**74.35**
Use_Toilet_Downstairs	15.38	0.00	05.26	**22.22**
Use_toilet_upstairs	13.33	0.00	16.66	**18.38**
Shave	66.66	70.00	44.44	**78.88**
Take_medication	0.00	0.00	0.00	**28.42**
Take_shower	70.00	72.13	70.96	**74.65**
Total	53.96	51.68	47.56	**64.46**

4.5.3 Model Interpretability

In this section, permutation feature importance (PFI) as a useful mechanism of model interpretability [357–359] is conducted to show more insight of each LSTM and 1D ConvNet independently. PFI is used to compute and rank input feature importance from Ordonez smart home A and B datasets based on how useful features are at HAR for LSTM and 1D ConvNet. PFI reflects how each predictor variable is important from HAR for each LSTM and 1D ConvNet. Experiments that are based on PFI show that the most and least important features from both datasets are different in HAR for LSTM and 1D ConvNet. This indicates the most important features in recognition process varies between LSTM and 1D ConvNet, which has been considered to take advantage of the most important features in both models in the proposed joint learning. Hence, the proposed joint temporal model exposes different features from input data into the joint learning to improve the performance of HAR, particularly minority classes given the different rank of the important features from both models. The

rank of features in LSTM and 1D ConvNet based on the PFI is different, which is indicative of how much each LSTM and 1D ConvNet models relies on the features. This illustrates that the joint proposed model takes feature importance in both models into account in joint learning. Therefore, the joint learning can use a set of the most important features from one model and a different set from the other model to contribute in the recognition to generally improve the performance of HAR particularly minority classes. Algorithm 2 designed based on [360, 359] shows the process of PFI in detail. Firstly, the result scores, i.e., the F1-score of each LSTM and 1D ConvNet, are computed. Next, a single feature from the dataset is randomly shuffled to generate a permuted version of the dataset. This mechanism removes the relationship between the features and the true labels. Subsequently, the LSTM and 1D ConvNet models are independently applied on the permuted version of the dataset to compute the result score. Finally, we subtract the result score based on permuted data from the result score of the original data. This mechanism measures the features' importance by computing the error of the LSTM and 1D ConvNet after permuting the feature. Moreover, after applying permutation on a feature, the decrease of the F1-score indicates the model's dependence on the permuted feature. This mechanism is applied on all of the features to compute feature importance.Tables 4.8 and 4.9 show PFI on the Ordonez smart homes A and B datasets and the rank of the importance of each feature. Further, the value of the mean and standard deviation of the F1-score for N runs after permuting the feature is presented. Figures 4.4 and 4.5 show the mean results of f1-score with $N = 12$ run for 12 sensors after permuting the sensor features.

We further analyse the effect of the proposed method for the performance improvement to the minority classes: *Breakfast, Grooming, Lunch, showering, toileting, snack, Dinner.* For example, the feature PIR_Cooktop_Kitchen is the most important feature for LSTM but the least important feature for 1D ConvNet from smart home A, as shown in Table 4.8. In contrast, the feature Electric_Microwave_Kitchen is one of the most important features for 1D ConvNet, but the last ranked and least important feature for LSTM. In addition to the two aforementioned features, the features Magnetic_Fridge_Kitchen, Magnetic_Cupboard_Kitchen, and Electric_Toaster_Kitchen have different rankings from LSTM and 1D ConvNet, where they contribute to the recognition of minority classes of the kitchen area, such Breakfast, Lunch and snack. Moreover, the PIR_Basin_Bathroom, Flush_Toilet_Bathroom, and PIR_Shower_Bathroom features have different rankings where they contribute to the recognition of minority activities of the bathroom area, such as Grooming, showering, and toileting. Features from smart home B also have different rankings where they contribute to the recognition of minority activities from both kitchen and bathroom

areas. For example, PIR_Shower_ Bathroom is one of the most important features for LSTM, but the least important features for 1D ConvNet, as shown in Table 4.9. Moreover, the Magnetic_Fridge_Kitchen and Electric_Microwave_Kitchen are the most important features for 1D ConvNet where they can contribute to the recognition of the minority classes such as Breakfast, Lunch, Dinner, and snack. Hence, the proposed joint learning takes advantage of the complementary features to improve the performance of the minority classes.

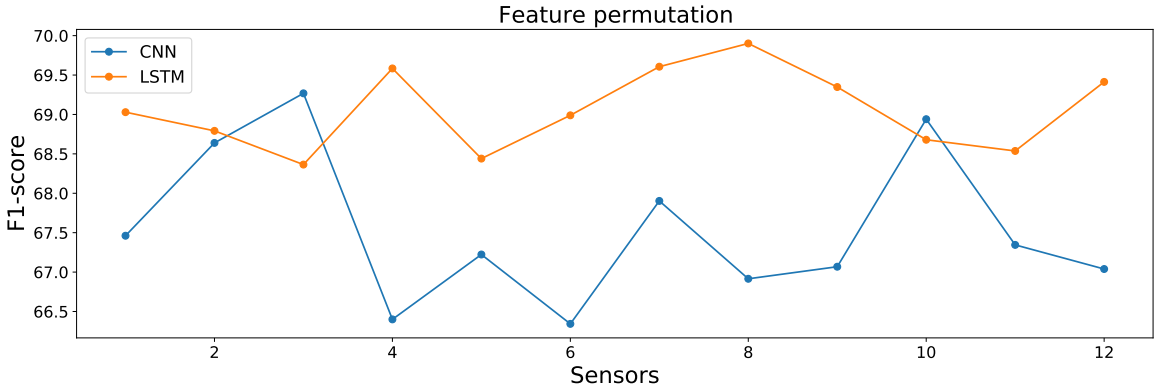


Fig. 4.4 Important features for LSTM and CNN of 12 sensors with $N = 12$ runs from Ordonez Home A

Table 4.8 Permutation feature importance of Ordonez Home A.

Sensors_Feature	**Feature Importance**				**F1-score of N Runs**	
Type_Location_Place of Sensors	**LSTM**	**Rank**	**1D ConvNet**	**Rank**	**Mean ± SD of LSTM**	**Mean ± SD of ConvNet**
PIR_Shower_Bathroom	13.13	7	12.43	8	69.02 ± 2.63	67.46 ± 1.30
PIR_Basin_Bathroom	13.36	5	11.26	10	68.79 ± 2.18	68.63 ±3.35
PIR_Cooktop_Kitchen	13.79	1	10.63	12	68.36 ± 2.71	69.26 ± 4.58
Magnetic_Main door_Entrance	12.57	10	13.50	2	69.58 ± 3.17	66.40 ± 2.93
Magnetic_Fridge_Kitchen	13.72	2	12.67	6	68.44 ± 2.96	67.22 ± 1.53
Magnetic_Cabinet_Bathroom	13.17	6	13.55	1	68.98 ± 5.04	66.34 ± 2.82
Magnetic_Cupboard_Kitchen	12.55	11	11.99	9	69.60 ± 2.14	67.90 ± 2.25
Electric_Microwave_Kitchen	12.25	12	12.98	3	69.90 ± 2.60	66.91 ± 1.94
Electric_Toaster_Kitchen	12.81	8	12.83	5	69.34 ± 2.43	67.06 ± 3.35
Pressure_Bed_Bedroom	13.48	4	10.96	11	68.67 ± 3.21	68.94 ± 4.04
Pressure_Seat_Living	13.62	3	12.55	7	68.53 ± 3.77	67.34 ± 4.10
Flush_Toilet_Bathroom	12.74	9	12.86	4	69.41 ± 2.10	67.03 ± 1.55

Table 4.9 Permutation feature importance of Ordonez Home B.

Sensors_Feature	**Feature Importance**				**F1-score of N Runs**	
Type_Location_Place of sensors	**LSTM**	Rank	**1D ConvNet**	Rank	**Mean ± SD of LSTM**	**Mean ± SD of ConvNet**
PIR_Shower_ Bathroom	9.56	2	7.62	9	62.79 ± 1.95	61.96 ±1.25
PIR_Basin_Bathroom	11.62	1	5.66	11	60.73 ± 1.80	63.92±1.61
PIR_Door_Kitchen	8.53	9	10.28	3	63.82 ± 1.70	59.30 ±1.26
PIR_Door_Bedroom	8.10	11	6.81	10	64.25 ± 1.48	62.77±1.35
PIR_Door_Living	8.65	7	9.39	7	63.70 ± 1.48	60.19 ± 1.63
Magnetic_ Maindoor_Entrance	8.12	10	9.60	6	64.23 ± 1.90	59.98± 1.41
Magnetic_Fridge_Kitchen	6.65	12	11.43	2	65.70 ± 1.68	58.15± 1.26
Magnetic_Cupboard_Kitchen	9.06	5	5.03	12	63.29 ± 1.78	64.55 ± 1.05
Electric_Microwave_Kitchen	8.71	6	12.41	1	63.64 ± 1.34	57.17± 1.90
Pressure_Bed_Bedroom	8.63	8	9.68	5	63.72 ± 1.24	59.90 ± 2.13
Pressure_Seat_Living	9.40	3	10.20	4	62.95 ± 2.18	59.38 ± 1.15
Flush_Toilet_Bathroom	9.33	4	9.35	8	63.02 ± 2.70	60.24±2.38

Algorithm 2: Compute Permutation Feature Importance (PFI)

1: **Input:** Train model M on original dataset D, label vector Y, error measure L(Y,$\hat{Y}$) (original datasets is the dataset without permutation ,$\hat{Y}$ is the predicted label)
2: Compute result score $RS_{original}(\hat{M}) = L(Y, \hat{M}(D))$ (e.g. F1-score)
3: **for** for each feature j from D **do**
4: **for** for each repetition $n = 1toN$ **do**
5: feature permutation on D to generate $\hat{D}_{n,j}$ (This removes the relationship between D_j and Y)
6: Compute result score $RS_{permuted,nj}(\hat{M}) = L(Y, \hat{M}(D_{n,j}))$ (e.g. F1-score on the permuted data)
7: **end for**
8: Compute feature importance i_j for feature j, $i_j = RS_{original}(\hat{M})$ - $\frac{1}{n}\sum_{n=1}^{N} RS_{permuted,nj}(\hat{M})$
9: **end for**
10: **Output:** feature importance i for all the features

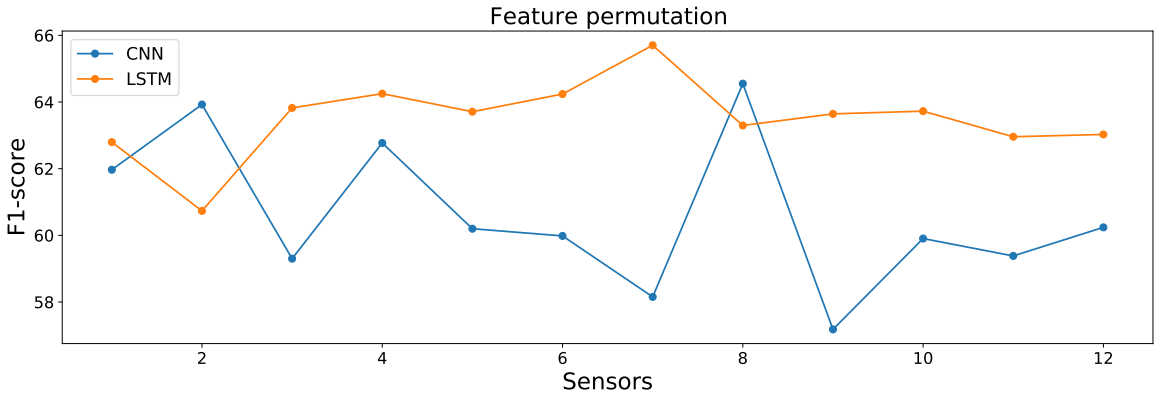


Fig. 4.5 Important features for LSTM and CNN of 12 sensors with $N = 12$ runs from Ordonez Home B

4.6 Discussions

This thesis proposes joint learning of the temporal model in an effort to improve the classification results for minority classes, as well as for the majority classes of HAR tasks in smart homes. The data are preprocessed using FTWs to segment the raw sensor data and build the input datasets. The proposed model is built upon LSTM and 1D CNN in parallel as one joint model. Extensive evaluations have been conducted in order to compare the proposed joint model with individual temporal models, i.e., LSTM, 1D CNN, and their hybridisations to show the superiority of the proposed model. The experimental results also confirm that our model has better performance for imbalanced data. The F-score results of the joint temporal model outperform 1D CNN, LSTM, and the hybrid learner by up to 4%.

The proposed joint learning outperforms the temporal models by 4% to 10%, which can be considered a substantial margin, since building accurate HAR systems is challenging due to the large diversity of activities since different sensors record human movements and inherently imbalanced the frequency of activities. Besides that, the proposed method is evaluated against state-of-the-art based on five standard benchmark datasets of activity recognition. An accepted threshold for a model to be considered to be successful is mainly based on the application scenario. Because activity recognition could be used to security perspective or elderly monitoring, the minimal 4% improvement is a substantial margin.

4.7 Conclusion

In this chapter, we propose two approaches to address imbalanced class problems in HAR systems based on algorithm-level and data-level. For the algorithm-level, we propose joint learning of the temporal models i.e. LSTM and 1D ConvNet in an effort to improve the classification results for minority classes, as well as for the majority classes of HAR tasks in smart homes. Extensive evaluations have been conducted in order to compare the proposed joint model with individual temporal models, i.e., LSTM, 1D CNN, and their hybridisations to show the superiority of the proposed model. The experimental results also confirm that our model has better performance for imbalanced data. The F-score results of the joint temporal model outperform 1D CNN, LSTM, and the hybrid learner by up to 4%.

For the data-level, to balance human activity datasets, the synthetic minority oversampling technique (SMOTE) [302] as a common method is mostly used. However, SMOTE in generating new instances fails to consider neighbouring instances of different classes which can append further noise and expand the overlapping of the classes. To overcome this problem, we propose iSMOTE that computes k-nearest neighbours of each generated instance to make sure the new instances are accurately annotated. Each new instance of the minority classes with its k-nearest neighbours must have the same class. For example, generated new instances of *Snack* activity must have k *Snack* activity as nearest neighbors.

Future work will investigate a newly proposed method for HAR to handle imbalanced human activity problems by integrating a data level and an algorithm level. Handling imbalanced class problems from the data level in addition to applying the proposed joint learning will further improve the recognition rate, particularly for extremely rare activities. One approach could be oversampling using the iSMOTE technique to only generate new samples from infrequent activities.

Chapter 5

Shared Representation for Human Activity Recognition

5.1 Introduction

Numerous studies have been conducted for HAR, majority of them are reporting the state-of-the-art performance and handling different problems of HAR systems. However, yet many challenges of HAR require further exploration to be properly tackled. One of the most major challenges of accurately developing HAR systems is to acquire considerable amount of annotated data which is not always available and expensive. Preparing a sufficient amount of annotated sensor data for HAR requires additional effort such as recording human activities with a synchronized video which requires manual effort and is very time-consuming.

Several studies are conducted and proposed different approaches for HAR systems with preserving reasonable accuracy to address the need for large annotated data. Mainly the approaches are working based on transfer learning which is capability to extend what has been learned in the training process in one domain to new and related domains. Transfer learning in machine learning is performed based on various approaches [313] which include multi-domain learning [314–316], self-supervised learning [317–319], domain adaptation [320], multi-task learning [321, 322], sharing of knowledge and representations [323].

In this thesis, shared representation is used to perform transfer learning and to reduce the need for rich quantity of well-curated human activity data through two different techniques: cross-domain learning and self-supervised learning to develop adaptable and robust HAR systems. Cross-domain learning is about transferring knowledge from a domain to another different but related domains to reduce the need for labelled data, reduce the training time and enhance the performance of HAR [313]. Besides, a self-supervised learning network is

proposed to learn a good representation of human activities from unlabeled data, enhance accuracy and substantially minimize the labelled data required for a downstream task, i.e. human activity classification. A brief introduction of the proposed methods is provided below.

1. Cross-domain activity recognition using shared representation is proposed to reduce the difference between the source and target domains and transfer knowledge across different but related human activity domains. We develop a multi-domain learning approach to jointly learn activity recognition based on different but related domains rather than learning each domain in isolation. The proposed method is a deep neural network that consists of two identical sub-networks. The sub-networks have the same configuration in terms of the number of layers, hyper-parameters and weights. Parameter updating is mirrored using a shared layer across both sub-networks. The proposed network consists of two 1D causal ConvNet layers for each of the sub-networks and is followed by a shared fully connected layer. The shared layer projects learned features to a common latent space to transfer knowledge between the source and target domains. A linear attention layer that increases focus on the important time steps is appended to the shared layer. The proposed method simultaneously processes two related domains: source and target domains. The benefit of our proposed method is firstly to reduce the need and effort for labelled data of the target domain. The proposed network uses the training data of the target domain with restricted size and the full training data of the source domain, yet provided better performance than using the full training data in a single domain setting. Secondly, the proposed multi-domain learning (MDL) network reduces the training time by rendering a generic model for related domains compared to fitting a model for each domain separately. Moreover, the proposed network can also be used to train on small datasets as a target domain by supporting the source domain. The effectiveness of transfer learning using the proposed cross-domain learning network is evaluated by the performance of the existing single-domain learning (SDL) models on the target datasets. The proposed MLD network outperforms the existing state-of-the-art SDL methods that are trained directly on the full target domain data. Moreover, the results demonstrate that the proposed MDL network can improve the performance of the source domain in addition to the performance of the target domain.

2. This thesis also proposes a self-supervised learning network for Human Activity Recognition (SHAR) that requires only unlabeled data to formulate a pre-training

model and learn a good representation of human activities for a downstream task i.e. activity recognition. The self-supervised pre-training model is then fine-tuned with a small amount of labelled data for supervised learning. The proposed self-supervised pre-training network renders human activity representations that are semantically meaningful and provides a good initialization for supervised fine-tuning. The pre-training unsupervised representation is then re-trained to perform well for HAR with less manually-labelled data than networks that are initialized randomly.

Our proposed network also handles imbalanced datasets for representation learning. To explore the effects of the imbalanced class problem on the proposed self-supervised network, we show experiments and study the outcomes of the proposed network on imbalanced and balanced human activities. To balance human activity datasets, we use over-sampled using synthetic minority over-sampling technique (SMOTE) [302]. However, SMOTE in generating new instances fails to consider neighbouring instances of different classes which can append further noise and expand the overlapping of the classes. To overcome this problem, border limited link SMOTE (BLL-SMOTE) is proposed, to avoid the misgenerated new samples [304]. BLL-SMOTE focuses on the distances between the newly generated samples with their k-nearest neighbours and the nearest sample in the dataset. To mitigate the distance calculations of the BLL-SMOTE and improve the oversampling process, we propose an improved SMOTE (iSMOTE) that computes k-nearest neighbours of each generated instance to make sure the new instances are accurately annotated. Each new instance of the minority classes with its k-nearest neighbours must have the same class. For example, generated new instances of *Shower* activity must have k *Shower* activity as nearest neighbors.

The proposed approaches to reduce the need for large annotated data and enhance HAR systems using shared representation are further delineated in the following sections.

5.2 Cross-Domain Activity Recognition Using Shared Representation

The proposed MDL network aims to improve performance of HAR systems, decrease the learning time and reduce the number of learned models. The distinctive characteristics of the proposed network are: (1) the proposed network simultaneously processes different but related datasets for HAR and provides a recognition performance for each of the datasets; (2) the proposed network preserves the ordering of temporal sequential input data and avoids

information flow from future time steps to past time steps using causal convolution; (3) the network uses a linear attention mechanism to focus more on most pertinent information in the sequence input. The following subsections provide details about the proposed network.

5.2.1 MDL Network

One of the major reasons for the misrecognition of human activities is the unavailability of complementary features which yield semantic information about human activities. In different but related domains, the complementary features are present with various intensities and scales [361]. We propose an MDL network to learn generic and robust feature representations from multiple domains that outperform state-of-the-arts of SDL on all of the domains by large margins. Suppose X and Y are the features and the label spaces respectively. A joint probability distribution P(X,Y), represents a defined domain on $X \times Y$. P_k(X) and P_k(X,Y) denote the marginal distribution and joint distribution of the datasets in the k-th domain [362]. Every single dataset is associated with a sample D_k=$\{ x_i, y_i \}_{i=1}^{L_k}$ where L_k is the sample size of the k-th domain. Given N related domains $P_{d=1}^{N}$(X, Y) and their corresponding datasets D_k=$\{ x_i, y_i \}_{i=1}^{L_k}$ from multiple human activity domains, the goal of the proposed MDL model is to learn a robust and multi-branch model $f : X_k \rightarrow Y_k$, k = {1, 2, . . . , N} and to perform HAR on all the domains in parallel. The proposed model is able to learn features from multiple domains using a shared representation and inference of all the domains. Fusing multi-domain features enable the proposed network to learn better features from multiple datasets for HAR rather than learning each domain in isolation. The shared representation of MDL aims to project learned features to shared feature space and to transfer knowledge between the domains. The proposed network shares data between two domains: source and target domains. Besides, the proposed network can be extended for sharing knowledge among many domains with different but related domain data.

