

**DUBLIN CITY UNIVERSITY**

**SCHOOL OF ELECTRONIC ENGINEERING**

**A Project Template**

**and Report Writing Guide**

**Mark Kane**

August 2019

BACHELOR OF ENGINEERING

IN

ELECTRONIC SYSTEMS

Supervised by Prof. N. O’Connor

# Acknowledgements

# Declaration

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations set out in the module documentation. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

I have read and understood the DCU Academic Integrity and Plagiarism at <https://www4.dcu.ie/sites/default/files/policy/1%20-%20integrity_and_plagiarism_ovpaa_v3.pdf> and IEEE referencing guidelines found at https://loop.dcu.ie/mod/url/view.php?id=448779.

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

Table of Contents

[Acknowledgements ii](#_Toc443488856)

[Declaration ii](#_Toc443488857)

[Abstract iii](#_Toc443488858)

[Chapter 1 - Introduction 1](#_Toc443488859)

[1.1 Equipment and Software 2](#_Toc443488860)

[1.1.1 The Intelligent Camera (ICAM) 2](#_Toc443488861)

[1.2 Writing an academic report 2](#_Toc443488862)

[1.2.1 Issues in Technical Writing 2](#_Toc443488863)

[1.2.2 Figures and Diagrams 6](#_Toc443488864)

[1.2.3 The Document Structure 7](#_Toc443488865)

[1.2.4 Summary 9](#_Toc443488866)

[Chapter 2 - Technical Background 10](#_Toc443488867)

[2.1 Using Heading Styles 10](#_Toc443488868)

[2.1.1 Heading 1 Style 10](#_Toc443488869)

[2.1.2 Heading 2 Style 11](#_Toc443488870)

[2.1.3 Heading 3 Style 11](#_Toc443488871)

[2.1.4 Other Headings 11](#_Toc443488872)

[2.2 Summary 11](#_Toc443488873)

[Chapter 3 - Design of … 12](#_Toc443488874)

[Chapter 4- Implementation and Testing of … 13](#_Toc443488875)

[Chapter 5 - Results and Discussion 14](#_Toc443488876)

[Chapter 6 – Ethics 15](#_Toc443488877)

[Chapter 7 - Conclusions and Further Research 16](#_Toc443488878)

[References 17](#_Toc443488879)

[Appendix 1 18](#_Toc443488880)

[Glossary 19](#_Toc443488881)

Table of Figures

# Chapter 1 - Introduction

## 1.1 Human Activity Recognition

Human activity recognition (HAR) is the term coined to define an ability to interpret a gesture or motion of the human body, and from this make a determination of the human activity or action being performed [1]. Recognizing a human’s activity, automatically, has become a significant problem in ubiquitous computing, human-computer interaction and human behavioural analysis [2]. Three scientific research contexts HAR finds itself an important component of are: surveillance, healthcare and human computer interaction [1].

### 1.1.1 Applications of HAR

**Surveillance**

HAR has been adopted in surveillance systems at public places, i.e. airports, banks, etc [1]. The findings in [3] confirm that proposed approaches are able to recognize ongoing human-human interactions at the earlier stage [1]. Furthermore, Legion: AR, a system proposed by [4], supplements existing recognition systems with on-demand, real-time activity identification to produce robust, deployable activity recognition [1].

**Healthcare**

The literature indicates that HAR has so far been employed in healthcare systems introduced in hospitals, rehabilitation centres, and even in residential environments [1]. HAR has extraordinary potential within the healthcare sector with a wide range of applications already seen today and is of particular use for aiding the elderly and vulnerable of society. By monitoring the activities of elderly people cared for in rehabilitation centres, HAR can be an effective way of monitoring chronic diseases as well as aiding disease prevention [1]. HAR is also utilised for monitoring patients at home to measure aspects of daily living such as energy expenditure to assist obesity prevention, treatment [5] and lifelogging [1]. Another use case of HAR is its application in monitoring stereotypical motion conditions suffered by children with Autism Spectrum Disorder (ASD) [6]. HAR can also be used for monitoring other behaviours such as those derived from abnormal conditions in cardiac patients [7], and the detection of early signs of illness [8], providing clinicians with an alarm mechanism for early intervention. More healthcare related HAR like fall detection is demonstrated by [9].

**Human Computer Interaction**

The human computer interaction field has seen the most well-known introductions of HAR, being adopted by gaming and exergaming such as Kinect and the Nintendo Wii [1]. The gestures and movements recognized through HAR are used by the machine to carry out specific tasks. HAR is also used in full-body motion-based games for older adults with neurological injury [1].

## 1.2 Types of HAR

There are two main types of HAR: video-based and sensor-based HAR [10]. More specifically, the sensing technologies used in HAR can be classified into the following three categories: RGB camera-based, depth sensor-based and wearable sensor-based [1]. These sensing technologies are discussed further in chapter two. The following figure illustrates the general structure of a HAR system for any of these sensing technologies:

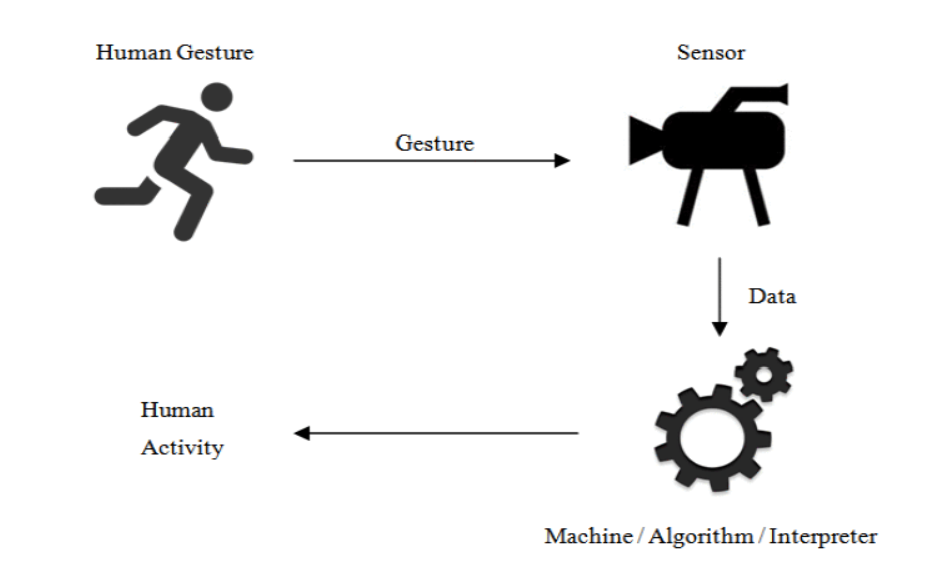


Figure 1: Structure of HAR system

Due to the rapid advancements in ubiquitous computing and the growth in concern for privacy protection, sensor-based HAR via wearable devices has seen itself rise in popularity. Because of this, and the reasons documented in chapter two, this project focuses on sensor-based HAR using wearable devices.

## 1.3 HAR: Machine Learning vs Deep Learning

HAR has most commonly been treated as a pattern recognition (PR) task, and to much avail. Machine learning algorithms such as the support vector machine have produced very satisfying results, most notably in controlled environments and where the volume of labelled data is low [10]. These machine learning methods do, however, rely on heuristic hand-crafted feature extraction which are limited to human domain knowledge. Hand-crafted features are hard to compute and difficult to scale [11], and therefore only shallow features can be learned using this type of approach [10]. In relation to HAR, training an accurate classifier is a particularly difficult task due to the high inter- and intra-class variability of human actions [11]. In other words, humans often perform different tasks differently to each other [11], which makes classification more difficult. Furthermore, using shallow, hand-crafted features reduces the performance of unsupervised learning methods [10]. Therefore, it is clear there are limited classification accuracies and model generalization available to the conventional pattern recognition approaches associated with the HAR problem [10].

The recent surge of pervasive computing has brought with it a rapid development and advancement of deep learning, achieving real success in a variety of domains [12]. In contrast to the conventional PR methods, deep learning is capable of learning high level features which are far more meaningful than the shallow features designed manually in conventional PR methods [10]. Convolutional neural networks (CNNs) have become very popular for their high accuracy in image classification [12]. By training a CNN, feature extraction and classification are combined into an encapsulated, end-to-end approach [11]. With respect to wearable-sensor based HAR, the features extracted through non-linear transformations are learned directly from the raw data, being more discriminative between the different human activity classes than those crafted in PR methods [11]. A drawback to deep learning, however, is its necessity for large amounts of labelled data to train a deep network. Although there are now many publicly available datasets particularly in the domain of object recognition, classification, detection and captioning such as ImageNet and MS-COCO, alternative tasks with a more specific scope may struggle to find labelled datasets on which to train their networks.

This report is structured as follows. Chapter two discusses the three different sensing technologies used in HAR detailing the advantages and disadvantages to each. Chapter two also covers some of the most common approaches to the HAR problem in the domain of time-series data generated from wearable sensors, and depicts the challenges faced by this domain of HAR and the suitability of CNNs for time-series data. Chapter two finally defines the problem being addressed by this project, which experiments with different adaptations to time-series data input to a suitable CNN architecture. Chapter three details the dataset used and describes the baseline approach taken by this project to the HAR problem. It then also details the design of the experiments, which attempt to improve upon the results obtained by the baseline approach. Chapter four is dedicated to the implementation details of this project, the results of which are discussed in chapter five. Chapter six is a piece about the ethics which are adhered to by this project. Chapter seven then draws some conclusions on the work documented by this project and discusses possible future work.