5.2.2 Architecture of MDL Network

The structure of the proposed network consists of two identical sub-networks which accept distinct inputs but are joined by a shared representation. Each of the sub-networks consists of two layers of causal 1D CNN followed by a fully connected shared layer. Then the linear attention mechanism is appended to the shared layer followed by a domain-specific layer for each of the domains. The feature maps are fed into the softmax output layer for HAR. Figure 5.1 shows the structure of the proposed network. Causal 1D CNN processes temporal sequential inputs independently and performs operations in parallel with avoiding

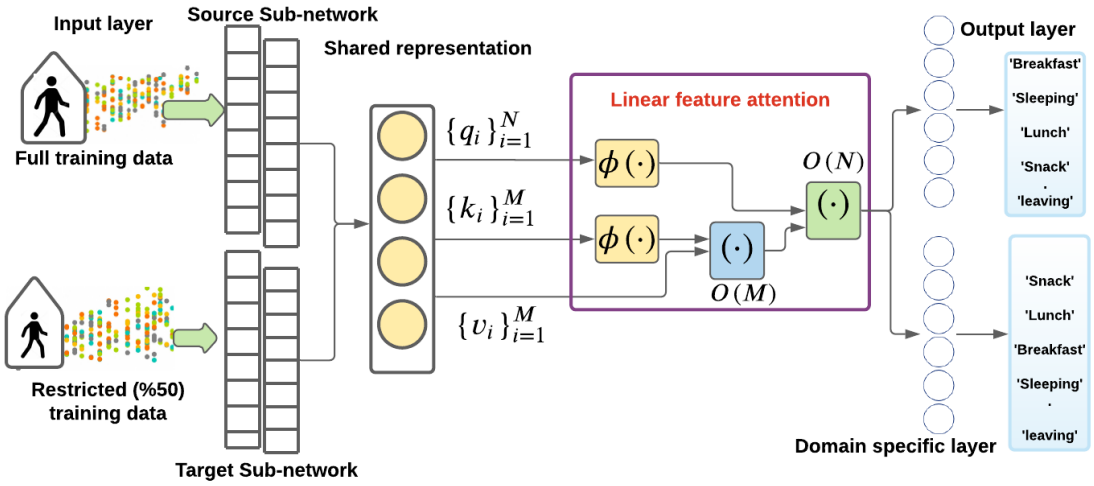


Fig. 5.1 Cross-domain learning using a shared representation to transfer knowledge among different but related domains

information leakage from future time steps to past time steps n the sequence input data [12]. The number of the filters of the convolutional layers is 64 for each of the sub-networks and the kernel size of the convolutional layers which specify the length of the convolution window is equal to 3. The stride of the convolutional filters that control how the filter convolves around the temporal sequential inputs is equal to 1 in order to shift the convolutions one unit at a time.

The advantage of the proposed network is first to reduce the need and effort for labelled data. Thereby the network uses the full training data of the source domain and a restricted size of the training data of the target domain. Yet the proposed network can achieve better performance than using the full training data of the target domain in a single domain setting. Secondly, the proposed method reduces the training time by training a generic model for two domains compared to fitting a model for each domain separately. The proposed network enriches the diversity of the training data due to the domain discrepancies using the shared representation. The diversity of the training data from the proposed network ensures that the training process can provide more discriminative features to the model. Parameter updating is mirrored using a shared representation between the sub-networks. The proposed network improves recognition performance for both domains by transferring knowledge across the domains. The effectiveness of transfer learning using the proposed approach is evaluated by the performance of the SDL model on the target domain using full training data.

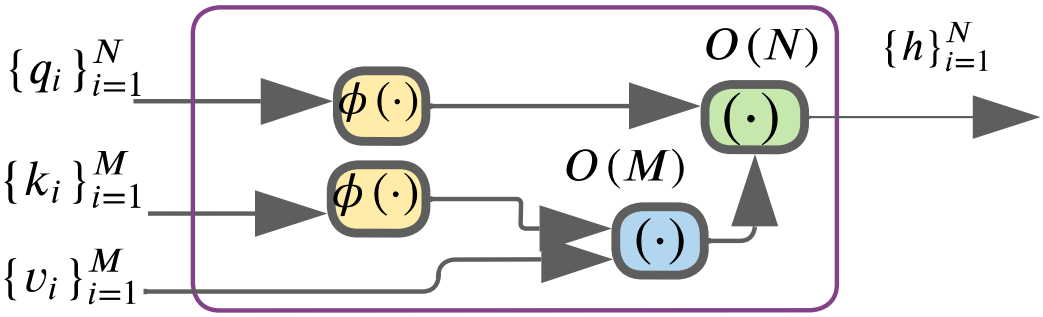


Fig. 5.2 Linear attention mechanism to focus more on the important time steps

5.2.3 Linear Attention Mechanism

The self-attention technique scales quadratically with the length of the input temporal data and adds more weight parameters to methods that increase learning time and requires more memory. To address this problem, random feature attention (RFA) is proposed as a linear attention mechanism [363] that uses random feature methods to approximate the softmax function. The linear attention mechanism has three learned matrix: queries $Q \in R^{N \times D_k}$, keys $K \in R^{M \times D_k}$, and values $V \in R^{M \times D_v}$, where N and M are the lengths of the queries and keys (or values), D_k and D_v are the dimensions of keys (or queries) and values as shown in Figure 5.2. RFA disentangle the softmax (QK^T) into QK^T and compute QK^TV in reverse order $Q(K^TV)$ leading to a linear time and space attention. RFA approximate or replace the unnormalized attention matrix exp (QK^T) with $\phi(Q)\ \phi(K)$ where ϕ is a feature map that is applied in row-wise manner. Therefore, RFA linearly computes unnormalized attention matrix by $\phi(Q)(\phi(K)^TV))$, as illustrated in Figure 5.2. The vector form of the self-attention mechanism is shown in Equation 5.1. RFA vector form is shown in Equation 5.2. RFA as a linear attention mechanism is adopted for our proposed network to increase the capability of our proposed network in extracting fine-grained features of human activities.

$$Att(Q_t, \{K_i\}, \{V_i\}) = \sum_i \frac{exp(Q_t \cdot K_i / \mathcal{T})}{\sum_j exp(Q_t \cdot K_j / \mathcal{T})} V_i^{\top}. \tag{5.1}$$

$$RFA(Q_t, \{K_i\}, \{V_i\}) = \frac{\phi(Q_t)^{\top} \sum_i \phi(K_i) \otimes V_i}{\phi(Q_t) \cdot \sum_j \phi(K_j)}. \tag{5.2}$$

5.3 Self-supervised Learning Based on Datasets with Imbalanced Classes

inspired by the success of the self-supervised methods in HAR and other domains, we aim to explore and enhance HAR from sensors data of intelligent environments and smart wearable devices. Our proposed method handles the imbalanced dataset problem for representation learning and reduces the need for labelled data in the fine-tunning. To address this problem, we firstly apply the iSMOTE technique to handle imbalanced datasets by generating and correctly labelling synthetic samples for the activities with infrequent instances. Secondly, we apply our proposed random masking technique on the unlabeled balanced datasets to remove identity mappings and to build corrupted or masked datasets for a downstream task. Finally, the proposed network trains on the corrupted and unlabeled balanced data to learn representations of human activities from transformed sensors signal data. Then the learned representations are fine-tuned with a small amount of labelled data for supervised learning. We demonstrate the ability of handle imbalanced datasets for improving representation learning compared to applying our proposed self-supervised network on the imbalanced datasets. The proposed network improves the performance of the activities with infrequent samples when the imbalanced datasets are handled. Figure 5.3 shows the complete architecture of the proposed network.

In the following sections, the proposed self-supervised network for HAR (SHAR) is presented. First, we present the proposed iSMOTE technique for oversampling minority class samples to feed a balanced unlabelled data into the proposed SHAR network to render a good representation. Second, the random masking technique to mask the input data and to remove identity mappings are described. We also provided a detailed description of the proposed SHAR network structure.

5.3.1 Random Masking

Random masking (RM) as a signal transformation technique is proposed in this thesis to remove identity mappings and build generic semantic representations for a downstream task. The RM technique corrupts the data based on a masking probability that transforms the sensor signals and generates new sensor signals. Algorithm 3 shows the proposed random masking technique to corrupt the data for self-supervised learning. The proposed RM technique is compared with many signal transformation methods that include: permutation, rotation, scaling, magnitude-warping, jittering and time-warping [364].

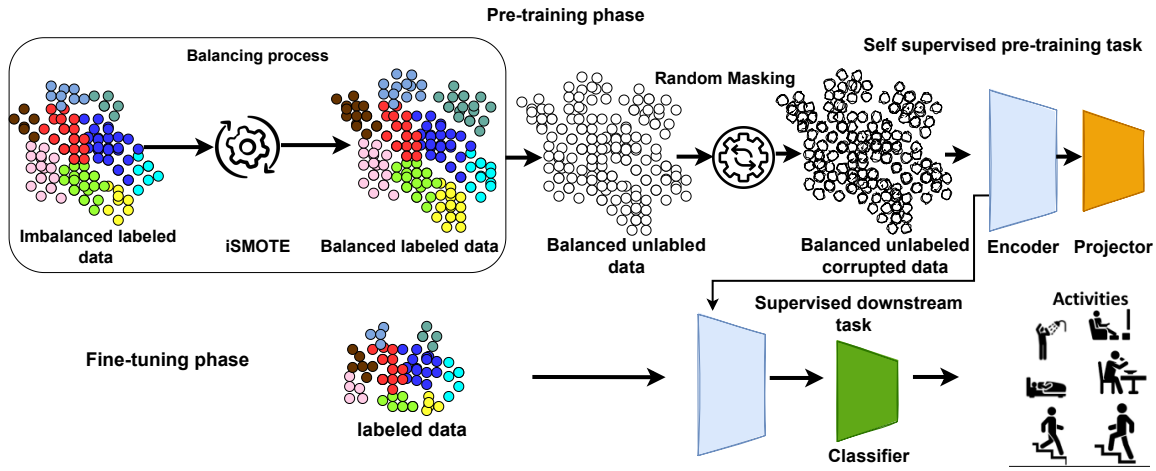


Fig. 5.3 Proposed self-supervised network for human activity recognition

Algorithm 3: Random Mask

```
RandomeMasking(x,prob):
rand_number ← randomint(10, x.size)
if rand_number > 10 ∗ prob then
    Mask ← 1
else
    Mask ← 0
end if
return x × Mask
```

5.3.2 Self-Supervised Network

Since acquiring a sufficient amount of manually labelled data for supervised learning is cumbersome, self-supervised learning has been very effectively employed in the representation learning (pre-training phase). Self-supervised learning provides an encouraging learning model since it allows learning from a considerable amount of readily accessible unlabelled data. In this thesis, SHAR is proposed to enhance the result score and minimise the need for the rich quantity of the annotated data. The proposed network has pre-training and fine-tuning phases which are described as follows.

5.3.3 Pre-training Phase

The pre-training phase trains an encoder followed by a projector to learn a meaningful representation of the underlying data. The encoder contains two 1D causal ConvNet followed by the lambda network layer [365]. The 1D ConvNet-based networks have shown

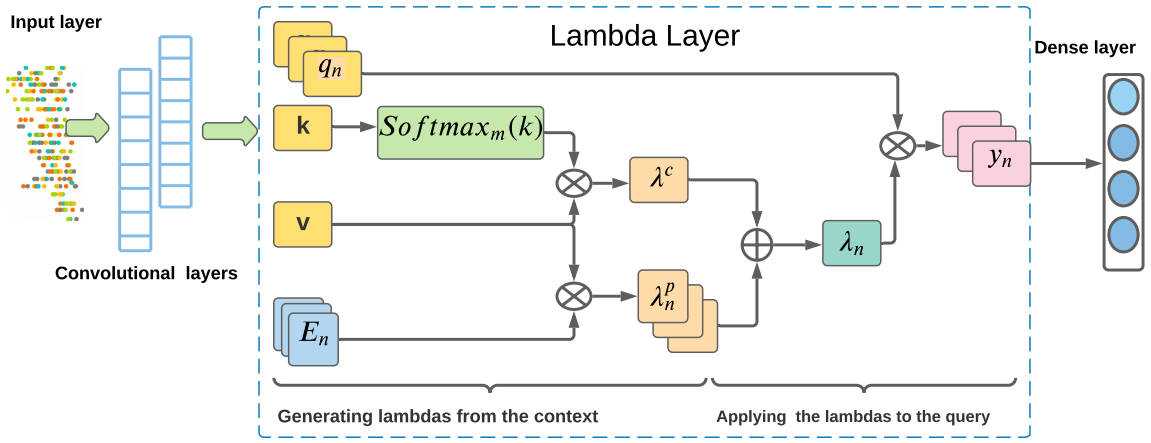


Fig. 5.4 Complete overview of the encoder

encouraging performance for HAR systems due to the capability of the 1D ConvNet in capturing rich relational features and local dependency from the input sensors data [19, 366]. Furthermore, 1D ConvNet layers are able to learn representations of hierarchical human activities that improve the accuracy of the HAR systems. Causal convolution is used for the encoder to preserve the ordering of temporal data that is significant for HAR systems [210]. The lambda layer [365] is proposed as an efficient self-attention [342] alternative to focus more on both position and content-based interactions that outperforms the self-attention mechanism. Moreover, the lambda layer as a linear function that reduces complexity of self-attention which is commonly used in HAR systems [14, 351, 367]. The lambda layer as self-attention mechanism composed of these learned components: query Q, key $K \in \mathbb{R}^{n \times d_k}$, and values $V \in \mathbb{R}^{n \times d_v}$. These components are generated by a linear projection of the lambda layer input. Different from the self-attention, the softmax function is only implemented to the $\sigma(K)$ Key matrix over the context positions. The K and V information are encapsulated in the content lambda λ^c to convert queries according to the context content. The positional lambda λ_n^p that encapsulates the V matrix and learnable parameter of positional embedding $E_n \in \mathbb{R}^{n \times d_k}$ transforms the query q_n based on their relative positions. The contextual lambda function $\lambda_n \in \mathbb{R}^{d_k \times d_v}$ is then multiplied to each query as shown in Equation 5.3

$$\lambda_n = \sigma(K)^T V + E_n^T V = \lambda^c + \lambda_n^p \tag{5.3}$$

The lambda layer outputs the matrix $Y \in \mathbb{R}^{n \times d_v}$ in which any element y_n of the matrix is the application of each contextual lambda to its query as shown in Equation 5.4.

$$y_n = \lambda_n^T q_n = (\lambda^c + \lambda_n^p)^T q_n \tag{5.4}$$

A fully connected (i.e dense) layer is appended on the lambda layer. The details of the encoder are shown in Figure 5.4. The learning layers are fed to the batch normalization technique [265] to normalize the feature map, make the proposed network faster and further stable during learning. Also, the dropout regularization technique [350] is utilized to avoid over-fitting. The projector adds a fully connected layer to the encoder. The projector outputs a vector with a size the same as the dimension of the unlabelled and balanced input datasets. The projector is discarded for the fine-tuning phase. The pre-training encoder is only using unlabelled data to generate good human activity representations that are semantically meaningful and provide a good initialization for the supervised fine-tuning.

5.3.4 Fine-tuning Phase

The pre-trained encoder is fine-tuned and followed by a classifier for the supervised downstream task (i.e. activity recognition). Labelled data are fed into the fine-tuning phase to adjust the pre-trained encoder and properly distinguish human activities. The pre-trained encoder leverages unlabelled data in the fine-tuning phase and transfers acquired knowledge to improve the supervised activity recognition and minimize the need for considerable amount of labelled data. The proposed SHAR requires a small amount of labelled data to outperform state-of-the-art methods.

5.4 Experimental Setup for the Proposed MDL Network

5.4.1 Datasets

Four HAR datasets based on smart homes are used to perform experiments of MDL. Table 5.1 shows the details of the Ordonez smart homes A and B and Kasteren homes A and B. Furthermore, two HAR datasets based wearable sensor are also used to conduct experiments. The details of these dataasets are all described in Chapter 2.4.1.

Table 5.1 Details of the datasets

Smart Homes	Setting	Activities	Rooms	Sensors	Duration (days)
Ordonez A	Home	10	4	12	14
Ordonez B	Home	11	5	12	21
Kastern A	Apartment	10	3	14	25
Kastern B	Apartment	13	2	23	14

Table 5.2 Number of activities in wearable wireless identification and sensing datasets

Activity	RoomSet1	RoomSet2
Ambulating	1956	335
Lying	30983	20537
Sit on bed	15162	1244
Sit on chair	4381	530

5.4.2 Implementation Details of the MDL Network

Hyperparameter tuning is a powerful optimization procedure to reach maximum effectiveness and to render accurate deep learning models. After conducting an extensive trial and error process, 0.0001 as the learning rate, 128 as the batch size with a 25% dropout rate are used for the proposed network to converge. Early stopping as a regularization technique is used to determine the number of epochs and to prevent overfitting by stopping the training when the validation error of the proposed network starts increasing. The 25% dropout rate as a regularization technique after each learning layer is used to further avoid overfitting [350]. Batch normalization that normalizes the input data across the batches is used after each learning layer [265] to make deep learning models faster and more stable during training.

5.5 Results of MDL Network

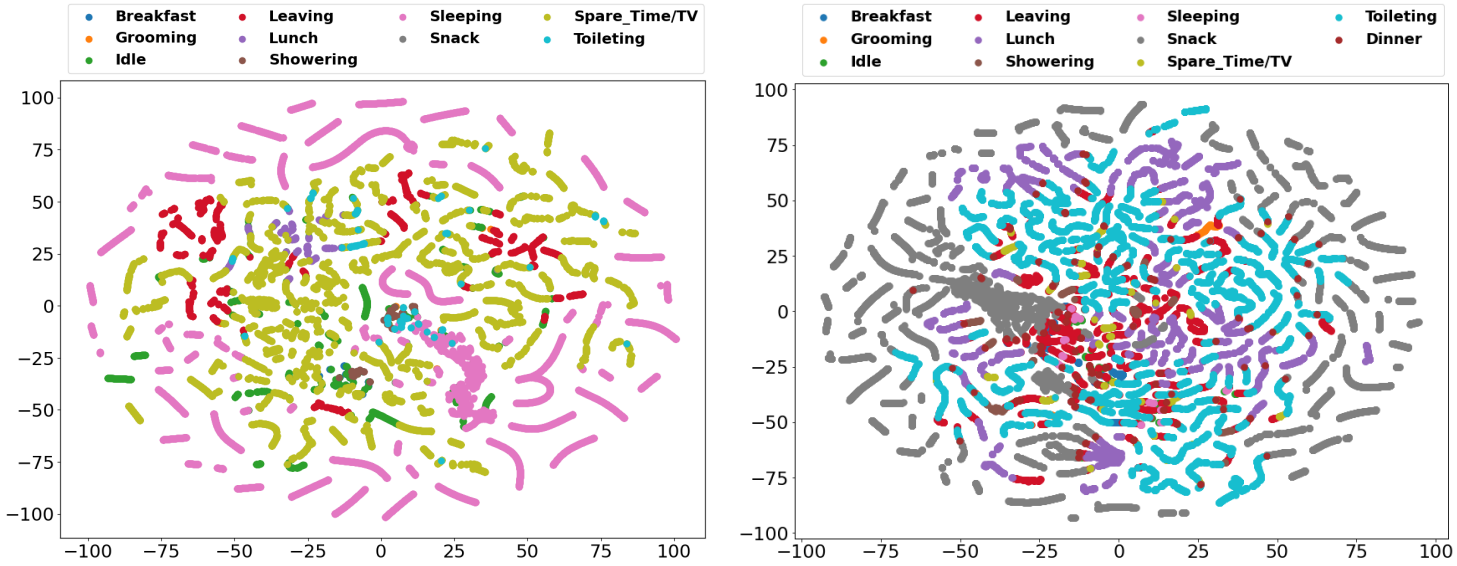
The experimental results and findings of the proposed MDL network are exposed and discussed. In this section, a comparative analysis is presented for the proposed MDL network and the state-of-the-art methods. The effectiveness of the proposed MDL network is evaluated by the existing SDL state-of-the-art methods. The proposed network processes full training data of a source domain while using only 50% training data of a target domain. The proposed MDL network using 50% training data of a target domain can outperform the SDL method on the same datasets with all the training data. Hence the proposed MDL network can reduce the need for labelled data using cross-domain learning by jointly training several domains. The results of the source domain in addition to the target domain are also improved compared to the existing state-of-the-art methods. The results of the target domains based on the proposed MDL network are highlighted in the tables to be easily compared with the results of the current methods. An ablation study is performed to show the performance of the proposed network where the training target domain data is 50%, 25% or 10%.

5.5.1 Results of Ordonez Smart Home Datasets

Tables 5.3 shows results of the proposed MDL network and existing methods from Ordonez home A and B datasets. The results of the proposed MDL network is compared with the existing SDL state-of-the-art methods. The proposed MDL network have achieved better results compared to the results obtained by the state-of-the-art methods. The results of the target smart home domains with 50% of their training data based on the proposed MDL network are significantly improved compared to their results obtained by the SDL approach with full training data. In addition to the improvement of the target domains, the results of the source domains are also improved. This is revealed that the proposed MDL network transfers knowledge between both the source and target human activity domains using the shared representation. The shared layer exposes different features from source and target domains to render strong mutual complementary features. Complementarity in the proposed MDL using a shared representation boosts the recognition performance. This is because the model in the MDL approach delivers distinctive features from both source and target human activity domains to enrich the training and each model refines the earlier layers of the other model and reduces their weaknesses. The joint optimization of the MDL approach enhances the functionality of the proposed network to gain more insight into the source and target domain features to increase the recognition result score. Figure 5.5 shows the t-SNE map for human activity recognition from Ordonez smart home A and B datasets, while Figure 5.6 shows a t-SNE map for human activity recognition from the proposed MDL network without the last softmax layer. The plot shows how the proposed network properly distinguishes the activities.

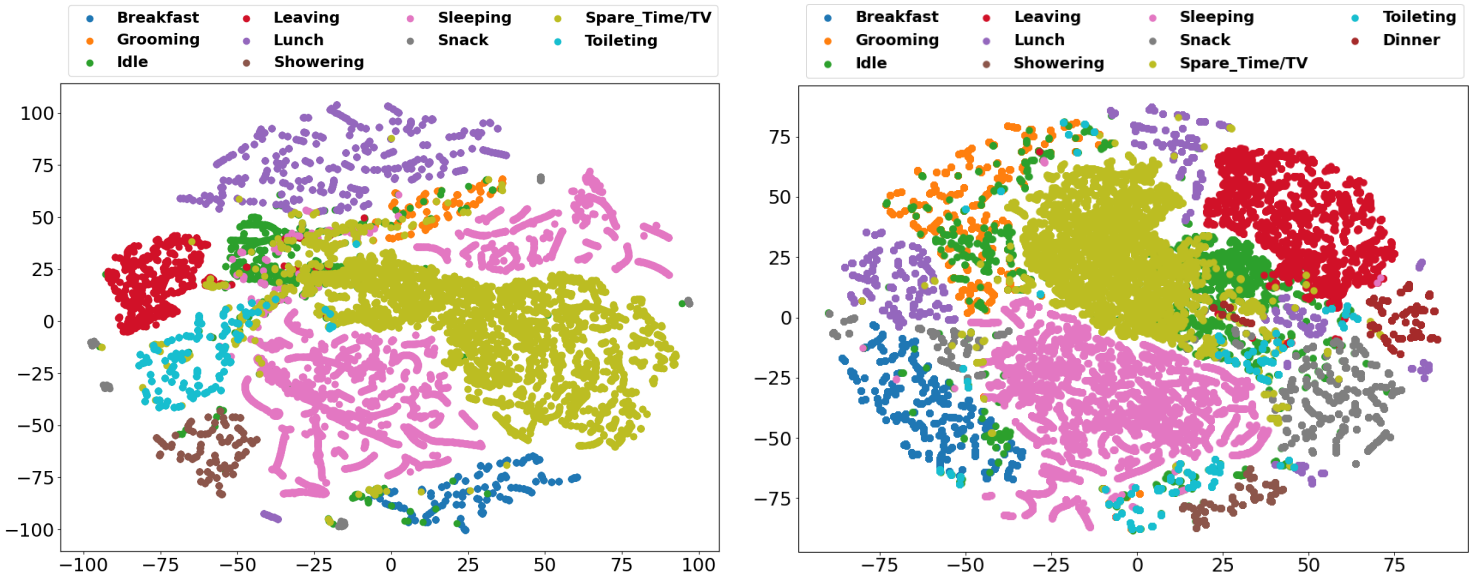
5.5.2 Results of Kasteren Datasets

Table 5.4 shows the results of the proposed MDL network compared to existing methods from the Kasteren smart home A and B datasets. The results of the proposed MDL network is also compared with the existing state-of-the-art methods. The common activities from source and target domains are shaded. The results of the source domains using the full training data and target domains using 50% of their training data based on the proposed MDL network are improved compared with the results obtained by the existing methods.



(a) Ordonez Home A (b) Ordonez Home B

Fig. 5.5 t-SNE map for human activity from input datasets



(a) Ordonez Home A (b) Ordonez Home B

Fig. 5.6 t-SNE map for human activity from proposed MDL

Table 5.3 F1-score results of the proposed network from Ordonez home A and B datasets: source domain uses all its training data, target domain uses 50% of its training data

Activity	Proposed MDL network				Existing SDL methods			
					Home A		Home B	
	Home A source domain	Home B target domain	Home B source domain	Home A target domain	Deep ConvLSTM	HAR+ Attention	Deep ConvLSTM	HAR+ Attention
Sleeping	99.77	**98.06**	99.71	**98.74**	97.53	97.11	96.37	95.42
Breakfast	88.59	**83.76**	85.11	**86.96**	83.11	84.51	74.87	75.39
Lunch	98.42	**99.23**	99.72	**97.09**	95.44	94.39	95.21	95.31
Grooming	85.00	**90.33**	92.98	**82.14**	75.32	80.00	85.33	87.91
Spare Time	99.16	**84.29**	86.15	**98.65**	96.83	97.21	78.21	79.32
Leaving	98.05	**96.20**	98.00	**96.36**	95.29	95.51	89.79	92.09
Snack	86.59	**80.91**	83.32	**85.31**	70.74	82.63	76.16	76.31
Showering	95.54	**84.11**	86.19	**97.34**	80.65	86.89	79.43	79.12
Toileting	83.85	**87.63**	89.57	**81.29**	69.89	77.25	83.56	83.24
Dinner		**89.11**	91.70				86.19	86.49
Average	**92.77**	**89.36**	**91.24**	**91.54**	84.97	88.55	84.51	85.06

Table 5.4 F1-score results of the proposed network from Kasteren home A and B datasets: source domain uses all its training data, target domain uses 50% of its training data, common activities in both domains are shaded

Activity	Proposed MDL network				Existing SDL methods			
					Home A		Home B	
	Home A source domain	Home B target domain	Home B source domain	Home A target domain	Deep ConvLSTM	HAR+ Attention	Deep ConvLSTM	HAR+ Attention
Get_Snack	75.29			**71.73**	57.22	58.71		
Prepare_breakfast	86.46			**83.72**	76.97	79.54		
Brush_teeth	60.35	**58.98**	60.52	**56.96**	43.59	52.22	42.89	47.82
Get_drink	70.87	**62.86**	65.93	**67.61**	59.33	59.54	44.15	44.75
Go_to_bed	89.06	**99.62**	99.81	**87.42**	80.16	81.76	96.32	94.48
Leave_house	88.95	**96.97**	97.19	**85.98**	80.02	82.19	92.98	93.21
Prepare_Dinner	97.91	**97.63**	99.51	**95.92**	89.56	91.87	96.21	95.31
Take_shower	92.75	**84.92**	86.49	**90.26**	85.13	89.23	81.95	82.13
Use_toilet	75.79	**65.03**	69.06	**72.82**	67.82	69.34	56.13	54.19
Eat_brunch		**94.1**	96.21				90.93	91.11
Eat_dinner		**88.06**	91.68				86.79	86.29
Prepare_brunch		**85.26**	89.85				85.62	84.29
Get_dressed		**48.32**	50.54				31.79	41.11
Wash_dishes		**81.63**	84.92				75.38	77.25
Average	**81.93**	**80.28**	**82.45**	**79.50**	71.09	73.82	73.42	74.32

5.5.3 Results of Wearable Sensors Datasets

Tables 5.5 shows the results of the proposed MDL network and existing methods from the wearable sensors datasets. The results of source and target domains are improved using the proposed MDL network compared to the existing state-of-the-art methods. The results indicate that the proposed MDL network shares knowledge between source and target domains using the shared representation.

Table 5.5 F1-score results of the proposed network from Wearable sensors Roomset 1 and Roomset 2: source domain uses all its training data, target domain uses 50% of its training data

	Proposed MDL network				**Existing SDL methods**			
					Home A		**Home B**	
Activity	Home A source domain	Home B target domain	Home B source domain	Home A target domain	Deep ConvLSTM	HAR+ Attention	Deep ConvLSTM	HAR+ Attention
Ambulating	98.79	**93.90**	95.23	**97.73**	93.67	95.22	87.11	89.79
Sit on bed	99.30	**98.61**	99.35	**99.46**	95.31	96.25	97.52	99.79
Lying	98.86	**95.70**	96.37	**97.71**	94.12	95.21	89.32	94.85
Sit on chair	75.31	**74.31**	76.89	**73.45**	60.52	70.11	68.87	69.87
Average	**93.06**	**90.63**	**91.96**	**92.08**	85.90	89.19	83.40	85.70

Table 5.6 F1-score results of the proposed network where target domain uses 50%, 25% and 10% of its training data

	50% target data		25% target data		10% target data		50% target data		25% target data		10% target data	
Datasets	source domain	target domain	source domain	target domain	source domain	target domain	source domain	target domain	source domain	target domain	source domain	target domain
	Home A	Home B	Home A	Home B	Home A	Home B	Home B	Home A	Home B	Home A	Home B	Home A
Ordonez	92.77	89.36	88.67	83.21	87.81	80.11	91.24	91.54	89.54	84.85	85.13	80.04
Kasteren	81.93	80.28	80.45	77.50	77.12	73.53	82.11	79.98	79.45	74.11	77.87	69.54
	Roomset 1	Roomset 2	Roomset 1	Roomset 2	Roomset 1	Roomset 2	Roomset 2	Roomset 1	Roomset 2	Roomset 1	Roomset 2	Roomset 1
Wearable sensors	93.06	90.63	90.28	85.21	86.88	82.23	91.96	92.08	89.15	86.17	86.76	80.18

Table 5.7 F1-score results of the proposed network and state-of-the-art methods with 25% or 10% of training target domain data

Datasets	State-of-the-art methods				Proposed network							
	25% training data		10% training data		25% target data				10% target data			
	Deep ConvLSTM	HAR+ Attention	Deep ConvLSTM	HAR+ Attention	source domain	target domain	source domain	target domain	source domain	target domain	source domain	target domain
	Home A		Home B		Home A	Home B	Home B	Home A	Home B	Home A	Home B	Home A
Ordonez	73.59	71.25	66.19	62.86	88.67	83.21	89.54	84.85	87.81	80.11	85.13	80.04
Kasteren	69.29	70.93	61.11	58.22	80.45	77.50	79.45	74.11	77.12	73.53	77.87	69.54
	Roomset 1		Roomset 2		Roomset 1	Roomset 2	Roomset 2	Roomset 1	Roomset 2	Roomset 1	Roomset 2	Roomset 1
Wearable sensors	86.25	82.00	76.19	71.54	90.28	85.21	89.15	86.17	86.88	82.23	86.76	80.18

5.5.4 Ablation study

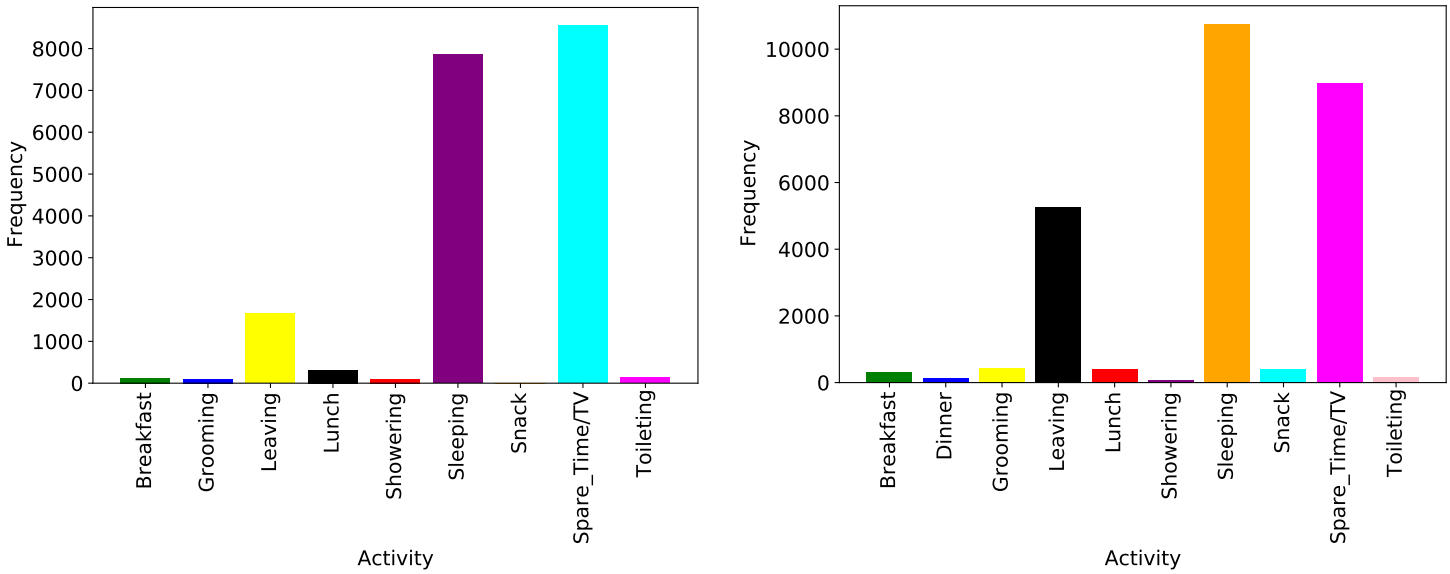
An ablation study is performed to show the performance of the proposed network where the training data of the target domains is 50%, 25%, or 10%. The results show that the 50% of the training data is the best configuration and renders better results as shown in Table 5.6. Moreover, the performance of the proposed network based on the 25% or 10% of training target domain data is compared with the state-of-the-art methods when the training data is 25% or 10%. Table 5.7 shows the performances of the proposed network are yet better compared to the performances achieved by the existing state-of-the-art methods. This is due to the proposed network uses a shared representation to transfer knowledge between the source and target domains.

5.6 Experimental Setup for the proposed SHAR network

In the experiments section, we describe the sensor-based datasets that are used to evaluate our proposed method. Furthermore, the prepossessing of the datasets is explained. We also provided the hyperparameters of the proposed network.

5.6.1 Datasets

To evaluate the SHAR network, 12 public datasets are used to evaluate the above proposed networks for HAR systems. Five smart home sensors collected datasets and seven wearable



(a) Ordonez smart home A

(b) Ordonez smart home B

Fig. 5.7 Frequency of activities from Ordonez Smart homes

sensor collected datasets which described in Chapter 2.4.1 on Page 20 are used in the evaluation process. The proposed networks are evaluated on the datasets with multiple human activities ranging from 4 to 13 from collected sensors data. The five smart home datasets are collected from Ordonez home A and B [111] and Kasteren home A, B and C datasets. Table 5.8 shows details of these five smart home datasets with respect to the residents, sensors, and the number of activities. Besides Tables 5.9 and 5.10 show the frequency of the human activities from Ordonez homes A and B as well as Kasteren homes A, B and C datasets after preprocessing and segmentation using FTWs which is delineate in Chapter 2.5.2. Figure 5.7 shows the distribution of the activities from the Ordonez smart homes A and B datasets. The frequency of activities from these datasets are highly imbalanced.

The wearable sensor datasets: UCI-HAR dataset, Roomset1 Roomset2 datasets, HHAR dataste, UniMiB SHAR dataset, MotionSense dataset, and WISDM dataset are also used in the evaluation process. The distribution of UCI-HAR wearable dataset is shown in Table 5.11. The distribution of human activities in the RoomSet1 and RoomSet2 wearable datasets is shown in Table 5.12. Table 5.13 shows the datasets that are particularly used to compare the proposed SHAR network with other self-supervised methods. Figure 5.8 shows the distribution of the activities in the UniMiB SHAR dataset which are highly imbalanced.

Table 5.8 Information about five smart home environment datasets

	Ordonez Home A -	Ordonez Home B -	Kastern Home A -	Kastern Home B -	Kastern Home C
Setting	Home	Home	Apartment	Apartment	House
Gender	-	-	Male	Male	Male
Activities	10	11	10	13	16
Age	-	-	26	28	57
Rooms	4	5	3	2	6
Sensors	12	12	14	23	21
Duration	14 days	21 days	25 days	14 days	19 days

Table 5.9 Distribution of performed human daily life activities in the Ordonez smart homes

Human Activity	Smart Home A	Smart Home B
Leaving	1,664	5,268
Snack	6	408
Grooming	98	427
Breakfast	120	309
Toileting	138	167
Showering	96	75
Lunch	315	395
Spare Time/ TV	8,555	8,984
Dinner	-	120
Sleeping	7,866	10,763

Table 5.13 shows that this dataset is collected from nine activities of 30 participants. Figure 5.9a presents the distribution of the human activities in the WISDM dataset and the samples of the human activities are imbalanced. Table 5.13 shows that this dataset is collected from six activities of 36 participants. Figure 5.9b shows the frequency the activities in the MotionSense dataset which are highly imbalanced. Table 5.13 shows that this dataset is collected from six activities of 24.

5.6.2 Implementation Details of the SHAR Network

Hyper-parameters of the proposed SHAR method are selected by a series of trial and error experiments. A learning rate of 0.002, and a batch size of 128, a dropout rate [350] of 20% after each learning layer are selected to the proposed network. Early stopping is conducted as one of the regularization mechanisms to avoid overfitting by halting the training of the network once the performance of the proposed SHAR network on the validation set stops improving. Batch normalization is used in the proposed network as a normalization method

Table 5.10 Distribution of human physical daily activities in the Kasteren smart homes

Activities	Home C	Activities	Home B	Activities	Home A
Get_dressed	70	Eat_brunch	132	Go_to_bed	11,599
prepare_Dinner	300	Prepare_brunch	82	Get_snack	24
Prepare_Breakfast	78	Prepare_dinner	87	Prepare_Breakfast	59
Eating	345	Brush_teeth	25	Take_shower	221
Get_snack	8	Eat_dinner	46	Leave_house	19,693
Leave_house	11,915	Go_to_bed	6,050	Prepare_Dinner	325
Prepare_Lunch	58	Wash_dishes	25	Use_toilet	154
Go_to_bed	7,395	Get_a_drink	6	Brush_teeth	21
Take_shower	184	Leaving_the_house	12,223	Get_drink	21
Get_drink	20	Use_toilet	39		
Use_toilet_upstairs	35	Take_shower	109		
Take_medication	6	Get_dressed	27		
Shave	57				
Brush_teeth	57				
Use_toilet_downstairs	75				

Table 5.11 Human activities in the UCI HAR dataset

Human Activity	Training samples	Testing samples
Walking_downstairs	986	420
Standing	1,374	532
Walking_upstairs	1,073	471
Laying	1,407	537
Sitting	1,286	491
Walking	1,226	496

Table 5.12 Distribution of human activities in the RoomSet1 and RoomSet2 wearable datasets

Human Activity	RoomSet1	RoomSet2
Sit on chair	4,381	530
Lying	30,983	20,537
Sit on bed	15,162	1,244
Ambulating	1,956	335

to normalize the feature maps during the learning process across the batches [265] to make the proposed classification SHAR network quicker and further stable during learning.

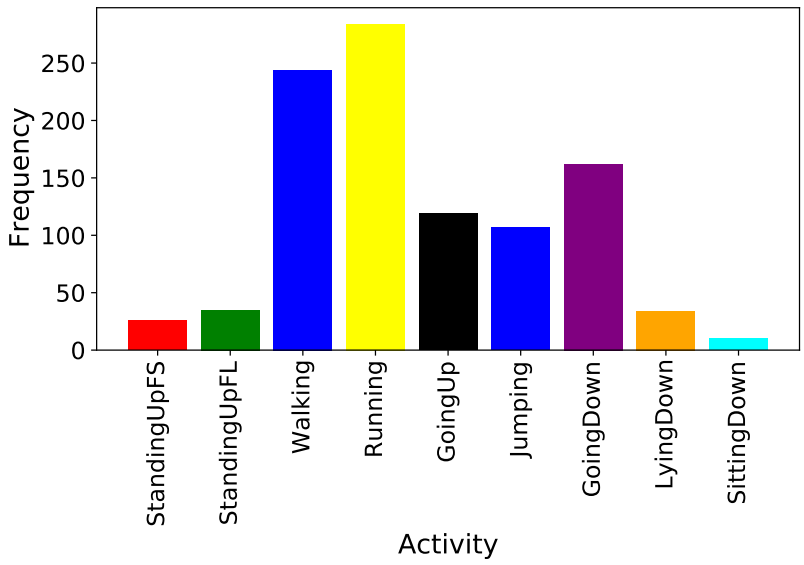


Fig. 5.8 Frequency of activities from UniBim datasets

Table 5.13 Datasets used in the evaluation of the proposed network against other Self-supervised methods

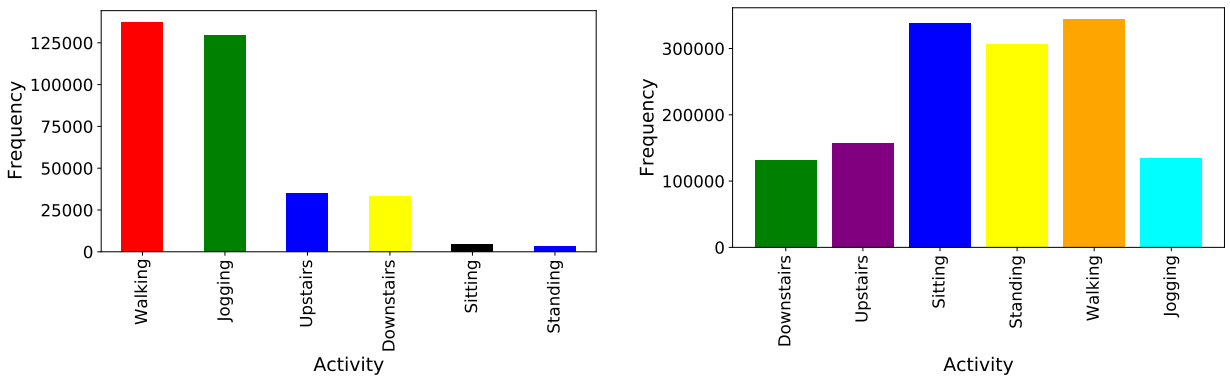
Datasets	Number of activities	number of users
UCI HAR	6	30
HHAR	6	9
UniMiB	9	30
WISDM	6	36
MotionSense	6	24

5.7 Results and Discussions for SHAR Network

In this section, the results of the experiments and evaluations that are conducted based on 12 public data are discussed. The datasets are recorded from sensors of smart home environments and wearable devices. The results of the SHAR are compared and evaluated with the state-of-the-art supervised and self-supervised methods. Ablation studies for various data corruption are also provided to further reveal the ability of the proposed classification network based on the proposed random masking and other signal transformation.

5.7.1 Results of Proposed Network Against Supervised Methods

Exhaustive experiments for HAR are performed to show the effectiveness of the proposed network and the results are shown. In this thesis, we proposed a framework that contains SHAR network and iSMOTE oversampling technique for HAR. The SHAR network is a



(a) WISDM dataset (b) MotionSense dataset

Fig. 5.9 Frequency of activities from wearable sensor datasets

self-supervised learning method that uses only unlabeled data for training and requires a small amount of annotated data for fine-tuning. Yet, the proposed network can outperform the existing methods. The iSMOTE oversampling technique is used to handle imbalanced class problems from the input data. The proposed SHAR network is evaluated and compared with the existing state-of-the-art methods [351] and DeepConvLSTM+Attention [14], using F1-score based on eight public human activity dataset. Tables 5.14 to 5.23 demonstrate that the Proposed SHAR outperforms the existing methods from all the datasets. The proposed iSMOTE oversampling technique is also evaluated and compared with the SMOTE oversampling technique and the original data based on the SHAR network. Furthermore, recognition of the minority classes is boosted using the proposed network in comparison with the current classification methods. Importantly the proposed network reduced the need for large labelled data for HAR systems. Figure 5.10 shows the summary of the experimental results from eight HAR sensor based datasets for the SHAR method and the state-of-the models. The results revealed that the proposed SHAR network outperforms the state-of-the-art classification models for HAR from eight datasets. The obtained classification outcomes of the SHAR method based on each dataset are presented in the below sections.