# Chapter 2 - Technical Background & Literature Review

## 2.1 Sensing Technologies

In order to choose a suitable sensing technology on which to design a HAR system, it is useful to discuss each of them and their pros and cons. The following three sensing technologies are discussed below: RGB video, depth sensors, wearable sensors.

### 2.1.1 RGB Video

Video-based HAR accepts videos/images capturing human motion as the input. Where the

camera is attached to the environment, the HAR system processes the image sequences using supervised learning. This involves the previous trainings of a system prior to its application [1]. This training stage consists of feeding image sequences along with the names of the human activity performed in those images into the system [1]. Feature extraction and classification are central components to the conventional HAR systems seen using this sensing technology [1] [\*13,48\*]. The RGB camera has been less preferred to other sensing technologies in the research of HAR [1]. There are a few reasons for this. Firstly, its ability to capture human motion in 3D space is limited [1] [\*39\*]. The performance of a real time HAR system may also suffer due to the high computational cost in the machine processing required to extract human movements from an image [1] [\*28\*] [\*5\*]. The biggest concern, however, regarding the employment of RGB camera data by a HAR system is the issue of privacy, i.e. users of such a system may not be comfortable with consenting to being recorded at all times [1].

### 2.1.2 Depth Sensor

A depth sensor can be referred to as an infrared sensor [1] [\*49\*]. This sends an infrared beam into a scene and recaptures it using the infrared sensor. The depth of an object can be calculated using the distance travelled by the beam [1]. Microsoft’s Kinect sensor is commonly employed as a depth sensor in HAR, according to reviews in [1] [\*33\*]. The Microsoft Kinect sensor is capable of detecting twenty human body joints using a real-world coordinate system [1] [\*40\*], which makes it very suitable for use in the classification of human movements. The literature is adamant that depth sensors are an improvement upon the RGB camera as a sensing technology [1]. The low cost of depth sensors has contributed to a rise in its popularity as a sensing technology for HAR [1] [\*19,20,27,37,39\*]. So too has its high sampling rate and capability of combining visual and depth information [1] [\*10\*]. Furthermore, the recognition processes appear to be far less expensive in comparison to those derived under RGB camera data [1] [\*10\*]. This said, depth sensors have so far been unable to solve the limitation of sensor viewpoint [1] [\*30\*] and sensor obstruction faced by the RGB camera [1][\*39\*].

### 2.1.3 Wearable Sensor

HAR via the use of wearable sensors requires only the subject wearing single or multiple wearables on their body [1]. The sensors typically found in these wearables are 3-dimensional accelerometer, gyroscope and magnetometers as well as temperature sensors. The emergence of wearables from smartphones (most smartphones now come with an accelerometer, gyroscope and magnetometer) to fitbits has opened many avenues for HAR. Wearables can also overcome the barriers of limited sensor viewpoint and sensor obstruction already discussed for RGB cameras and depth sensors. Additionally, wearables have the potential to act as a non-obtrusive [1] [\*26\*] solution to HAR as they have already proven to be effective in monitoring other aspects of daily living without infringing on the user’s quality of daily living. Further advantages of wearable sensors are their low monetary cost and high power efficiency [1] [\*46\*]. There are still some challenges facing wearable sensors, however. The accuracy of activity recognition using a wearable sensor may not be sufficient as current wearable sensor-based HAR systems require subjects to wear multiple sensors across different body parts [11] [1][\*23\*]. This is undoubtedly a sub-optimal solution to HAR as it is quite inconvenient and mildly intrusive for a subject to be required to wear multiple sensors across different body parts.

## 2.3 Sensor Modalities

Most HAR approaches work with just one specific type of sensor only. [10](Chavarriaga et al., 2013) suggests that we classify the sensor modalities into the following categories: *body-worn sensors, object sensors, ambient sensors.* These sensor modalities are further detailed below:

**Body-worn Sensor**

The body-worn sensor is of the most commonly used sensor modalities in HAR. As their name suggests, these sensors are most often worn by the user and include accelerometers, gyroscopes and magnetometers. These sensors are widely distributed throughout ubiquitous computing and are now very often found in watches, smart phones, bands, helmets and glasses. Body-worn sensors are very frequently used for deep learning based HAR [2] [11] [13] [14]. This related work shows that body-worn sensors are primarily adopted for recognizing activities of daily living (ADL) and sport. One notable point regarding the use of body-worn sensors in deep learning based HAR is that it is the original signal produced by the sensors that is used as the input to the network, as opposed to the approach of traditional machine learning methods where the inputs are statistical/frequency features extracted from the motion data.