5.7.2 Results of SHAR from Ordonez Datasets

Tables 5.14 and 5.15 present the classification outcomes of the proposed network against the current state-of-the-art methods [14, 351] from Ordonez smart environments A and B. The presented results reveal that the proposed network outperforms the results obtained by the current state-of-the-art methods. The proposed network is evaluated using different labelling

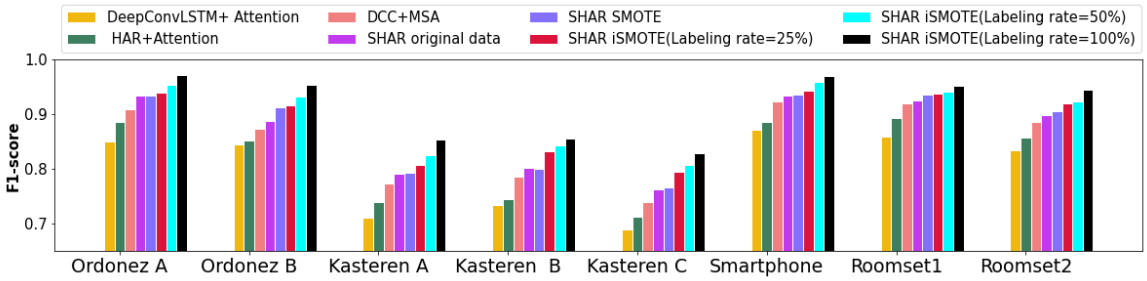


Fig. 5.10 Classification results from eight datasets using the proposed SHAR network and the state-of-the-art supervised methods

rates i.e., 100%, 50%, 25% and 10% for the fine-tuning phase. The SHAR network is also evaluated based on the original input data, SMOTE technique for oversampling and proposed iSMOTE technique for oversampling. The classification outcomes uncover that the our SHAR method based on the iSMOTE improves activity recognition compared with the current state-of-the-art methods, the SHAR based on the original data and the SHAR based on the SMOTE technique. Furthermore, the proposed SHAR network based on the iSMOTE technique with 25% of labelled data for fine-tuning phase has yet enhanced activity recognition compared with the current state-of-the-art methods, the SHAR based on the original data and the SHAR based on the SMOTE technique. The results reveal that the proposed network even based on the original data outperforms the current state-of-the-art methods. This shows the efficacy of our SHAR method proposed for human activity classification. In addition, The proposed framework, SHAR based on the iSMOTE technique, has further improved HAR, particularly minority classes which uncovers the ability of the iSMOTE in handling imbalanced classes. The minority classes from these two smart homes data include *Dinner, Grooming, Showering, Snack, Breakfast, and Toileting* as shown in Table 5.9 are well improved based on the SHAR and iSMOTE technique against the current methods. The results of the classification show that our network obtained better average results in both of the smart home datasets for all classes besides the results of each activity.

5.7.3 Results of SHAR from Kasteren Datasets

The proposed SHAR network is evaluated on the Kasteren intelligent environments A, B and C, the results of the proposed network against the state-of-the-art methods [14, 351] are shown in Tables 5.16, 5.17 and 5.18. The results of the experiments indicate that our proposed SHAR network obtained better results against the current methods and reduced the need for large labelled data. Different labelling rates i.e., 100%, 50%, 25% and 10%

Table 5.14 F1-score for the classification results of HAR in the Ordonez home A dataset

Labeling rate	100%	100%	100%	100%	100%	50%	25%	10%
Activity	DeepConvLSTM + Attention	HAR+ Attention	SHAR original data	SHAR SMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE
Breakfast	83.11	84.51	88.96	89.27	**94.56**	92.88	91.81	89.11
Grooming	75.32	80.00	87.75	88.11	**93.98**	92.12	90.05	88.24
Leaving	95.29	95.51	99.45	99.26	**99.61**	99.21	98.63	97.91
Lunch	95.44	94.39	97.16	97.49	**98.39**	97.00	96.59	95.88
Showering	80.65	86.89	94.91	95.21	**97.64**	94.76	93.12	92.68
Sleeping	97.53	97.11	98.36	98.12	**99.55**	99.17	98.19	97.33
Snack	70.74	82.63	88.22	88.51	**95.52**	92.94	90.00	88.18
Spare Time	96.83	97.21	99.69	98.24	**99.89**	97.32	97.00	96.11
Toileting	69.89	77.25	86.26	86.77	**95.14**	92.85	89.49	87.83
Average	84.97	88.55	93.41	93.44	**97.14**	95.36	93.87	92.58

Table 5.15 F1-score for the classification results of HAR in the Ordonez home B dataset

Labeling rate	100%	100%	100%	100%	100%	50%	25%	10%
Activity	DeepConvLSTM + Attention	HAR+ Attention	SHAR original data	SHAR SMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE
Leaving	89.79	92.09	95.22	94.39	**98.89**	96.11	94.62	90.55
Sleeping	96.37	95.42	98.78	98.11	**99.52**	99.32	98.23	92.98
Grooming	85.33	87.91	91.11	92.27	**97.91**	95.42	93. 13	88.98
Breakfast	74.87	75.39	83.46	84.19	**90.17**	88.81	87.12	83.41
Showering	79.43	79.12	89.45	90.02	**94.76**	92.36	90.23	87.12
Lunch	95.21	95.31	99.18	98.81	**99.77**	98.15	97.46	95.32
Snack	76.16	76.31	83.14	84.25	**90.23**	87.95	86.24	82.46
Toileting	83.56	83.24	90.31	89.98	**94.33**	91.12	89.67	86.24
Spare Time	78.21	79.32	86.78	86.67	**91.11**	88.58	86.79	83.96
Dinner	86.19	86.49	91.58	92.58	**96.19**	94.87	92.12	88.71
Average	84.51	85.06	88.79	91.13	**95.27**	93.26	91.56	87.96

are used in the fine-tuning phase to evaluate and show the effectiveness of our proposed network. The SHAR network based on the proposed iSMOTE oversampling technique is evaluated with the SHAR based on the original input data, and the SHAR based on the SMOTE oversampling technique. The results show that the proposed SHAR network based on the iSMOTE improves activity recognition compared with the current state-of-the-art methods, the SHAR based on the original data and the SHAR based on the SMOTE technique. Moreover, The proposed framework, SHAR based on the iSMOTE oversampling technique, with only 25% of labelled data for fine-tuning phase has yet enhanced activity recognition compared with the current state-of-the-art methods, the SHAR based on the original data and the SHAR based on the SMOTE technique. The results show that the proposed network even based on the original input data outperforms the current state-of-the-art methods. This indicates the effectiveness of the proposed SHAR network. In addition, The proposed framework, SHAR based on the iSMOTE technique, has further improved HAR particularly less frequent classes which shows the capability of the iSMOTE in addressing imbalanced class problems. The minority classes such as *Brush teeth* is well improved based on the

SHAR and iSMOTE technique compared with the current methods. The results show that our network obtained better average results in the smart home datasets for all classes besides the results of individual activity.

5.7.4 Results of SHAR from Wearable Devices

The proposed SHAR network is to classify human activities that are recorded using wearable devices. Tables 5.19, 5.20 and 5.21 reveal the classification outcomes of the proposed SHAR method against the current classification methods of human activity. Particularly, the classification outcomes of the proposed SHAR method from wearable sensors data are presented in Table 5.19. Additionally, the classification results from sensor data of wearable devices in Roomset1 and Roomset2 are presented in Tables 5.20 and 5.21. The results uncover that our proposed SHAR method outperforms the state-of-the-art techniques. The results show that the proposed network even based on the original input data outperforms the current state-of-the-art methods. This indicates the effectiveness of the proposed SHAR network. The proposed SHAR method boosted the single activity results and the average result scores of all activities in comparison of the current HAR methods from all sensor datasets of wearable devices. Moreover, the proposed SHAR network has significantly reduced the need for a large labelled data fine-tuning phase. With 25% of labelled data for fine-tuning phase, the proposed SHAR network based on the iSMOTE oversampling technique renders better classification outcomes compared to the existing methods, the SHAR based on the original input data, and the SHAR based on the SMOTE oversampling technique. This uncovers the effectiveness of the proposed iSMOTE oversampling technique in addressing the imbalanced class problems in wearable sensor data. The minority classes such as *Ambulating and Sit on chair* are well improved based on the SHAR and iSMOTE technique compared with the current classification HAR methods. The classification outcomes present that the proposed SHAR method obtained better average results in the wearable sensor datasets for all classes besides the results of individual activity.

5.7.5 Results of the Proposed SHAR Network Against other Self-supervised Methods

The proposed SHAR network is evaluated and compared with two self-supervised learning methods [329, 338]. Table 5.22 shows the results of the proposed SHAR network agains other self-supervised methods based on five datasets. The results show that the proposed SHAR network significantly improved HAR. For example, the result of the UCI HAR dataset

Table 5.16 F1-score for the classification results of HAR in the Kasteren smart home A dataset

Labeling rate	100%	100%	100%	100%	100%	50%	25%	10%
Activity	DeepConvLSTM + Attention	HAR+ Attention	SHAR original data	SHAR SMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE
Brush_teeth	43.59	52.22	57.19	56.89	**69.32**	66.86	64.22	59.21
Get_drink	59.33	59.54	68.21	68.94	**76.34**	73.56	71.19	68.22
Get_Snack	57.22	58.71	65.24	66.15	**69.97**	68.11	66.46	63.69
Go_to_bed	80.16	81.76	87.99	88.02	**94.13**	91.71	89.38	86.93
Leave_house	80.02	82.19	86.76	87.59	**93.84**	89.98	87.73	84.55
Prepare_breakfast	76.97	79.54	85.31	86.31	**91.85**	88.21	86.42	84.76
Prepare_Dinner	89.56	91.87	96.66	94.18	**98.04**	96.47	95.77	93.84
Take_shower	85.13	89.23	90.87	90.73	**95.49**	92.11	90.63	88.21
Use_toilet	67.82	69.34	73.21	74.22	**78.83**	75.33	74.65	72.49
Average	71.09	73.82	79.04	79.22	**85.31**	82.48	80.71	77.99

Table 5.17 F1-score for the classification results of HAR in the Kasteren smart home B dataset

Labeling rate	100%	100%	100%	100%	100%	50%	25%	10%
Activity	DeepConvLSTM + Attention	HAR+ Attention	SHAR original data	SHAR SMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE
Brush_teeth	42.89	47.82	53.21	53.98	**59.72**	58.21	57.29	52.67
Eat_brunch	90.93	91.11	96.21	95.38	**99.32**	98.54	97.82	93.21
Eat_dinner	86.79	86.29	91.87	92.68	**95.17**	94.05	93.61	90.18
Get_a_drink	44.15	44.75	55.21	55.23	**61.32**	60.44	59.01	55.10
Go_to_bed	96.32	94.48	99.88	98.21	**99.91**	98.49	97.89	93.57
Leaving_the_house	92.98	93.21	97.78	97.13	**99.76**	98.17	97.52	94.51
Prepare_brunch	85.62	84.29	89.92	88.69	**95.38**	93.97	92.80	89.14
Get_dressed	31.79	41.11	45.89	46.53	**58.19**	56.48	55.12	50.25
Prepare_dinner	96.21	95.31	97.98	96.64	**99.18**	98.41	97.62	91.45
Take_shower	81.95	82.13	85.32	85.69	**91.28**	90.14	89.47	85.32
Use_toilet	56.13	54.19	64.86	65.32	**74.11**	72.89	71.02	67.23
Wash_dishes	75.38	77.25	84.29	85.12	**93.12**	92.03	90.43	88.26
Average	73.42	74.32	80.20	80.03	**85.53**	84.31	83.12	79.24

is improved by 6% compared to the selfHAR [338] and 7% compared to the the method proposed in [329]. Hence, our proposed SHAR network outperformed the aforementioned self-supervised methods.

5.7.6 Ablation Studies Based on Different Signal Transformations

Ablation studies to further uncover the capability of the proposed SHAR network and show the contribution of the iSMOTE in the SHAR method for HAR systems is performed. The proposed method based on the iSMOTE is compared with the proposed network based on different transformation methods. Table 5.23 demonstrates the classification outcomes of the proposed SHAR method based on the iSMOTE and different transformation methods. The proposed SHAR method based on the iSMOTE technique outperforms the proposed method based on the transformation methods which are permutation, rotation, scaling, magnitude-

Table 5.18 F1-score for the classification results of HAR in the Kasteren smart home C dataset

Labeling rate	100%	100%	100%	100%	100%	50%	25%	10%
Activity	DeepConvLSTM + Attention	HAR+ Attention	SHAR original data	SHAR SMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE
Eating	80.98	84.22	86.89	85.61	**92.27**	91.23	90.24	87.05
Brush_teeth	63.55	67.76	69.98	69.46	**75.26**	74.28	73.62	70.23
Get_dressed	56.82	60.27	64.19	65.44	**73.48**	71.20	70.32	67.41
Get_drink	48.11	50.37	55.25	56.78	**64.31**	61.55	60.03	55.71
Get_snack	67.86	70.45	74.22	74.61	**79.23**	77.68	77.07	75.19
Go_to_bed	95.41	93.21	97.59	96.23	**99.93**	98.44	97.12	94.17
Leave_house	92.57	91.39	95.76	95.49	**99.11**	97.26	96.15	93.12
Prepare_Breakfast	78.45	81.24	84.68	84.29	**90.66**	88.79	86.92	83.35
Prepare_Dinner	79.68	80.14	86.21	86.39	**92.81**	90.59	88.13	86.05
prepare_Lunch	78.39	79.31	87.63	87.32	**93.45**	91.56	90.67	86.48
Use_Toilet_Downstairs	45.17	49.55	62.59	63,14	**69.37**	67.84	65.62	62.22
Use_toilet_upstairs	46.21	48.64	53.44	54.26	**63.18**	60.34	57.77	53.13
Shave	78.25	77.82	83.41	84.56	**90.16**	87.69	86.32	82.24
Take_medication	45.21	56.52	61.36	62.45	**69.43**	66.22	64.72	60.18
Take_shower	75.42	76.61	80.99	81.35	**89.56**	86.27	85.46	81.79
Average	68.79	71.16	76.27	76.49	**82.81**	80.72	79.34	75.88

Table 5.19 F1-score results in UCI HAR dataset

Labeling rate	100%	100%	100%	100%	100%	50%	25%	10%
Activity	DeepConvLSTM + Attention	HAR+ Attention	SHAR original data	SHAR SMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE
Laying	89.67	89.91	96.79	96.57	**98.93**	98.02	97.16	95.75
Sitting	86.45	90.25	95.99	94.31	**98.69**	97.22	96.32	94.02
Standing	88.92	88.56	94.32	95.05	**96.89**	95.92	95.31	94.86
Walking	80.89	81.11	88.89	88.21	**95.42**	95.01	93.19	89.67
Walking_downstairs	80.11	83.91	87.38	86.81	**92.11**	89.29	85.55	83.18
Walking_upstairs	96.93	97.23	100.00	99.91	**100**	100	98.12	94.15
Average	87.16	88.49	93.39	93.47	**97.01**	95.91	94.27	91.93

Table 5.20 F1-score performance of our proposed SHAR network and other supervised methods in RoomSet1

Labeling rate	100%	100%	100%	100%	100%	50%	25%	10%
Activity	DeepConvLSTM + Attention	HAR+ Attention	SHAR original data	SHAR SMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE
Ambulating	93.67	95.22	98.01	98.34	**99.59**	99.04	98.43	95.12
Lying	94.12	95.21	97.92	97.32	**99.79**	99.02	98.93	94.39
Sit_on_bed	95.31	96.25	99.93	99.82	**99.96**	99.91	99.17	96.68
Sit_on_chair	60.52	70.11	74.41	78.37	**81.56**	78.13	78.05	73.28
Average	85.90	89.19	92.56	93.46	**95.22**	94.02	93.64	89.86

warping (MagW), jittering (Jitter) and time-warping (TimeW) [364]. These transformation methods are described below:

Table 5.21 F1-score performance of our proposed SHAR network and other supervised methods in RoomSet2

Labeling rate	100%	100%	100%	100%	100%	50%	25%	10%
Activity	DeepConvLSTM + Attention	HAR+ Attention	SHAR original data	SHAR SMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE	SHAR iSMOTE
Ambulating	85.43	87.11	92.93	92.58	95.44	93.06	92.95	90.16
Lying	89.42	89.32	96.53	96.38	98.51	98.00	96.82	93.57
Sit_on_bed	96.21	97.52	99.84	99.87	99.91	98.89	99.74	94.76
Sit_on_chair	62.56	68.87	73.31	73.29	83.60	79.15	78.32	74.68
Average	83.40	85.70	89.87	90.53	**94.36**	92.27	91.95	88.17

Table 5.22 F1-score performance between our proposed SHAR network and other self-supervised methods

Datasets	Transformation Discrimination [329]	SelfHAR [338]	SHAR
UCI HAR	90.53	91.35	**97. 01**
HHAR	79.61	78.46	**84.38**
UniMiB	80.98	87.31	**90.17**
WISDM	89.48	90.81	**92.59**
MotionSense	92.95	96.31	**97.72**

i. Permutation: a simple transformation method that randomly perturbs the events in order to produce a new pattern within a temporal window through swapping and slicing various segments of the data. Permutation can enable the model to develop permutation invariance properties.

ii. Rotation: is applying arbitrary rotations on the input signal. The rotation inverts the the sign of the sensor without changing the associated class-label. This may even happen if the sensor is held upside down.

iii. Scaling: is a transformation that changes the the magnitude of the instances by multiplying a random scalar in a window.

iv. Magnitude-warping (MagW): is the transformation that changes the magnitude of each instance by convolving the time steps with a smooth curve varying around one.

v. Jittering (Jitter): is the transformation that changes the sensor readings by adding random noise to the sensor input signals.

vi. Time-warping (TimeW): is perturbing the temporal data location through a smooth distortion of time intervals between samples.

Table 5.23 Ablation studies performance of the SHAR based on different signal transformation techniques

Datasets	Proposed RM	Permutation	Rotation	Scaling	MagW	Jitter	TimeW
Ordonez Smart Home A	**97.14**	92.34.	93.21	90.53	91.88	93.22	89.34
Ordonez Smart Home B	**95.27**	91.56	89.71	92.68	92.11	92.84	90.14
Kastern Smart Home A	**85.31**	83.08	80.43	82.92	83.04	81.11	82.89
Kastern Smart Home B	**85.53**	82.11	83.54	79.29	81.10	82.38	81.22
Kastern SmartHome C	**82.81**	78.16	79.01	78.19	79.85	78.54	77.92
UCI HAR dataset	**97.01**	93.42	91.13	92.67	93.39	90.34	89.35
Wearable RoomSet1	**95.22**	92.47	92.82	91.79	92.27	92.91	91.28
Wearable RoomSet2	**94.36**	89.91	92.18	91.76	90.11	91.29	92.32
HHAR	**84.38**	79.91	82.11	80.23	80.15	80.29	80.56
UniMiB	**90.17**	88.11	88.29	87.94	89.38	88.95	88.54
WISDM	**92.59**	90.11	90.84	91.46	90.79	91.20	90.97
MotionSense	**97.72**	95.44	95.21	95.91	96.83	96.10	96.34

5.8 Discussions

This chapter proposes two deep learning approaches to transfer knowledge and reduce the need for a rich quantity of annotated human activity data. The proposed MDL network shares representation between different but related domains to train a generic model rather than learning each domain in isolation to enhance the performance of HAR systems. Besides, this thesis also proposes a self-supervised network which is called SHAR to reduce the need and effort for large labelled data and boost the performance of HARsystems for SDL tasks.

The proposed MDL network further improves the accuracy of human activity recognition from smart home and wearable sensor data compared to the existing methods. The proposed MDL network learns on two different but related sensor generated domains using a shared representation. The shared representation transfers knowledge across the domains to make strong mutual complementary features that improve the recognition rate and mitigate the data scarcity problems. The proposed MDL network uses the source domain with its full training data and the target domain with 50% of its training data to reduce the need for labelled training data. Six human activity datasets are used to evaluate the proposed MDL network. Extensive experiments are conducted to make a comparative analysis for the results that are obtained by the proposed MDL network and existing methods. The experimental results confirm that the results of the source and target domains achieved based on the proposed MDL network outperform the results achieved by the existing methods. The efficacy of the proposed MDL approach is validated by the results of the target domains even though 50% of their training data is used in the learning compared to using the full training data by the existing methods. In addition to improving the results of the target domains, the

results of the source domains are also improved using the proposed network compared to the state-of-the-art methods.

In addition, this thesis proposes, SHAR, a self-supervised learning network to improve human activity recognition and minimize the amount of required annotated data from intelligent environments and wearable sensors. The proposed SHAR network uses unlabelled data for the pre-training task and a small portion of the annotated data for the fine-tuning task. Extensive experiments are performed on 12 datasets and demonstrated the evaluation of the proposed SHAR method in comparison with the current classification state-of-the-art models. The classification outcomes of the experiments uncover that the proposed SHAR method outperformed multiple classification methods and minimized the need for large labelled data in comparison with the current classification state-of-the-art methods. we also proposed a method to address imbalanced class problems based on SMOTE technique which we call improved SMOTE (iSMOTE). We further presented ablation studies to reveal the effectiveness of the iSMOTE for activity classification. The results of the ablation studies uncover that the proposed SHAR method with our proposed iSMOTE renders better results compared with the SHAR based on the original imbalanced data and the SHAR based on the SMOTE oversampling technique. We also propose random masking, a signal transformation technique to mask the input data to remove identity mapping from the datasets and build a generic semantic representation for HAR. The proposed framework based on the SHAR, iSMOTE, and random masking outperforms the state-of-the-art methods for HAR systems.

5.9 Conclusion

To reduce the dependency on labelled data and build accurate HAR systems, we have conducted transfer learning using two deep learning approaches. First, we proposed MDL network to share representation between different but related domains and train a generic model rather than learning each domain in isolation for enhancing the performance of HAR systems. The proposed MDL network outperformed the existing methods form smart home and wearable sensor data. The proposed MDL network learns on two different but related sensor generated domains using a shared representation. The shared representation transfers knowledge across the domains to make strong mutual complementary features that improve the recognition rate and mitigate the data scarcity problems. The proposed MDL network uses the source domain with its full training data and the target domain with 50% of its training data to reduce the need for labelled training data. Six human activity datasets are used to evaluate the proposed MDL network. Besides, we proposed a self-supervised

network which is called SHAR to reduce the dependency on large labelled data and boost the performance of HAR systems for SDL tasks. SHAR improves HAR systems and minimize the dependency on annotated data from intelligent environments and wearable sensors. The proposed SHAR network uses unlabelled data for the pre-training task and a small portion of the annotated data for the fine-tuning task. The proposed SHAR network outperforms the state-of-the-art methods for HAR systems.

Chapter 6

Conclusion

This chapter concludes the thesis and highlights the possible future directions of human activity recognition. The summary of the thesis about human activity recognition is presented. Human activity recognition which aims to accurately recognise human daily activities is an active and challenging research field in ubiquitous computing and plays a significant role in several applications such as healthcare monitoring, security surveillance systems, and resident situation assessment. In this thesis, the proposed approaches draw on several major research challenges of human activity recognition. The challenges are further enhancing accuracy against the-state-of-art methods to accurately recognise human activities, skewed distribution of human activities, and also the need for a rich quantity of well-curated human activity data. This thesis proposes robust and more accurate networks to further enhance human activity recognition, address long-tailed distributions of human activities, and minimise the need for a rich quantity of well-curated human activity data. Particularly, in this chapter, we briefly summarise the key contributions of this PhD thesis, indicate the limitations of the proposed networks, and highlight future research developments.

6.1 Summary of Thesis

This thesis proposed robust and more accurate temporal sequential learning models to enhance the performance of human activity recognition systems compared with the existing approaches to sensors collected data. The proposed methods integrate causal convolution to prevent information leakage from the future to the past of time steps and preserve the ordering of the time steps of the temporal sensors data. Besides different attention mechanisms are proposed to effectively expose deep semantic correlations from action sequences involving human activities and to focus more on significant information. Furthermore, di-

lated convolutions within the proposed method are used to maximise the receptive field by orders of magnitude and aggregate multi-scale contextual information without increasing computational cost. Extensive experiments are conducted using eight benchmark sensor datasets to validate the proposed networks. The results of the experiments demonstrate that the proposed networks outperformed the current-state-of-the art methods.

In this thesis, different methods are proposed to address the imbalanced class problems for human activity recognition systems. First, joint learning of sequential deep learning algorithms, i.e., long short-term memory and convolutional neural networks is proposed to improve the performance of human activity recognition, particularly for infrequent human activities. The proposed method combines the learning processes of two temporal models in a single joint training mechanism to enhance the accuracy of minority classes in addition to maintaining the accuracy of majority classes. The two temporal learners of the jointly proposed methods exploit different features from the input data to render a strong mutual complementary model. Complementarity in joint learning based on different models can greatly boost the performance of minority activities. This is because each base learner brings different features into the joint learner to enrich the joint learning process and each learner improves the earlier layers of the other learner, but at the same time, the weaknesses of each individual learner are avoided. The joint optimization that leads to increasing functionality of the proposed joint temporal models to gain more insight into the input data and features reduces the recognition error rate. Thereby, the proposed model increases the performance of human activity recognition, particularly the minority classes.

In addition to that, we also proposed a data-level solution to address imbalanced class problems by extending the synthetic minority over-sampling technique (SMOTE) which we named (iSMOTE) to avoid misgenerated new samples. The proposed iSMOTE computes k-nearest neighbours of each generated instance to make sure the new instances are accurately annotated. Each new instance of the minority classes with its k-nearest neighbours must have the same class. These methods have enhanced the results of the minority human activities and outperformed the current state-of-the-art methods.

Moreover, sequential deep learning networks are proposed to boost the performance of human activity recognition and reduce the need for a large quantity of annotated human activity data by transfer learning techniques. A multi-domain learning network is proposed to process data from multi-domains, transfer knowledge across different but related domains of human activities and mitigate isolated learning paradigms using a shared representation. The advantage of the proposed method is firstly to reduce the need and effort for labelled data of the target domain. The proposed network uses the training data of the target domain with

restricted size and the full training data of the source domain, yet provided better performance than using the full training data in a single domain setting. Secondly, the proposed method can be used for small datasets. Lastly, the proposed multi-domain learning network reduces the training time by rendering a generic model for related domains compared to fitting a model for each domain separately.

Finally, this thesis also proposed a self-supervised model to further minimise the need for a large quantity of annotated human activity data. The self-supervised method is pre-trained entirely on unlabeled data and fine-tuned on a small amount of labelled data for supervised learning. A random masking method is proposed and applied on the input datasets to remove identity mappings from the datasets and build generic representations for the human activity recognition task. The proposed self-supervised pre-training network renders human activity representations that are semantically meaningful and provides a good initialization for supervised fine-tuning. The developed network enhances the performance of human activity recognition in addition to minimizing the need for a considerable amount of labelled data.

6.2 Limitations

Even though the proposed approaches in the thesis enhanced human activity recognition and minimised the need for large labelled data, the thesis has some limitations. Firstly, the proposed multi-domain learning model processes different but related source and target domains, yet the proposed multi-domain learning model is unable to appropriately process wearable sensor data and smart sensor data together for the source and target domains of human activity recognition. Rather, both domains ought to be based on either wearable sensor data or smart home sensor data to be similar to each other. When learning the proposed network based on either wearable sensor data or smart home sensors data for the source domain makes the learning of the target domain harder which leads to negative transfer [368].

Secondly, the proposed joint learning of sequential learning models boosted the performance of human activity recognition and addressed the imbalanced class problems as an algorithm-level solution. However, the proposed joint learning network has not adequately addressed the imbalanced class problems since some of the activities such as *Snack* and $Breakfast$ have extremely few samples which make their learning by the network harder. Hence, the datasets with long-tailed distributions of human activities require data-level solutions in addition to an algorithm-level solution for human activity recognition.

6.3 Future Work

Future research will further explore imbalanced class problems to develop a hybrid model of data-level and algorithm-level for human activity recognition. The hybrid model is required to better address the imbalanced class problems and further improve the results of the minority classes, particularly the activities with extremely few samples.

Besides, a better attention mechanism is required compared with the current attention mechanism to precisely capture important information, particularly for fine-grained human activities such as Breakfast and Snack, or Walking downstairs and Walking upstairs to further improve human activity recognition. Several human activities are similar and share some characteristics such as snack, breakfast, lunch or dinner that cause the overlapping problem. These activities are less discriminative due to their overlaps in the feature space which makes the recognition process much harder. Furthermore, often human activity recognition is performed for single-user activities at a time, however, in real-life scenarios, concurrently multiple activities could be conducted by numerous individuals. Thus, recognizing multi-user activities and their interactions is still an open research problem and requires further research.

Considering the limitations of this thesis, a self-supervised multi-domain learning model is required to further minimise the need for large labelled data and process multi-domains jointly rather than learning each domain in isolation. The method must consider multi-modal sensor data and follow data fusion strategies to further boost the performance of human activity recognition systems and render a pre-trained multi-domain learning model entirely on unlabelled data.

Appendix A

Sequential Temporal Models

A.1 LSTM

In this thesis project, LSTM as a temporal model is used to be compared with the proposed methods. Two layers of LSTM with a flattened layer are stacked. Then the outputs of the flattened layer are passed into a fully connected layer with the ReLU activation function and followed by a softmax layer. Figure A.1 shows the architecture of the LSTM model.

Fast LSTM implementation backed by cuDNN (CUdNNLSTM) [369] is also used in this study with the same architecture of LSTM model. CUdNNLSTM is a version of LSTM that uses the CuDNN library and it can only be run on a GPU to accelerate training and inference time.

A.2 ConvNet

In this thesis, 1D ConvNet is employed and its results are shown and compared with the results of the proposed methods. The 1D ConvNet model is designed by stacking two convolutional layers each with 64 filters. The kernel size of the 1D ConvNet in this study is equal to 3 which indicates the length of the 1D convolution window with a stride size of 1. A Max-pooling layer with a window size equal to 2 is applied after the convolution layers to down-sample the features maps. The feature maps are flattened to be processed by the fully connected, i.e., a dense layer with ReLU activation function followed by a soft-max layer. Figure A.2 shows the architecture of 1D ConvNet.

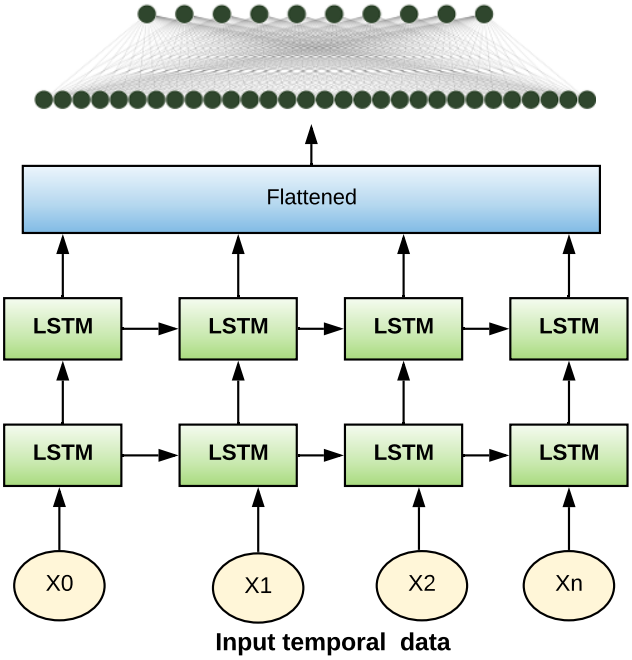


Fig. A.1 Architecture of the LSTM model

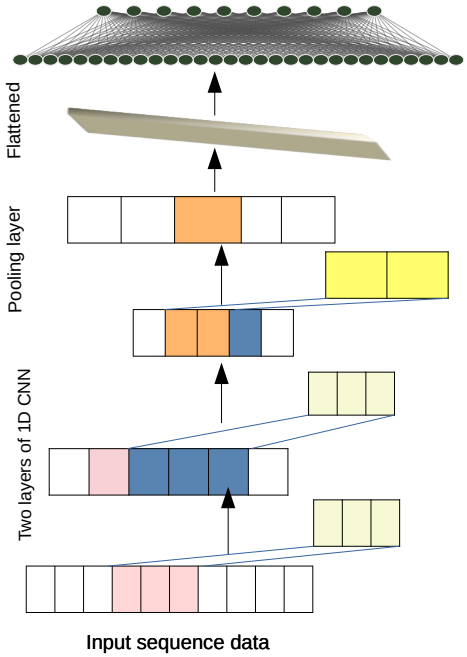


Fig. A.2 Architecture of the 1D ConvNet model

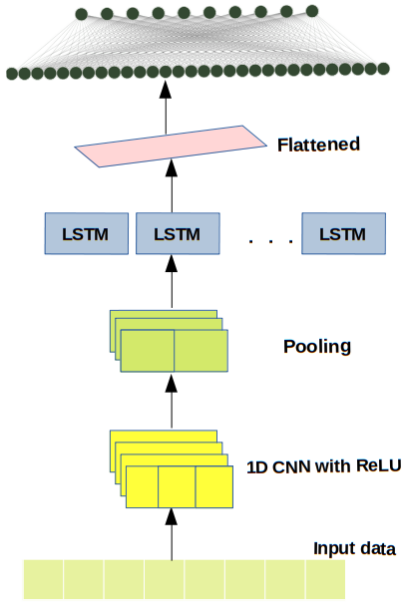


Fig. A.3 Hybrid 1D ConvNet + LSTM model

A.3 Hybrid of ConvNet and LSTM

In this study, the hybrid model is employed by stacking one layer of each 1D ConvNet and LSTM to human activities from smart home data. Figure A.3 shows the architecture of the hybrid model. The input data are firstly fed into the 1D ConvNet layer to extract features before the LSTM layer to support sequence recognition. The input sub-sequences sensor data are processed independently by 1D ConvNet hence timestep orders are not considered. The feature maps of 1D ConvNet are down-sampled by a max-pooling layer with the window size equal to 2 before the LSTM layer. The feature maps are processed by the LSTM and then flattened followed by fully-connected layers, i.e., a dense layer with ReLU activation function and a soft-max layer.

However, order sensitivity is not considered in the extracted features by the 1D CNN. Hence the hybrid of 1D CNN and LSTM is not the most acceptable solution to improve the performance of activity recognition.

A.4 Bidirectional LSTM

Bidirectional LSTM trains input data in forward and backward directions by using previous and subsequent information of a specific time step in two separate recurrent layers [370]. Figure A.4 shows bidirectional LSTM where inputs of backward states are not connected to the

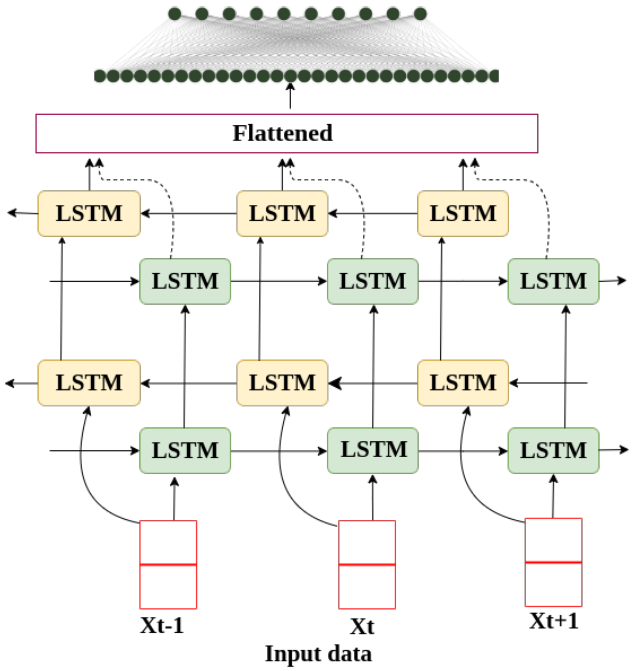


Fig. A.4 Bidirectional LSTM model

outputs of the forward states. Including future information in addition to past information in bidirectional LSTM appears at first sight to violate causality [371]. Although Bidirectional LSTM has been successfully proposed in HAR and achieved satisfying results, Bidirectional LSTM is indeed expensive to train since it has a double recurrent setting in each layer [372]. Bidirectional LSTM is used in this thesis by stacking two forward and backwards LSTM layers. The outputs of these two layers are flattened and then fed to a fully-connected layer, i.e., a dense layer with ReLU activation function and a soft-max layer.

Table A.1 Architectures of temporal models and state-of-the-art methods

Model	Architectures
LSTM	Two consecutive LSTM layers are flattened and followed by a fully connected layer and a softmax layer.
1D CNN	Two 1D convolutional layers each with 64 filters with the kernel size equal to 3 followed by a max-pooling layer are used to create the feature maps. Then the feature maps are flattened and passed into a fully-connected layer followed by a softmax layer.
Hybrid 1D CNN and LSTM	A 1D convolutional layer followed by a max-pooling layer is stacked with an LSTM layer to create feature maps that are flattened and fed into a fully-connected layer followed by a softmax layer.
Bi-LSTM and CuDNN LSTM	The structures of Bi-LSTM and CuDNN LSTM are the same as the LSTM model but instead LSTM, Bi-LSTM and CuDNN LSTM are used.
DeepConvLSTM + Attention [14]	A 1D convolution followed by an LSTM layer is used to create feature maps. Then the self-attention mechanism followed by a softmax layer is applied on the feature maps.
HAR+Attention [351]	A 1D convolution layer followed by positional encoding and N self-attention blocks are used to create feature maps. Then a global attention mechanism followed by a softmax layer is applied on the feature maps.

References

[1] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, 46(3):1–33, 2014.

[2] Diane J Cook and Narayanan C Krishnan. *Activity learning: discovering, recognizing, and predicting human behavior from sensor data*. John Wiley & Sons, 2015.

[3] Oscar D Lara and Miguel A Labrador. A survey on human activity recognition using wearable sensors. *IEEE communications surveys & tutorials*, 15(3):1192–1209, 2012.

[4] Kalpa Kharicha, Steve Iliffe, Danielle Harari, Cameron Swift, Gerhard Gillmann, and Andreas E Stuck. Health risk appraisal in older people 1: are older people living alone an 'at-risk'group? *British Journal of General Practice*, 57(537):271–276, 2007.

[5] Florenc Demrozi, Graziano Pravadelli, Azra Bihorac, and Parisa Rashidi. Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. *IEEE Access*, 8:210816–210836, 2020.

[6] Rodolfo S Antunes, Lucas A Seewald, Vinicius F Rodrigues, Cristiano A Da Costa, Luiz Gonzaga Jr, Rodrigo R Righi, Andreas Maier, Bjoern Eskofier, Malte Ollenschlaeger, Farzad Naderi, et al. A survey of sensors in healthcare workflow monitoring. *ACM Computing Surveys (CSUR)*, 51(2):1–37, 2018.

[7] Qin Ni, Ana García Hernando, and Iván de la Cruz. The elderly's independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development. *Sensors*, 15(5):11312–11362, 2015.

[8] 10 facts on ageing and health. https://www.who.int/news-room/fact-sheets/detail/10-facts-on-ageing-and-health. Accessed: 2022-09-30.

[9] Thinagaran Perumal, YL Chui, Mohd Anuaruddin Bin Ahmadon, and Shingo Yamaguchi. Iot based activity recognition among smart home residents. In *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*, pages 1–2. IEEE, 2017.

[10] Adls (activities of daily living). https://www.medicinenet.com/script/main/art.asp?articlekey=2152. Accessed: 2020-07-19.

[11] Ahatsham Hayat, Fernando Morgado-Dias, Bikram Pratim Bhuyan, and Ravi Tomar. Human activity recognition for elderly people using machine and deep learning approaches. *Information*, 13(6):275, 2022.

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

[13] Aiguo Wang, Guilin Chen, Xi Wu, Li Liu, Ning An, and Chih-Yung Chang. Towards human activity recognition: A hierarchical feature selection framework. *Sensors*, 18 (11):3629, 2018.

[14] Satya P Singh, Madan Kumar Sharma, Aimé Lay-Ekuakille, Deepak Gangwar, and Sukrit Gupta. Deep convlstm with self-attention for human activity decoding using wearable sensors. *IEEE Sensors Journal*, 21(6):8575–8582, 2020.

[15] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th annual international conference on machine learning*, pages 609–616, 2009.

[16] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6):82–97, 2012.

[17] Honglak Lee, Peter Pham, Yan Largman, and Andrew Y Ng. Unsupervised feature learning for audio classification using convolutional deep belief networks. In *Advances in neural information processing systems*, pages 1096–1104, 2009.

[18] Rui Zhao, Jinjiang Wang, Ruqiang Yan, and Kezhi Mao. Machine health monitoring with lstm networks. In *2016 10th International Conference on Sensing Technology (ICST)*, pages 1–6. IEEE, 2016.

[19] Rebeen Ali Hamad, Alberto Salguero Hidalgo, Mohamed-Rafik Bouguelia, Macarena Espinilla Estevez, and Javier Medina Quero. Efficient activity recognition in smart homes using delayed fuzzy temporal windows on binary sensors. *IEEE journal of biomedical and health informatics*, 24(2):387–395, 2019.

[20] Anindya Das Antar, Masud Ahmed, and Md Atiqur Rahman Ahad. Challenges in sensor-based human activity recognition and a comparative analysis of benchmark datasets: a review. In *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, pages 134–139. IEEE, 2019.

[21] Kaixuan Chen, Dalin Zhang, Lina Yao, Bin Guo, Zhiwen Yu, and Yunhao Liu. Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)*, 54(4):1–40, 2021.

[22] Henry Friday Nweke, Ying Wah Teh, Mohammed Ali Al-Garadi, and Uzoma Rita Alo. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications*, 2018.

[23] Rebeen Ali Hamad, Masashi Kimura, and Jens Lundström. Efficacy of imbalanced data handling methods on deep learning for smart homes environments. *SN Computer Science*, 1(4):1–10, 2020.

[24] Yinan Guo, Yaoqi Chu, Botao Jiao, Jian Cheng, Zekuan Yu, Ning Cui, and Lianbo Ma. Evolutionary dual-ensemble class imbalance learning for human activity recognition. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2021.

[25] Donghui Wu, Zhelong Wang, Ye Chen, and Hongyu Zhao. Mixed-kernel based weighted extreme learning machine for inertial sensor based human activity recognition with imbalanced dataset. *Neurocomputing*, 190:35–49, 2016.

[26] Jun Qi, Po Yang, Martin Hanneghan, Stephen Tang, and Bo Zhou. A hybrid hierarchical framework for gym physical activity recognition and measurement using wearable sensors. *IEEE Internet of Things Journal*, 6(2):1384–1393, 2018.

[27] Carlos Aviles-Cruz, Eduardo Rodriguez-Martinez, Juan Villegas-Cortez, and Andrés Ferreyra-Ramirez. Granger-causality: An efficient single user movement recognition using a smartphone accelerometer sensor. *Pattern Recognition Letters*, 125:576–583, 2019.

[28] Ganbayar Batchuluun, Jong Hyun Kim, Hyung Gil Hong, Jin Kyu Kang, and Kang Ryoung Park. Fuzzy system based human behavior recognition by combining behavior prediction and recognition. *Expert Systems with Applications*, 81:108–133, 2017.

[29] S Sankar, P Srinivasan, and R Saravanakumar. Internet of things based ambient assisted living for elderly people health monitoring. *Research Journal of Pharmacy and Technology*, 11(9):3900–3904, 2018.

[30] Eftim Zdravevski, Petre Lameski, Vladimir Trajkovik, Andrea Kulakov, Ivan Chorbev, Rossitza Goleva, Nuno Pombo, and Nuno Garcia. Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering. *Ieee Access*, 5:5262–5280, 2017.

[31] Nicole A Capela, Edward D Lemaire, and Natalie Baddour. Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients. *PloS one*, 10(4):e0124414, 2015.

[32] Andrea Prati, Caifeng Shan, and Kevin I-Kai Wang. Sensors, vision and networks: From video surveillance to activity recognition and health monitoring. *Journal of Ambient Intelligence and Smart Environments*, 11(1):5–22, 2019.

[33] An-An Liu, Ning Xu, Yu-Ting Su, Hong Lin, Tong Hao, and Zhao-Xuan Yang. Single/multi-view human action recognition via regularized multi-task learning. *Neurocomputing*, 151:544–553, 2015.

[34] Brian K Hensel, George Demiris, and Karen L Courtney. Defining obtrusiveness in home telehealth technologies: A conceptual framework. *Journal of the American Medical Informatics Association*, 13(4):428–431, 2006.

[35] Gita Sukthankar, Christopher Geib, Hung Bui, David Pynadath, and Robert P Goldman. *Plan, activity, and intent recognition: Theory and practice*. Newnes, 2014.

[36] Yan Wang, Shuang Cang, and Hongnian Yu. A survey on wearable sensor modality centred human activity recognition in health care. *Expert Systems with Applications*, 137:167–190, 2019.

[37] Im Y Jung. A review of privacy-preserving human and human activity recognition. *International Journal on Smart Sensing & Intelligent Systems*, 13(1), 2020.

[38] 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*, 114:441–453, 2018.

[39] Javier Medina Quero, Claire Orr, Shuai Zang, Chris Nugent, Alberto Salguero, and Macarena Espinilla. Real-time recognition of interleaved activities based on ensemble classifier of long short-term memory with fuzzy temporal windows. In *Multidisciplinary digital publishing institute proceedings*, volume 2, page 1225, 2018.

[40] Alvina Anjum and Muhammad U Ilyas. Activity recognition using smartphone sensors. In *2013 ieee 10th consumer communications and networking conference (ccnc)*, pages 914–919. IEEE, 2013.

[41] Grazia Cicirelli, Roberto Marani, Antonio Petitti, Annalisa Milella, and Tiziana D'Orazio. Ambient assisted living: A review of technologies, methodologies and future perspectives for healthy aging of population. *Sensors*, 21(10):3549, 2021.

[42] Shibo Zhang, Yaxuan Li, Shen Zhang, Farzad Shahabi, Stephen Xia, Yu Deng, and Nabil Alshurafa. Deep learning in human activity recognition with wearable sensors: A review on advances. *Sensors*, 22(4):1476, 2022.

[43] Sarbagya Ratna Shakya, Chaoyang Zhang, and Zhaoxian Zhou. Comparative study of machine learning and deep learning architecture for human activity recognition using accelerometer data. *Int. J. Mach. Learn. Comput*, 8(6):577–582, 2018.

[44] Abdul Rehman Javed, Muhammad Usman Sarwar, Suleman Khan, Celestine Iwendi, Mohit Mittal, and Neeraj Kumar. Analyzing the effectiveness and contribution of each axis of tri-axial accelerometer sensor for accurate activity recognition. *Sensors*, 20(8): 2216, 2020.

[45] Tal Shany, Stephen J Redmond, Michael R Narayanan, and Nigel H Lovell. Sensors-based wearable systems for monitoring of human movement and falls. *IEEE Sensors Journal*, 12(3):658–670, 2011.

[46] Yao-Chiang Kan and Chun-Kai Chen. A wearable inertial sensor node for body motion analysis. *IEEE Sensors Journal*, 12(3):651–657, 2011.

[47] Edward S Sazonov, George Fulk, James Hill, Yves Schutz, and Raymond Browning. Monitoring of posture allocations and activities by a shoe-based wearable sensor. *IEEE Transactions on Biomedical Engineering*, 58(4):983–990, 2010.

[48] Benoit Mariani, Mayté Castro Jiménez, François JG Vingerhoets, and Kamiar Aminian. On-shoe wearable sensors for gait and turning assessment of patients with parkinson's disease. *IEEE transactions on biomedical engineering*, 60(1):155–158, 2012.

[49] Subhas Chandra Mukhopadhyay. Wearable sensors for human activity monitoring: A review. *IEEE sensors journal*, 15(3):1321–1330, 2014.

[50] Catherine Tong, Shyam A Tailor, and Nicholas D Lane. Are accelerometers for activity recognition a dead-end? In *Proceedings of the 21st International Workshop on Mobile Computing Systems and Applications*, pages 39–44, 2020.

[51] Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data. In *International conference on pervasive computing*, pages 1–17. Springer, 2004.

[52] Maria Cornacchia, Koray Ozcan, Yu Zheng, and Senem Velipasalar. A survey on activity detection and classification using wearable sensors. *IEEE Sensors Journal*, 17 (2):386–403, 2016.

[53] Derick A Johnson and Mohan M Trivedi. Driving style recognition using a smartphone as a sensor platform. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1609–1615. IEEE, 2011.

[54] Shyamal Patel, Hyung Park, Paolo Bonato, Leighton Chan, and Mary Rodgers. A review of wearable sensors and systems with application in rehabilitation. *Journal of neuroengineering and rehabilitation*, 9(1):1–17, 2012.

[55] Pierluigi Casale, Oriol Pujol, and Petia Radeva. Human activity recognition from accelerometer data using a wearable device. In *Iberian conference on pattern recognition and image analysis*, pages 289–296. Springer, 2011.

[56] Yonglei Zheng, Weng-Keen Wong, Xinze Guan, and Stewart Trost. Physical activity recognition from accelerometer data using a multi-scale ensemble method. In *Twenty-Fifth IAAI Conference*, 2013.

[57] Hristijan Gjoreski and Matjaž Gams. Accelerometer data preparation for activity recognition. In *Proceedings of the International Multiconference Information Society, Ljubljana, Slovenia*, volume 1014, page 1014, 2011.

[58] Ming Jiang, Hong Shang, Zhelong Wang, Hongyi Li, and Yuechao Wang. A method to deal with installation errors of wearable accelerometers for human activity recognition. *Physiological measurement*, 32(3):347, 2011.

[59] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.

[60] Chun Zhu and Weihua Sheng. Motion-and location-based online human daily activity recognition. *Pervasive and Mobile Computing*, 7(2):256–269, 2011.

[61] Pekka Siirtola and Juha Röning. User-independent human activity recognition using a mobile phone: Offline recognition vs. real-time on device recognition. In *Distributed computing and artificial intelligence*, pages 617–627. Springer, 2012.

[62] C Sweetlin Hemalatha and Vijay Vaidehi. Frequent bit pattern mining over tri-axial accelerometer data streams for recognizing human activities and detecting fall. *Procedia Computer Science*, 19:56–63, 2013.

[63] Andrea Mannini, Stephen S Intille, Mary Rosenberger, Angelo M Sabatini, and William Haskell. Activity recognition using a single accelerometer placed at the wrist or ankle. *Medicine and science in sports and exercise*, 45(11):2193, 2013.

[64] Lei Gao, AK Bourke, and John Nelson. Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Medical engineering & physics*, 36(6):779–785, 2014.

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

[66] Juan Carlos Davila, Ana-Maria Cretu, and Marek Zaremba. Wearable sensor data classification for human activity recognition based on an iterative learning framework. *Sensors*, 17(6):1287, 2017.

[67] 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*, 81:307–313, 2018.

[68] Shaohua Wan, Lianyong Qi, Xiaolong Xu, Chao Tong, and Zonghua Gu. Deep learning models for real-time human activity recognition with smartphones. *Mobile Networks and Applications*, 25(2):743–755, 2020.

[69] Sakorn Mekruksavanich and Anuchit Jitpattanakul. Lstm networks using smartphone data for sensor-based human activity recognition in smart homes. *Sensors*, 21(5):1636, 2021.

[70] Chaolei Han, Lei Zhang, Yin Tang, Wenbo Huang, Fuhong Min, and Jun He. Human activity recognition using wearable sensors by heterogeneous convolutional neural networks. *Expert Systems with Applications*, 198:116764, 2022.

[71] Zhenghua Chen, Chaoyang Jiang, Shili Xiang, Jie Ding, Min Wu, and Xiaoli Li. Smartphone sensor-based human activity recognition using feature fusion and maximum full a posteriori. *IEEE Transactions on Instrumentation and Measurement*, 69 (7):3992–4001, 2019.

[72] Muhammad Shoaib, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul JM Havinga. Complex human activity recognition using smartphone and wrist-worn motion sensors. *Sensors*, 16(4):426, 2016.

[73] Yufei Chen and Chao Shen. Performance analysis of smartphone-sensor behavior for human activity recognition. *Ieee Access*, 5:3095–3110, 2017.

[74] Marcin Straczkiewicz, Peter James, and Jukka-Pekka Onnela. A systematic review of smartphone-based human activity recognition methods for health research. *NPJ Digital Medicine*, 4(1):1–15, 2021.

[75] 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)*, 42(6):790–808, 2012.

[76] Weslllen Sousa Lima, Eduardo Souto, Khalil El-Khatib, Roozbeh Jalali, and Joao Gama. Human activity recognition using inertial sensors in a smartphone: An overview. *Sensors*, 19(14):3213, 2019.

[77] Michael R Narayanan, Stephen J Redmond, Maria Elena Scalzi, Stephen R Lord, Branko G Celler, Nigel H Lovell, et al. Longitudinal falls-risk estimation using triaxial accelerometry. *IEEE Transactions on Biomedical Engineering*, 57(3):534–541, 2009.

[78] Barry R Greene, Alan O'Donovan, Roman Romero-Ortuno, Lisa Cogan, Cliodhna Ni Scanaill, and Rose A Kenny. Quantitative falls risk assessment using the timed up and go test. *IEEE Transactions on Biomedical Engineering*, 57(12):2918–2926, 2010.

[79] John Paul Varkey, Dario Pompili, and Theodore A Walls. Human motion recognition using a wireless sensor-based wearable system. *Personal and Ubiquitous Computing*, 16(7):897–910, 2012.

[80] Nagender K Suryadevara and Subhas C Mukhopadhyay. Determining wellness through an ambient assisted living environment. *IEEE Intelligent Systems*, 29(3):30–37, 2014.

[81] Narayanan C Krishnan and Diane J Cook. Activity recognition on streaming sensor data. *Pervasive and mobile computing*, 10:138–154, 2014.

[82] Prafulla Nath Dawadi, Diane Joyce Cook, and Maureen Schmitter-Edgecombe. Automated cognitive health assessment from smart home-based behavior data. *IEEE journal of biomedical and health informatics*, 20(4):1188–1194, 2015.

[83] Guilin Chen, Aiguo Wang, Shenghui Zhao, Li Liu, and Chih-Yung Chang. Latent feature learning for activity recognition using simple sensors in smart homes. *Multimedia Tools and Applications*, 77(12):15201–15219, 2018.

[84] Albert Haque, Arnold Milstein, and Li Fei-Fei. Illuminating the dark spaces of healthcare with ambient intelligence. *Nature*, 585(7824):193–202, 2020.

[85] Maria Ahmed Qureshi, Kashif Naseer Qureshi, Gwanggil Jeon, and Francesco Piccialli. Deep learning-based ambient assisted living for self-management of cardiovascular conditions. *Neural Computing and Applications*, pages 1–19, 2021.

[86] Rebeen Ali Hamad, Eric Järpe, and Jens Lundström. Stability analysis of the t-sne algorithm for human activity pattern data. In *2018 IEEE international conference on systems, man, and cybernetics (SMC)*, pages 1839–1845. IEEE, 2018.

[87] Jin-Hyuk Hong, Julian Ramos, Choonsung Shin, and Anind K Dey. An activity recognition system for ambient assisted living environments. In *International Competition on Evaluating AAL Systems Through Competitive Benchmarking*, pages 148–158. Springer, 2012.

[88] Holger Storf, Thomas Kleinberger, Martin Becker, Mario Schmitt, Frank Bomarius, and Stephan Prueckner. An event-driven approach to activity recognition in ambient assisted living. In *European conference on ambient intelligence*, pages 123–132. Springer, 2009.

[89] Aiguo Wang, Shenghui Zhao, Chundi Zheng, Jing Yang, Guilin Chen, and Chih-Yung Chang. Activities of daily living recognition with binary environment sensors using deep learning: A comparative study. *IEEE Sensors Journal*, 21(4):5423–5433, 2020.

[90] Zhiqiang Zhang, Xuebin Gao, Jit Biswas, and Jian Kang Wu. Moving targets detection and localization in passive infrared sensor networks. In *2007 10th International Conference on Information Fusion*, pages 1–6. IEEE, 2007.

[91] Geetika Singla, Diane J Cook, and Maureen Schmitter-Edgecombe. Tracking activities in complex settings using smart environment technologies. *International journal of biosciences, psychiatry, and technology (IJBSPT)*, 1(1):25, 2009.

[92] Barnan Das, Narayanan C Krishnan, and Diane J Cook. Handling class overlap and imbalance to detect prompt situations in smart homes. In *2013 IEEE 13th international conference on data mining workshops*, pages 266–273. IEEE, 2013.

[93] Monica-Andreea Dragan and Irina Mocanu. Human activity recognition in smart environments. In *2013 19th International Conference on Control Systems and Computer Science*, pages 495–502, 2013. doi: 10.1109/CSCS.2013.78.

[94] Aiguo Wang, Guilin Chen, Cuijuan Shang, Miaofei Zhang, and Li Liu. Human activity recognition in a smart home environment with stacked denoising autoencoders. In *International conference on web-age information management*, pages 29–40. Springer, 2016.

[95] Abdelhamid Salih and A Abraham. A review of ambient intelligence assisted healthcare monitoring. *International Journal of Computer Information Systems and Industrial Management (IJCISIM)*, 5:741–750, 2013.

[96] Gennaro Tartarisco, Giovanni Baldus, Daniele Corda, Rossella Raso, Antonino Arnao, Marcello Ferro, Andrea Gaggioli, and Giovanni Pioggia. Personal health system architecture for stress monitoring and support to clinical decisions. *Computer Communications*, 35(11):1296–1305, 2012.

[97] Pema Chodon, Devi Maya Adhikari, Gopal Chandra Nepal, Rajen Biswa, Sangay Gyeltshen, et al. Passive infrared (pir) sensor based security system. *International Journal of Electrical, Electronics & Computer Systems*, 14(2), 2013.

[98] Dipak Surie, Olivier Laguionie, and Thomas Pederson. Wireless sensor networking of everyday objects in a smart home environment. In *2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pages 189–194. IEEE, 2008.

[99] Xin Hong and Chris D Nugent. Partitioning time series sensor data for activity recognition. In *2009 9th international conference on information technology and applications in biomedicine*, pages 1–4. IEEE, 2009.

[100] Huiru Zheng, Haiying Wang, and Norman Black. Human activity detection in smart home environment with self-adaptive neural networks. In *2008 IEEE International Conference on Networking, Sensing and Control*, pages 1505–1510. IEEE, 2008.

[101] Fco Javier Ordóñez, José Antonio Iglesias, Paula De Toledo, Agapito Ledezma, and Araceli Sanchis. Online activity recognition using evolving classifiers. *Expert Systems with Applications*, 40(4):1248–1255, 2013.

[102] Chao Chen, Barnan Das, and Diane J Cook. A data mining framework for activity recognition in smart environments. In *2010 Sixth International Conference on Intelligent Environments*, pages 80–83. IEEE, 2010.

[103] Nicholas Foubert, Anita M McKee, Rafik A Goubran, and Frank Knoefel. Lying and sitting posture recognition and transition detection using a pressure sensor array. In *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings*, pages 1–6. IEEE, 2012.

[104] Jonathan Lester, Tanzeem Choudhury, and Gaetano Borriello. A practical approach to recognizing physical activities. In *International conference on pervasive computing*, pages 1–16. Springer, 2006.

[105] Joon-Ho Lim, Hyunchul Jang, Jaewon Jang, and Soo-Jun Park. Daily activity recognition system for the elderly using pressure sensors. In *2008 30th annual international conference of the ieee engineering in medicine and biology society*, pages 5188–5191. IEEE, 2008.

[106] Dan Ding, Rory A Cooper, Paul F Pasquina, and Lavinia Fici-Pasquina. Sensor technology for smart homes. *Maturitas*, 69(2):131–136, 2011.

[107] Xin Hong, Chris Nugent, Maurice Mulvenna, Sally McClean, Bryan Scotney, and Steven Devlin. Evidential fusion of sensor data for activity recognition in smart homes. *Pervasive and Mobile Computing*, 5(3):236–252, 2009.

[108] Matthai Philipose, Kenneth P Fishkin, Mike Perkowitz, Donald J Patterson, Dieter Fox, Henry Kautz, and Dirk Hahnel. Inferring activities from interactions with objects. *IEEE pervasive computing*, (4):50–57, 2004.

[109] Joshua R Smith, Kenneth P Fishkin, Bing Jiang, Alexander Mamishev, Matthai Philipose, Adam D Rea, Sumit Roy, and Kishore Sundara-Rajan. Rfid-based techniques for human-activity detection. *Communications of the ACM*, 48(9):39–44, 2005.

[110] Daniel H Wilson and Chris Atkeson. Simultaneous tracking and activity recognition (star) using many anonymous, binary sensors. In *International Conference on Pervasive Computing*, pages 62–79. Springer, 2005.

[111] Fco Ordóñez, Paula De Toledo, Araceli Sanchis, et al. Activity recognition using hybrid generative/discriminative models on home environments using binary sensors. *Sensors*, 13(5):5460–5477, 2013.

[112] Chandrashekhar MC Kushbu and MZ Kurian. Design and implementation of child activity recognition using accelerometer and rfid cards. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 3(4):1437–1440, 2014.

[113] Wenjie Ruan. Unobtrusive human localization and activity recognition for supporting independent living of the elderly. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 1–3. IEEE, 2016.

[114] Christian Meißner, Jürgen Meixensberger, Andreas Pretschner, and Thomas Neumuth. Sensor-based surgical activity recognition in unconstrained environments. *Minimally Invasive Therapy & Allied Technologies*, 23(4):198–205, 2014.

[115] Sara Amendola, Luigi Bianchi, and Gaetano Marrocco. Movement detection of human body segments: Passive radio-frequency identification and machine-learning technologies. *IEEE Antennas and Propagation Magazine*, 57(3):23–37, 2015.

[116] Sumi Helal, William Mann, Hicham El-Zabadani, Jeffrey King, Youssef Kaddoura, and Erwin Jansen. The gator tech smart house: A programmable pervasive space. *Computer*, 38(3):50–60, 2005.

[117] Diane J Cook, Michael Youngblood, Edwin O Heierman, Karthik Gopalratnam, Sira Rao, Andrey Litvin, and Farhan Khawaja. Mavhome: An agent-based smart home. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.(PerCom 2003).*, pages 521–524. IEEE, 2003.

[118] Beth Logan, Jennifer Healey, Matthai Philipose, Emmanuel Munguia Tapia, and Stephen Intille. A long-term evaluation of sensing modalities for activity recognition. In *International conference on Ubiquitous computing*, pages 483–500. Springer, 2007.

[119] TLM Van Kasteren, Gwenn Englebienne, and Ben JA Kröse. Activity recognition using semi-markov models on real world smart home datasets. *Journal of ambient intelligence and smart environments*, 2(3):311–325, 2010.

[120] Emmanuel Munguia Tapia, Stephen S Intille, and Kent Larson. Activity recognition in the home using simple and ubiquitous sensors. In *International conference on pervasive computing*, pages 158–175. Springer, 2004.

[121] Tim Van Kasteren, Athanasios Noulas, Gwenn Englebienne, and Ben Kröse. Accurate activity recognition in a home setting. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 1–9, 2008.

[122] Timotheus Leonhard Martinus van Kasteren et al. Activity recognition for health monitoring elderly using temporal probabilistic models. ASCI, 2011.

[123] Diane Cook, Maureen Schmitter-Edgecombe, Aaron Crandall, Chad Sanders, and Brian Thomas. Collecting and disseminating smart home sensor data in the casas project. In *Proceedings of the CHI workshop on developing shared home behavior datasets to advance HCI and ubiquitous computing research*, pages 1–7, 2009.

[124] Geetika Singla, Diane J Cook, and Maureen Schmitter-Edgecombe. Recognizing independent and joint activities among multiple residents in smart environments. *Journal of ambient intelligence and humanized computing*, 1(1):57–63, 2010.

[125] Qing Zhang, Mohan Karunanithi, Rajib Rana, and Jiajun Liu. Determination of activities of daily living of independent living older people using environmentally placed sensors. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 7044–7047. IEEE, 2013.

[126] Ali Chelli and Matthias Pätzold. A machine learning approach for fall detection and daily living activity recognition. *IEEE Access*, 7:38670–38687, 2019.

[127] Godwin Ogbuabor and Robert La. Human activity recognition for healthcare using smartphones. In *Proceedings of the 2018 10th international conference on machine learning and computing*, pages 41–46, 2018.

[128] Rani Baghezza, Kévin Bouchard, Abdenour Bouzouane, and Charles Gouin-Vallerand. From offline to real-time distributed activity recognition in wireless sensor networks for healthcare: A review. *Sensors*, 21(8):2786, 2021.

[129] UM Kamthe and CG Patil. Suspicious activity recognition in video surveillance system. In *2018 Fourth international conference on computing communication control and automation (ICCUBEA)*, pages 1–6. IEEE, 2018.

[130] Tanzila Saba, Amjad Rehman, Rabia Latif, Suliman Mohamed Fati, Mudassar Raza, and Muhammad Sharif. Suspicious activity recognition using proposed deep l4-branched-actionnet with entropy coded ant colony system optimization. *IEEE Access*, 9:89181–89197, 2021.

[131] Zehao Sun, Shaojie Tang, He Huang, Zhenyu Zhu, Hansong Guo, Yu-e Sun, and Liusheng Huang. Sos: Real-time and accurate physical assault detection using smartphone. *Peer-to-Peer Networking and Applications*, 10(2):395–410, 2017.

[132] Jonghwa Choi, Dongkyoo Shin, and Dongil Shin. Research and implementation of the context-aware middleware for controlling home appliances. *IEEE Transactions on Consumer Electronics*, 51(1):301–306, 2005.

[133] L Mary Gladence, Hari Haran Sivakumar, Gobinath Venkatesan, and S Shanmuga Priya. Home and office automation system using human activity recognition. In *2017 International conference on communication and signal processing (ICCSP)*, pages 0758–0762. IEEE, 2017.

[134] Devon Sigurdson and Eleni Stroulia. Activity recognition for smart-lighting automation at home. In *2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pages 1–6, 2018. doi: 10.1109/IISA.2018.8633667.

[135] Damien Bouchabou, Sao Mai Nguyen, Christophe Lohr, Benoit LeDuc, and Ioannis Kanellos. A survey of human activity recognition in smart homes based on iot sensors algorithms: Taxonomies, challenges, and opportunities with deep learning. *Sensors*, 21(18):6037, 2021.

[136] Emiro De-La-Hoz-Franco, Paola Ariza-Colpas, Javier Medina Quero, and Macarena Espinilla. Sensor-based datasets for human activity recognition–a systematic review of literature. *IEEE Access*, 6:59192–59210, 2018.

[137] Miguel Angel Lopez Medina, Macarena Espinilla, Cristiano Paggeti, and Javier Medina Quero. Activity recognition for iot devices using fuzzy spatio-temporal features as environmental sensor fusion. *Sensors*, 19(16):3512, 2019.

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

[139] Rex Liu, Albara Ah Ramli, Huanle Zhang, Erik Henricson, and Xin Liu. An overview of human activity recognition using wearable sensors: Healthcare and artificial intelligence. In *International Conference on Internet of Things*, pages 1–14. Springer, 2021.

[140] 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*, pages 165–186. Springer, 2011.

[141] TL Kasteren, Gwenn Englebienne, and BJ Kröse. An activity monitoring system for elderly care using generative and discriminative models. *Personal and ubiquitous computing*, 14(6):489–498, 2010.

[142] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. In *Esann*, volume 3, page 3, 2013.

[143] Jorge-L Reyes-Ortiz, Luca Oneto, Albert Samà, Xavier Parra, and Davide Anguita. Transition-aware human activity recognition using smartphones. *Neurocomputing*, 171:754–767, 2016.

[144] Roberto Luis Shinmoto Torres, Damith C Ranasinghe, and Qinfeng Shi. Evaluation of wearable sensor tag data segmentation approaches for real time activity classification in elderly. In *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*, pages 384–395. Springer, 2013.

[145] Roberto L Shinmoto Torres, Damith C Ranasinghe, Qinfeng Shi, and Alanson P Sample. Sensor enabled wearable rfid technology for mitigating the risk of falls near beds. In *2013 IEEE International Conference on RFID (RFID)*, pages 191–198. IEEE, 2013.

[146] Asanga Wickramasinghe and Damith C Ranasinghe. Recognising activities in real time using body worn passive sensors with sparse data streams: To interpolate or not to interpolate? In *proceedings of the 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services on 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 21–30, 2016.

[147] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM conference on embedded networked sensor systems*, pages 127–140, 2015.

[148] Daniela Micucci, Marco Mobilio, and Paolo Napoletano. Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences*, 7(10):1101, 2017.

[149] Mohammad Malekzadeh, Richard G Clegg, Andrea Cavallaro, and Hamed Haddadi. Protecting sensory data against sensitive inferences. In *Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems*, pages 1–6, 2018.

[150] Norbert Noury and Tareq Hadidi. Computer simulation of the activity of the elderly person living independently in a health smart home. *Computer methods and programs in biomedicine*, 108(3):1216–1228, 2012.

[151] Toufik Guettari, Jérôme Boudy, B E Benkelfat, Gérard Chollet, J L Baldinger, Pascal Doré, and Dan Istrate. Thermal signal analysis in smart home environment for detecting a human presence. In *2014 1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pages 334–339. IEEE, 2014.

[152] Muhammad Muaaz and René Mayrhofer. An analysis of different approaches to gait recognition using cell phone based accelerometers. In *Proceedings of International Conference on Advances in Mobile Computing & Multimedia*, page 293. ACM, 2013.

[153] Thorsten Rodner and Lothar Litz. Data-driven generation of rule-based behavior models for an ambient assisted living system. In *2013 IEEE Third International Conference on Consumer Electronics¿ Berlin (ICCE-Berlin)*, pages 35–38. IEEE, 2013.

[154] Nawel Yala, Belkacem Fergani, and Anthony Fleury. Feature extraction for human activity recognition on streaming data. In *2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*, pages 1–6. IEEE, 2015.

[155] Bronagh Quigley, Mark Donnelly, George Moore, and Leo Galway. A comparative analysis of windowing approaches in dense sensing environments. *Multidisciplinary Digital Publishing Institute Proceedings*, 2(19):1245, 2018.

[156] Fadi Al Machot, Heinrich C Mayr, and Suneth Ranasinghe. A windowing approach for activity recognition in sensor data streams. In *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 951–953. IEEE, 2016.

[157] Oresti Banos, Juan-Manuel Galvez, Miguel Damas, Hector Pomares, and Ignacio Rojas. Window size impact in human activity recognition. *Sensors*, 14(4):6474–6499, 2014.

[158] Qin Ni, Timothy Patterson, Ian Cleland, and Chris Nugent. Dynamic detection of window starting positions and its implementation within an activity recognition framework. *Journal of biomedical informatics*, 62:171–180, 2016.

[159] Javier Ortiz Laguna, Angel García Olaya, and Daniel Borrajo. A dynamic sliding window approach for activity recognition. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 219–230. Springer, 2011.

[160] Anzah H Niazi, Delaram Yazdansepas, Jennifer L Gay, Frederick W Maier, Lakshmish Ramaswamy, Khaled Rasheed, and Matthew P Buman. Statistical analysis of window sizes and sampling rates in human activity recognition. In *HEALTHINF*, pages 319–325, 2017.

[161] Jonathan Liono, A Kai Qin, and Flora D Salim. Optimal time window for temporal segmentation of sensor streams in multi-activity recognition. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 10–19. ACM, 2016.

[162] Minoru Yoshizawa, Wataru Takasaki, and Ren Ohmura. Parameter exploration for response time reduction in accelerometer-based activity recognition. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 653–664. ACM, 2013.

[163] Muhammad Fahim. Evolutionary learning models for indoor and outdoor human activity recognition. *Ph. D. Thesis*, 2014.

[164] Xin Hong and Chris D Nugent. Segmenting sensor data for activity monitoring in smart environments. *Personal and ubiquitous computing*, 17(3):545–559, 2013.

[165] Nida Saddaf Khan and Muhammad Sayeed Ghani. A survey of deep learning based models for human activity recognition. *Wireless Personal Communications*, 120(2): 1593–1635, 2021.

[166] Wen Qi, Hang Su, Chenguang Yang, Giancarlo Ferrigno, Elena De Momi, and Andrea Aliverti. A fast and robust deep convolutional neural networks for complex human activity recognition using smartphone. *Sensors*, 19(17):3731, 2019.

[167] Ganapati Bhat, Ranadeep Deb, Vatika Vardhan Chaurasia, Holly Shill, and Umit Y Ogras. Online human activity recognition using low-power wearable devices. In *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 1–8. IEEE, 2018.

[168] Andrey Ignatov. Real-time human activity recognition from accelerometer data using convolutional neural networks. *Applied Soft Computing*, 62:915–922, 2018.

[169] Tahmina Zebin, Matthew Sperrin, Niels Peek, and Alexander J Casson. Human activity recognition from inertial sensor time-series using batch normalized deep lstm recurrent networks. In *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 1–4. IEEE, 2018.

[170] 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*, 2018, 2018.

[171] Damla Arifoglu and Abdelhamid Bouchachia. Activity recognition and abnormal behaviour detection with recurrent neural networks. *Procedia Computer Science*, 110: 86–93, 2017.

[172] Antonio Bevilacqua, Kyle MacDonald, Aamina Rangarej, Venessa Widjaya, Brian Caulfield, and Tahar Kechadi. Human activity recognition with convolutional neural networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 541–552. Springer, 2018.

[173] Fernando Moya Rueda, René Grzeszick, Gernot A Fink, Sascha Feldhorst, and Michael Ten Hompel. Convolutional neural networks for human activity recognition using body-worn sensors. In *Informatics*, volume 5, page 26. MDPI, 2018.

[174] Shuochao Yao, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek Abdelzaher. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In *Proceedings of the 26th international conference on world wide web*, pages 351–360, 2017.

[175] Seyed Ali Rokni, Marjan Nourollahi, and Hassan Ghasemzadeh. Personalized human activity recognition using convolutional neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[176] Taylor R Mauldin, Marc E Canby, Vangelis Metsis, Anne HH Ngu, and Coralys Cubero Rivera. Smartfall: A smartwatch-based fall detection system using deep learning. *Sensors*, 18(10):3363, 2018.

[177] Yong Zhang, Yu Zhang, Zhao Zhang, Jie Bao, and Yunpeng Song. Human activity recognition based on time series analysis using u-net. *arXiv preprint arXiv:1809.08113*, 2018.

[178] Kun Wang, Jun He, and Lei Zhang. Attention-based convolutional neural network for weakly labeled human activities' recognition with wearable sensors. *IEEE Sensors Journal*, 19(17):7598–7604, 2019.

[179] Abdulmajid Murad and Jae-Young Pyun. Deep recurrent neural networks for human activity recognition. *Sensors*, 17(11):2556, 2017.

[180] Schalk Wilhelm Pienaar and Reza Malekian. Human activity recognition using lstm-rnn deep neural network architecture. In *2019 IEEE 2nd wireless africa conference (WAC)*, pages 1–5. IEEE, 2019.

[181] Jian Sun, Yongling Fu, Shengguang Li, Jie He, Cheng Xu, and Lin Tan. Sequential human activity recognition based on deep convolutional network and extreme learning machine using wearable sensors. *Journal of Sensors*, 2018, 2018.

[182] Harish Haresamudram, David V Anderson, and Thomas Plötz. On the role of features in human activity recognition. In *Proceedings of the 23rd International symposium on wearable computers*, pages 78–88, 2019.

[183] Maryam Banitalebi Dehkordi, Abolfazl Zaraki, and Rossitza Setchi. Feature extraction and feature selection in smartphone-based activity recognition. *Procedia Computer Science*, 176:2655–2664, 2020.

[184] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[185] Hu Min and Wu Fangfang. Filter-wrapper hybrid method on feature selection. In *2010 Second WRI Global Congress on Intelligent Systems*, volume 3, pages 98–101. IEEE, 2010.

[186] Praneeth Vepakomma, Debraj De, Sajal K Das, and Shekhar Bhansali. A-wristocracy: Deep learning on wrist-worn sensing for recognition of user complex activities. In *2015 IEEE 12th International conference on wearable and implantable body sensor networks (BSN)*, pages 1–6. IEEE, 2015.

[187] Wanmin Wu, Sanjoy Dasgupta, Ernesto E Ramirez, Carlyn Peterson, Gregory J Norman, et al. Classification accuracies of physical activities using smartphone motion sensors. *Journal of medical Internet research*, 14(5):e2208, 2012.

[188] Aiguo Wang, Guilin Chen, Jing Yang, Shenghui Zhao, and Chih-Yung Chang. A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sensors Journal*, 16(11):4566–4578, 2016.

[189] Ferhat Attal, Samer Mohammed, Mariam Dedabrishvili, Faicel Chamroukhi, Latifa Oukhellou, and Yacine Amirat. Physical human activity recognition using wearable sensors. *Sensors*, 15(12):31314–31338, 2015.

[190] Sadiq Sani, Stewart Massie, Nirmalie Wiratunga, and Kay Cooper. Learning deep and shallow features for human activity recognition. In *International conference on knowledge science, engineering and management*, pages 469–482. Springer, 2017.

[191] Yan Wang, Shuang Cang, and Hongnian Yu. A data fusion-based hybrid sensory system for older people's daily activity and daily routine recognition. *IEEE Sensors Journal*, 18(16):6874–6888, 2018.

[192] Frédéric Li, Kimiaki Shirahama, Muhammad Adeel Nisar, Lukas Köping, and Marcin Grzegorzek. Comparison of feature learning methods for human activity recognition using wearable sensors. *Sensors*, 18(2):679, 2018.

[193] Annemarie Laudanski, Brenda Brouwer, and Qingguo Li. Activity classification in persons with stroke based on frequency features. *Medical engineering & physics*, 37 (2):180–186, 2015.

[194] Kazuya Murao and Tsutomu Terada. A recognition method for combined activities with accelerometers. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 787–796, 2014.

[195] Jozsef Suto, Stefan Oniga, and Petrica Pop Sitar. Feature analysis to human activity recognition. *International Journal of Computers Communications & Control*, 12(1): 116–130, 2016.

[196] Bobak Jack Mortazavi, Mohammad Pourhomayoun, Gabriel Alsheikh, Nabil Alshurafa, Sunghoon Ivan Lee, and Majid Sarrafzadeh. Determining the single best axis for exercise repetition recognition and counting on smartwatches. In *2014 11th international conference on wearable and implantable body sensor networks*, pages 33–38. IEEE, 2014.

[197] Mi Zhang and Alexander A Sawchuk. A feature selection-based framework for human activity recognition using wearable multimodal sensors. In *BodyNets*, pages 92–98, 2011.

[198] Bo Wei, Rebeen Ali Hamad, Longzhi Yang, Xuan He, Hao Wang, Bin Gao, and Wai Lok Woo. A deep-learning-driven light-weight phishing detection sensor. *Sensors*, 19(19):4258, 2019.

[199] Jianbo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiao Li Li, and Shonali Krishnaswamy. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Twenty-fourth international joint conference on artificial intelligence*, 2015.

[200] Md Abdullah Al Hafiz Khan, Nirmalya Roy, and Archan Misra. Scaling human activity recognition via deep learning-based domain adaptation. In *2018 IEEE international conference on pervasive computing and communications (PerCom)*, pages 1–9. IEEE, 2018.

[201] Carlos Avilés-Cruz, Andrés Ferreyra-Ramírez, Arturo Zúñiga-López, and Juan Villegas-Cortéz. Coarse-fine convolutional deep-learning strategy for human activity recognition. *Sensors*, 19(7):1556, 2019.

[202] Panagiotis Kasnesis, Charalampos Z Patrikakis, and Iakovos S Venieris. Perceptionnet: A deep convolutional neural network for late sensor fusion. In *Proceedings of SAI Intelligent Systems Conference*, pages 101–119. Springer, 2018.

[203] Charlene V San Buenaventura, Nestor Michael C Tiglao, and Rowel O Atienza. Deep learning for smartphone-based human activity recognition using multi-sensor fusion. In *International Wireless Internet Conference*, pages 65–75. Springer, 2018.

[204] Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115, 2016.

[205] Alireza Abedin, S Hamid Rezatofighi, Qinfeng Shi, and Damith C Ranasinghe. Sparsesense: Human activity recognition from highly sparse sensor data-streams using set-based neural networks. *arXiv preprint arXiv:1906.02399*, 2019.

[206] Daniele Ravi, Charence Wong, Benny Lo, and Guang-Zhong Yang. Deep learning for human activity recognition: A resource efficient implementation on low-power devices. In *2016 IEEE 13th international conference on wearable and implantable body sensor networks (BSN)*, pages 71–76. IEEE, 2016.

[207] Bandar Almaslukh, Abdel Monim Artoli, and Jalal Al-Muhtadi. A robust deep learning approach for position-independent smartphone-based human activity recognition. *Sensors*, 18(11):3726, 2018.

[208] Heeryon Cho and Sang Min Yoon. Divide and conquer-based 1d cnn human activity recognition using test data sharpening. *Sensors*, 18(4):1055, 2018.

[209] Mingtao Dong, Jindong Han, Yuan He, and Xiaojun Jing. Har-net: Fusing deep representation and hand-crafted features for human activity recognition. In *International Conference On Signal And Information Processing, Networking And Computers*, pages 32–40. Springer, 2018.

[210] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[211] Pradeep Hewage, Ardhendu Behera, Marcello Trovati, Ella Pereira, Morteza Ghahremani, Francesco Palmieri, and Yonghuai Liu. Temporal convolutional neural (tcn) network for an effective weather forecasting using time-series data from the local weather station. *Soft Computing*, 24(21):16453–16482, 2020.

[212] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553): 436–444, 2015.

[213] Geoffrey Hinton and Terrence J Sejnowski. *Unsupervised learning: foundations of neural computation*. MIT press, 1999.

[214] 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*, pages 216–223. Springer, 2012.

[215] KG Manosha Chathuramali and Ranga Rodrigo. Faster human activity recognition with svm. In *International conference on advances in ICT for emerging regions (ICTer2012)*, pages 197–203. IEEE, 2012.

[216] Adithyan Palaniappan, R Bhargavi, and V Vaidehi. Abnormal human activity recognition using svm based approach. In *2012 international conference on recent trends in information technology*, pages 97–102. IEEE, 2012.

[217] Jun Yang. Toward physical activity diary: motion recognition using simple acceleration features with mobile phones. In *Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics*, pages 1–10, 2009.

[218] Adil Mehmood Khan, Ali Tufail, Asad Masood Khattak, and Teemu H Laine. Activity recognition on smartphones via sensor-fusion and kda-based svms. *International Journal of Distributed Sensor Networks*, 10(5):503291, 2014.

[219] Timothy Sohn, Alex Varshavsky, Anthony LaMarca, Mike Y Chen, Tanzeem Choudhury, Ian Smith, Sunny Consolvo, Jeffrey Hightower, William G Griswold, and Eyal de Lara. Mobility detection using everyday gsm traces. In *International Conference on Ubiquitous Computing*, pages 212–224. Springer, 2006.

[220] Hong Lu, Jun Yang, Zhigang Liu, Nicholas D Lane, Tanzeem Choudhury, and Andrew T Campbell. The jigsaw continuous sensing engine for mobile phone applications. In *Proceedings of the 8th ACM conference on embedded networked sensor systems*, pages 71–84, 2010.

[221] Shumei Zhang, Paul McCullagh, Chris Nugent, and Huiru Zheng. Activity monitoring using a smart phone's accelerometer with hierarchical classification. In *2010 sixth international conference on intelligent environments*, pages 158–163. IEEE, 2010.

[222] Barnan Das, Adriana M Seelye, Brian L Thomas, Diane J Cook, Larry B Holder, and Maureen Schmitter-Edgecombe. Using smart phones for context-aware prompting in smart environments. In *2012 IEEE Consumer Communications and Networking Conference (CCNC)*, pages 399–403. IEEE, 2012.

[223] Jun-geun Park, Ami Patel, Dorothy Curtis, Seth Teller, and Jonathan Ledlie. Online pose classification and walking speed estimation using handheld devices. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 113–122, 2012.

[224] Quang Viet Vo, Minh Thang Hoang, and Deokjai Choi. Personalization in mobile activity recognition system using k-medoids clustering algorithm. *International Journal of Distributed Sensor Networks*, 9(7):315841, 2013.

[225] E Ramanujam, Thinagaran Perumal, and S Padmavathi. Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review. *IEEE Sensors Journal*, 21(12):13029–13040, 2021.

[226] John D Kelleher. *Deep learning*. MIT press, 2019.

[227] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[228] Dan Ciregan, Ueli Meier, and Jürgen Schmidhuber. Multi-column deep neural networks for image classification. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3642–3649. IEEE, 2012.

[229] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[230] Alex Graves. Supervised sequence labelling. In *Supervised sequence labelling with recurrent neural networks*, pages 5–13. Springer, 2012.

[231] Ivan Nunes Da Silva, Danilo Hernane Spatti, Rogerio Andrade Flauzino, Luisa Helena Bartocci Liboni, and Silas Franco dos Reis Alves. Artificial neural networks. *Cham: Springer International Publishing*, 39, 2017.

[232] Kouichi Murakami and Hitomi Taguchi. Gesture recognition using recurrent neural networks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 237–242, 1991.

[233] Peter Vamplew and Anthony Adams. Recognition and anticipation of hand motions using a recurrent neural network. 1995.

[234] Yun Long, Eui Min Jung, Jaeha Kung, and Saibal Mukhopadhyay. Reram crossbar based recurrent neural network for human activity detection. In *2016 international joint conference on neural networks (IJCNN)*, pages 939–946. IEEE, 2016.

[235] Michele Alessandrini, Giorgio Biagetti, Paolo Crippa, Laura Falaschetti, and Claudio Turchetti. Recurrent neural network for human activity recognition in embedded systems using ppg and accelerometer data. *Electronics*, 10(14):1715, 2021.

[236] Gonzalo Bailador, Daniel Roggen, Gerhard Tröster, and Gracián Triviño. Real time gesture recognition using continuous time recurrent neural networks. In *BodyNets*, page 15, 2007.

[237] Longfei Zheng, Shuai Li, Ce Zhu, and Yanbo Gao. Application of indrnn for human activity recognition: The sussex-huawei locomotion-transportation challenge. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pages 869–872, 2019.

[238] Pankaj Khatiwada, Matrika Subedi, Ayan Chatterjee, and Martin Wulf Gerdes. Automated human activity recognition by colliding bodies optimization-based optimal feature selection with recurrent neural network. *arXiv preprint arXiv:2010.03324*, 2020.

[239] Xufei Wang, Weixian Liao, Yifan Guo, Lixing Yu, Qianlong Wang, Miao Pan, and Pan Li. Perrnn: Personalized recurrent neural networks for acceleration-based human activity recognition. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2019.

[240] Mingqi Lv, Wei Xu, and Tieming Chen. A hybrid deep convolutional and recurrent neural network for complex activity recognition using multimodal sensors. *Neurocomputing*, 362:33–40, 2019.

[241] István Ketykó, Ferenc Kovács, and Krisztián Zsolt Varga. Domain adaptation for semg-based gesture recognition with recurrent neural networks. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2019.

[242] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[243] Yuwen Chen, Kunhua Zhong, Ju Zhang, Qilong Sun, and Xueliang Zhao. Lstm networks for mobile human activity recognition. In *2016 International conference on artificial intelligence: technologies and applications*, pages 50–53. Atlantis Press, 2016.

[244] Deepika Singh, Erinc Merdivan, Ismini Psychoula, Johannes Kropf, Sten Hanke, Matthieu Geist, and Andreas Holzinger. Human activity recognition using recurrent neural networks. In *International cross-domain conference for machine learning and knowledge extraction*, pages 267–274. Springer, 2017.

[245] Seungeun Chung, Jiyoun Lim, Kyoung Ju Noh, Gague Kim, and Hyuntae Jeong. Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning. *Sensors*, 19(7):1716, 2019.

[246] Shilong Yu and Long Qin. Human activity recognition with smartphone inertial sensors using bidir-lstm networks. In *2018 3rd international conference on mechanical, control and computer engineering (icmcce)*, pages 219–224. IEEE, 2018.

[247] Mohib Ullah, Habib Ullah, Sultan Daud Khan, and Faouzi Alaya Cheikh. Stacked lstm network for human activity recognition using smartphone data. In *2019 8th European workshop on visual information processing (EUVIP)*, pages 175–180. IEEE, 2019.

[248] LuKun Wang and RuYue Liu. Human activity recognition based on wearable sensor using hierarchical deep lstm networks. *Circuits, Systems, and Signal Processing*, 39 (2):837–856, 2020.

[249] Preeti Agarwal and Mansaf Alam. A lightweight deep learning model for human activity recognition on edge devices. *Procedia Computer Science*, 167:2364–2373, 2020.

[250] Xiaokang Zhou, Wei Liang, I Kevin, Kai Wang, Hao Wang, Laurence T Yang, and Qun Jin. Deep-learning-enhanced human activity recognition for internet of healthcare things. *IEEE Internet of Things Journal*, 7(7):6429–6438, 2020.

[251] Nafiul Rashid, Berken Utku Demirel, and Mohammad Abdullah Al Faruque. Ahar: Adaptive cnn for energy-efficient human activity recognition in low-power edge devices. *IEEE Internet of Things Journal*, 2022.

[252] Sara Ashry, Tetsuji Ogawa, and Walid Gomaa. Charm-deep: Continuous human activity recognition model based on deep neural network using imu sensors of smartwatch. *IEEE Sensors Journal*, 20(15):8757–8770, 2020.

[253] Sara Ashry Mohammed, Reda Elbasiony, and Walid Gomaa. An lstm-based descriptor for human activities recognition using imu sensors. In *ICINCO (1)*, pages 504–511, 2018.

[254] Walid Gomaa, Reda Elbasiony, and Sara Ashry. Adl classification based on autocorrelation function of inertial signals. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 833–837. IEEE, 2017.

[255] Kunihiko Fukushima and Sei Miyake. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern recognition*, 15(6): 455–469, 1982.

[256] Maxime W Lafarge, Erik J Bekkers, Josien PW Pluim, Remco Duits, and Mitko Veta. Roto-translation equivariant convolutional networks: Application to histopathology image analysis. *Medical Image Analysis*, 68:101849, 2021.

[257] Xiao-Xiao Niu and Ching Y Suen. A novel hybrid cnn–svm classifier for recognizing handwritten digits. *Pattern Recognition*, 45(4):1318–1325, 2012.

[258] Di-Xiu Xue, Rong Zhang, Hui Feng, and Ya-Lei Wang. Cnn-svm for microvascular morphological type recognition with data augmentation. *Journal of medical and biological engineering*, 36(6):755–764, 2016.

[259] Abien Fred Agarap. An architecture combining convolutional neural network (cnn) and support vector machine (svm) for image classification. *arXiv preprint arXiv:1712.03541*, 2017.

[260] Haifeng Wu, Qing Huang, Daqing Wang, and Lifu Gao. A cnn-svm combined model for pattern recognition of knee motion using mechanomyography signals. *Journal of Electromyography and Kinesiology*, 42:136–142, 2018.

[261] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[262] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

[263] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[264] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[265] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[266] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[267] Md Zahangir Alom, Tarek M Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Mahmudul Hasan, Brian C Van Essen, Abdul AS Awwal, and Vijayan K Asari. A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3):292, 2019.

[268] Deepika Singh, Erinc Merdivan, Sten Hanke, Johannes Kropf, Matthieu Geist, and Andreas Holzinger. Convolutional and recurrent neural networks for activity recognition in smart environment. In *Towards integrative machine learning and knowledge extraction*, pages 194–205. Springer, 2017.

[269] Yujia Qin, Fanchao Qi, Sicong Ouyang, Zhiyuan Liu, Cheng Yang, Yasheng Wang, Qun Liu, and Maosong Sun. Improving sequence modeling ability of recurrent neural networks via sememes. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:2364–2373, 2020.

[270] Yuqing Chen and Yang Xue. A deep learning approach to human activity recognition based on single accelerometer. In *2015 IEEE international conference on systems, man, and cybernetics*, pages 1488–1492. IEEE, 2015.

[271] Yin Tang, Qi Teng, Lei Zhang, Fuhong Min, and Jun He. Layer-wise training convolutional neural networks with smaller filters for human activity recognition using wearable sensors. *IEEE Sensors Journal*, 21(1):581–592, 2020.

[272] Xin Cheng, Lei Zhang, Yin Tang, Yue Liu, Hao Wu, and Jun He. Real-time human activity recognition using conditionally parametrized convolutions on mobile and wearable devices. *IEEE Sensors Journal*, 22(6):5889–5901, 2022.

[273] Federico Cruciani, Anastasios Vafeiadis, Chris Nugent, Ian Cleland, Paul McCullagh, Konstantinos Votis, Dimitrios Giakoumis, Dimitrios Tzovaras, Liming Chen, and Raouf Hamzaoui. Feature learning for human activity recognition using convolutional neural networks. *CCF Transactions on Pervasive Computing and Interaction*, 2(1): 18–32, 2020.

[274] Mark Nutter, Catherine H Crawford, and Jorge Ortiz. Design of novel deep learning models for real-time human activity recognition with mobile phones. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.

[275] Ran Zhu, Zhuoling Xiao, Ying Li, Mingkun Yang, Yawen Tan, Liang Zhou, Shuisheng Lin, and Hongkai Wen. Efficient human activity recognition solving the confusing activities via deep ensemble learning. *Ieee Access*, 7:75490–75499, 2019.

[276] Seyed Vahab Shojaedini and Mohamad Javad Beirami. Mobile sensor based human activity recognition: distinguishing of challenging activities by applying long short-term memory deep learning modified by residual network concept. *Biomedical Engineering Letters*, 10(3):419–430, 2020.

[277] Song-Mi Lee, Sang Min Yoon, and Heeryon Cho. Human activity recognition from accelerometer data using convolutional neural network. In *2017 ieee international conference on big data and smart computing (bigcomp)*, pages 131–134. IEEE, 2017.

[278] Fanyi Xiao, Ling Pei, Lei Chu, Danping Zou, Wenxian Yu, Yifan Zhu, and Tao Li. A deep learning method for complex human activity recognition using virtual wearable sensors. In *International Conference on Spatial Data and Intelligence*, pages 261–270. Springer, 2020.

[279] Kun Xia, Jianguang Huang, and Hanyu Wang. Lstm-cnn architecture for human activity recognition. *IEEE Access*, 8:56855–56866, 2020.

[280] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

[281] Wen Qi, Hang Su, and Andrea Aliverti. A smartphone-based adaptive recognition and real-time monitoring system for human activities. *IEEE Transactions on Human-Machine Systems*, 50(5):414–423, 2020.

[282] Huaijun Wang, Jing Zhao, Junhuai Li, Ling Tian, Pengjia Tu, Ting Cao, Yang An, Kan Wang, and Shancang Li. Wearable sensor-based human activity recognition using hybrid deep learning techniques. *Security and communication Networks*, 2020, 2020.

[283] Tianqi Lv, Xiaojuan Wang, Lei Jin, Yabo Xiao, and Mei Song. Margin-based deep learning networks for human activity recognition. *Sensors*, 20(7):1871, 2020.

[284] Debadyuti Mukherjee, Riktim Mondal, Pawan Kumar Singh, Ram Sarkar, and Debotosh Bhattacharjee. Ensemconvnet: a deep learning approach for human activity recognition using smartphone sensors for healthcare applications. *Multimedia Tools and Applications*, 79(41):31663–31690, 2020.

[285] Tongtong Su, Huazhi Sun, Chunmei Ma, Lifen Jiang, and Tongtong Xu. Hdl: Hierarchical deep learning model based human activity recognition using smartphone sensors. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019.

[286] Qingchang Zhu, Zhenghua Chen, and Yeng Chai Soh. A novel semisupervised deep learning method for human activity recognition. *IEEE Transactions on Industrial Informatics*, 15(7):3821–3830, 2018.

[287] Jun He, Qian Zhang, Liqun Wang, and Ling Pei. Weakly supervised human activity recognition from wearable sensors by recurrent attention learning. *IEEE Sensors Journal*, 19(6):2287–2297, 2018.

[288] Wenbin Gao, Lei Zhang, Qi Teng, Jun He, and Hao Wu. Danhar: Dual attention network for multimodal human activity recognition using wearable sensors. *Applied Soft Computing*, 111:107728, 2021.

[289] Wei Wei, Jinjiu Li, Longbing Cao, Yuming Ou, and Jiahang Chen. Effective detection of sophisticated online banking fraud on extremely imbalanced data. *World Wide Web*, 16(4):449–475, 2013.

[290] Matthew Herland, Taghi M Khoshgoftaar, and Richard A Bauder. Big data fraud detection using multiple medicare data sources. *Journal of Big Data*, 5(1):1–21, 2018.

[291] David A Cieslak, Nitesh V Chawla, and Aaron Striegel. Combating imbalance in network intrusion datasets. In *GrC*, pages 732–737. Citeseer, 2006.

[292] Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural networks*, 106: 249–259, 2018.

[293] Shuo Wang and Xin Yao. Multiclass imbalance problems: Analysis and potential solutions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(4):1119–1130, 2012.

[294] Haibo He and Edwardo A Garcia. Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, 21(9):1263–1284, 2009.

[295] Nathalie Japkowicz. The class imbalance problem: Significance and strategies. In *Proc. of the Int'l Conf. on Artificial Intelligence*, volume 56, pages 111–117. Citeseer, 2000.

[296] Justin M Johnson and Taghi M Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):1–54, 2019.

[297] Fayez Alharbi, Lahcen Ouarbya, and Jamie A Ward. Comparing sampling strategies for tackling imbalanced data in human activity recognition. *Sensors*, 22(4):1373, 2022.

[298] Jason Van Hulse, Taghi M Khoshgoftaar, and Amri Napolitano. Experimental perspectives on learning from imbalanced data. In *Proceedings of the 24th international conference on Machine learning*, pages 935–942, 2007.

[299] Inderjeet Mani and I Zhang. knn approach to unbalanced data distributions: a case study involving information extraction. In *Proceedings of workshop on learning from imbalanced datasets*, volume 126, pages 1–7. ICML, 2003.

[300] Miroslav Kubat, Stan Matwin, et al. Addressing the curse of imbalanced training sets: one-sided selection. In *Icml*, volume 97, page 179. Citeseer, 1997.

[301] Ricardo Barandela, Rosa M Valdovinos, J Salvador Sánchez, and Francesc J Ferri. The imbalanced training sample problem: Under or over sampling? In *Joint IAPR international workshops on statistical techniques in pattern recognition (SPR) and structural and syntactic pattern recognition (SSPR)*, pages 806–814. Springer, 2004.

[302] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[303] Hui Han, Wen-Yuan Wang, and Bing-Huan Mao. Borderline-smote: a new over-sampling method in imbalanced data sets learning. In *International conference on intelligent computing*, pages 878–887. Springer, 2005.

[304] Ky Trung Nguyen, François Portet, and Catherine Garbay. Dealing with imbalanced data sets for human activity recognition using mobile phone sensors. In *3rd International Workshop on Smart Sensing Systems*, 2018.

[305] Chumphol Bunkhumpornpat, Krung Sinapiromsaran, and Chidchanok Lursinsap. Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 475–482. Springer, 2009.

[306] Shoujin Wang, Wei Liu, Jia Wu, Longbing Cao, Qinxue Meng, and Paul J Kennedy. Training deep neural networks on imbalanced data sets. In *2016 international joint conference on neural networks (IJCNN)*, pages 4368–4374. IEEE, 2016.

[307] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[308] Haishuai Wang, Zhicheng Cui, Yixin Chen, Michael Avidan, Arbi Ben Abdallah, and Alexander Kronzer. Predicting hospital readmission via cost-sensitive deep learning. *IEEE/ACM transactions on computational biology and bioinformatics*, 15(6):1968–1978, 2018.

[309] Chong Zhang, Kay Chen Tan, and Ruoxu Ren. Training cost-sensitive deep belief networks on imbalance data problems. In *2016 international joint conference on neural networks (IJCNN)*, pages 4362–4367. IEEE, 2016.

[310] Salman H Khan, Munawar Hayat, Mohammed Bennamoun, Ferdous A Sohel, and Roberto Togneri. Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE transactions on neural networks and learning systems*, 29(8): 3573–3587, 2017.

[311] Mikel Galar, Alberto Fernandez, Edurne Barrenechea, Humberto Bustince, and Francisco Herrera. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4):463–484, 2011.

[312] Zhi-Hua Zhou. *Ensemble methods: foundations and algorithms*. CRC press, 2012.

[313] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3(1):9, 2016.

[314] Derek Hao Hu, Vincent Wenchen Zheng, and Qiang Yang. Cross-domain activity recognition via transfer learning. *Pervasive and Mobile Computing*, 7(3):344–358, 2011.

[315] Yongxin Yang and Timothy M Hospedales. A unified perspective on multi-domain and multi-task learning. *arXiv preprint arXiv:1412.7489*, 2014.

[316] Yunsheng Li and Nuno Vasconcelos. Efficient multi-domain learning by covariance normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5424–5433, 2019.

[317] Carl Doersch and Andrew Zisserman. Multi-task self-supervised visual learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2051–2060, 2017.

[318] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Colorization as a proxy task for visual understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6874–6883, 2017.

[319] Basura Fernando, Hakan Bilen, Efstratios Gavves, and Stephen Gould. Self-supervised video representation learning with odd-one-out networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3636–3645, 2017.

[320] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *European conference on computer vision*, pages 213–226. Springer, 2010.

[321] Jing Jiang. Multi-task transfer learning for weakly-supervised relation extraction. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 1012–1020, 2009.

[322] Trevor Standley, Amir R Zamir, Dawn Chen, Leonidas Guibas, Jitendra Malik, and Silvio Savarese. Which tasks should be learned together in multi-task learning? *arXiv preprint arXiv:1905.07553*, 2019.

[323] Antonio Torralba, Kevin P Murphy, and William T Freeman. Sharing visual features for multiclass and multiview object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):854–869, 2007.

[324] Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang. Learning deep feature representations with domain guided dropout for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1249–1258, 2016.

[325] Mahesh Joshi, Mark Dredze, William Cohen, and Carolyn Rose. Multi-domain learning: when do domains matter? In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1302–1312, 2012.

[326] Alice Schoenauer-Sebag, Louise Heinrich, Marc Schoenauer, Michele Sebag, Lani F Wu, and Steve J Altschuler. Multi-domain adversarial learning. *arXiv preprint arXiv:1903.09239*, 2019.

[327] Jindong Wang, Vincent W Zheng, Yiqiang Chen, and Meiyu Huang. Deep transfer learning for cross-domain activity recognition. In *proceedings of the 3rd International Conference on Crowd Science and Engineering*, pages 1–8, 2018.

[328] Jindong Wang, Yiqiang Chen, Lisha Hu, Xiaohui Peng, and S Yu Philip. Stratified transfer learning for cross-domain activity recognition. In *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–10. IEEE, 2018.

[329] Aaqib Saeed, Tanir Ozcelebi, and Johan Lukkien. Multi-task self-supervised learning for human activity detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(2):1–30, 2019.

[330] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[331] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[332] Zhen Peng, Yixiang Dong, Minnan Luo, Xiao-Ming Wu, and Qinghua Zheng. Self-supervised graph representation learning via global context prediction. *arXiv preprint arXiv:2003.01604*, 2020.

[333] Harish Haresamudram, Apoorva Beedu, Varun Agrawal, Patrick L Grady, Irfan Essa, Judy Hoffman, and Thomas Plötz. Masked reconstruction based self-supervision for human activity recognition. In *Proceedings of the 2020 International Symposium on Wearable Computers*, pages 45–49, 2020.

[334] Bulat Khaertdinov, Esam Ghaleb, and Stylianos Asteriadis. Contrastive self-supervised learning for sensor-based human activity recognition. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2021.

[335] Harish Haresamudram, Irfan Essa, and Thomas Plötz. Contrastive predictive coding for human activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(2):1–26, 2021.

[336] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[337] Huatao Xu, Pengfei Zhou, Rui Tan, Mo Li, and Guobin Shen. Limu-bert: Unleashing the potential of unlabeled data for imu sensing applications. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, pages 220–233, 2021.

[338] Chi Ian Tang, Ignacio Perez-Pozuelo, Dimitris Spathis, Soren Brage, Nick Wareham, and Cecilia Mascolo. Selfhar: Improving human activity recognition through self-training with unlabeled data. *arXiv preprint arXiv:2102.06073*, 2021.

[339] Wei Niu, Jiao Long, Dan Han, and Yuan-Fang Wang. Human activity detection and recognition for video surveillance. In *2004 IEEE International Conference on Multimedia and Expo (ICME)(IEEE Cat. No. 04TH8763)*, volume 1, pages 719–722. IEEE, 2004.

[340] Duckki Lee and Sumi Helal. From activity recognition to situation recognition. In *International Conference on Smart Homes and Health Telematics*, pages 245–251. Springer, 2013.

[341] Daniele Liciotti, Michele Bernardini, Luca Romeo, and Emanuele Frontoni. A sequential deep learning application for recognising human activities in smart homes. *Neurocomputing*, 396:501–513, 2020.

[342] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.

[343] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.

[344] Krzysztof Choromanski, Valerii Likhosherstov, David Dohan, Xingyou Song, Andreea Gane, Tamas Sarlos, Peter Hawkins, Jared Davis, Afroz Mohiuddin, Lukasz Kaiser, et al. Rethinking attention with performers. *arXiv preprint arXiv:2009.14794*, 2020.

[345] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 18661–18673. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/d89a66c7c80a29b1bdbab0f2a1a94af8-Paper.pdf.

[346] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9268–9277, 2019.

[347] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[348] Ming Zeng, Le T Nguyen, Bo Yu, Ole J Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang. Convolutional neural networks for human activity recognition using mobile sensors. In *6th international conference on mobile computing, applications and services*, pages 197–205. IEEE, 2014.

[349] Yu Guan and Thomas Plötz. Ensembles of deep lstm learners for activity recognition using wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(2):1–28, 2017.

[350] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[351] Saif Mahmud, M. T. H. Tonmoy, Kishor Kumar Bhaumik, A. M. Rahman, M. A. Amin, M. Shoyaib, Muhammad Asif Hossain Khan, and A. Ali. Human activity recognition from wearable sensor data using self-attention. In *ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain*, 2020.

[352] Jiayi Zhu, Ying Tan, Rude Lin, Jiaqing Miao, Xuwei Fan, Yafei Zhu, Ping Liang, Jinnan Gong, and Hui He. Efficient self-attention mechanism and structural distilling model for alzheimer's disease diagnosis. *Computers in Biology and Medicine*, 147: 105737, 2022.

[353] Iram Fatima, Muhammad Fahim, Young-Koo Lee, and Sungyoung Lee. Analysis and effects of smart home dataset characteristics for daily life activity recognition. *The Journal of Supercomputing*, 66(2):760–780, 2013.

[354] Luyang Jing, Taiyong Wang, Ming Zhao, and Peng Wang. An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox. *Sensors*, 17(2):414, 2017.

[355] Liang Cao, Yufeng Wang, Bo Zhang, Qun Jin, and Athanasios V Vasilakos. Gchar: An efficient group-based context—aware human activity recognition on smartphone. *Journal of Parallel and Distributed Computing*, 118:67–80, 2018.

[356] Aditya Devarakonda, Maxim Naumov, and Michael Garland. Adabatch: Adaptive batch sizes for training deep neural networks. *arXiv preprint arXiv:1712.02029*, 2017.

[357] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[358] Aaron Fisher, Cynthia Rudin, and Francesca Dominici. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20(177): 1–81, 2019.

[359] Christoph Molnar. *Interpretable Machine Learning*. Lulu. com, 2020.

[360] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[361] Zeeshan Ahmad and Naimul Mefraz Khan. Multidomain multimodal fusion for human action recognition using inertial sensors. In *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, pages 429–434. IEEE, 2019.

[362] Yajing Liu, Xinmei Tian, Ya Li, Zhiwei Xiong, and Feng Wu. Compact feature learning for multi-domain image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7193–7201, 2019.

[363] Hao Peng, Nikolaos Pappas, Dani Yogatama, Roy Schwartz, Noah A Smith, and Lingpeng Kong. Random feature attention. *arXiv preprint arXiv:2103.02143*, 2021.

[364] Terry T Um, Franz MJ Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 216–220, 2017.

[365] Irwan Bello. Lambdanetworks: Modeling long-range interactions without attention. *arXiv preprint arXiv:2102.08602*, 2021.

[366] Wenbo Huang, Lei Zhang, Qi Teng, Chaoda Song, and Jun He. The convolutional neural networks training with channel-selectivity for human activity recognition based on sensors. *IEEE Journal of Biomedical and Health Informatics*, 25(10):3834–3843, 2021.

[367] Carlos Betancourt, Wen-Hui Chen, and Chi-Wei Kuan. Self-attention networks for human activity recognition using wearable devices. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1194–1199. IEEE, 2020.

[368] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[369] Jeremy Appleyard, Tomas Kocisky, and Phil Blunsom. Optimizing performance of recurrent neural networks on gpus. *arXiv preprint arXiv:1604.01946*, 2016.

[370] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681, 1997.

[371] Alex Graves and Jürgen Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, 18(5-6):602–610, 2005.

[372] Fabio Hernández, Luis F Suárez, Javier Villamizar, and Miguel Altuve. Human activity recognition on smartphones using a bidirectional lstm network. In *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, pages 1–5. IEEE, 2019.