**Object Sensor**

Object sensors are different to body-worn sensors in that they infer human activity by detecting the movement of a certain object [10](Chavarriagaetal.,2013). Therefore, object sensors are placed on specific objects to detect their movement. For example, the activity of drinking water can be detected by placing an accelerometer on a cup, [10]. Radio frequency identifier (RFID) tags are often used and mainly employed in smart homes [10] (Vepakomma et al., 2015; Yang et al., 2015; Fang and Hu, 2014) and medical facilities [10] (Li et al., 2016b; Wang et al., 2016a). RFID tags are used as they can produce more fine-grained information aiding in the recognition of more complex activities [10]. Object sensors are not as popular as body-worn sensors for HAR as they are difficult to deploy in real-world environments, however, there is an emergence of using object sensors combined with other types for the recognition of more complex activities [10] (Yang, 2009).

**Ambient Sensor**

In contrast to object sensors which measure an object’s movement, ambient sensors attempt to capture changes in the environments. Human activity is inferred by capturing the interaction between humans and the environment. Radar, pressure sensors, sound sensors, and temperature sensors are all different examples of ambient sensors. The work with ambient sensors indicates they are usually embedded in users’ smart environment (smart home) and are used to recognize activities of daily living and hand gestures [10] (Lane et al., 2015; Wang et al., 2016a; Kim and Toomajian, 2016). Like object sensors, the deployment of ambient sensors is not easy and so they are often less preferred to other sensor types. Another drawback to ambient sensors is that they are very easily affected by the environment, and therefore, only specific types of activities are accurately inferable by ambient sensors.

**Hybrid Sensor**

The combination of sensor types has been shown to potentially improve the accuracy in HAR. Ambient sensors used alongside object sensors are advantageous as they provide information on object movements as well as the state of the environment. In [10] (Vepakomma et al., 2015), a smart home environment is designed in which a wide range of complex and fine-grained activities of multiple occupants are recognizable via the use of body-worn, object and ambient sensors. It is clear that using multiple sensor modalities can yield superior information of human activities. The work also shows that it is a deployable solution to HAR in certain environments like smart homes.

## 2.4 CNNs and HAR

Convolutional Neural Networks have been used extensively in the field of computer vision for their effective feature extraction and pattern recognition abilities. They have been widely adopted by deep learning and have produced very promising results in image classification, speech recognition and text analysis [10]. CNNs are of particular use in image classification as convolution leverages three important ideas: sparse interactions, parameter sharing and equivariant representations [15] [\*23\*], [datadriveninvestor?]. With these properties, CNNs can extract smaller features of greater significance, whilst at the same time reducing the storage requirements compared to those more densely connected neural networks [15]. In addition to this, the depth and breadth of convolutions are adjustable, making it easier to train CNNs compared to alternative feedforward neural networks [15]. CNNs do, however, raise the concern of overfitting the training set, particularly in image recognition. If overfitting has occurred, it means that the CNN is also learning the obsolete background features of the training images, aiding the classification accuracy of the training set [15]. This is not desirable as it means that the trained model will not generalise well, i.e. it has failed to capture the more general characteristics of the training set and so, on unseen data, the model will not produce a classification accuracy close to that found on the training set.

When used for the classification of time-series data, like in HAR, CNNs have two advantages over other models: local dependency and scale invariance [10], [14]. Local dependency refers to the correlation between nearby signals [10]. Scales invariance means the output is not sensitive to a variance in paces or frequencies in the input. When using CNNs as a solution to the HAR problem, the concepts of *input adaptation, pooling* and *weight sharing* must be considered [10].

### 2.4.1 Input Adaptation for CNNs

The sensors used in HAR, obviously, do not produce image data. These sensors produce time-series readings. The most commonly gathered data in wearable sensor-based HAR are accelerometer, gyroscope and magnetometer time-series signals. More specifically, these are 3-axial (x, y and z direction) 1D readings along the temporal dimension [10]. These readings must be adapted first before a CNN can be applied to them. The inputs must first be adapted to form a *virtual image* [10]. There are two ways of achieving this: *data-driven* and *model-driven*.

* **Data-driven:** This approach is straightforward and easy to implement. Each dimension is treated as a channel on which a 1D convolution is performed, i.e. each 1D sensor is treated as a 1D image [10]. Although uncomplicated, this approach does have a drawback: it does not account for dependencies between dimensions and sensors, which may impact performance [10].
* **Model-driven:** This approach looks to apply a 2D convolution, and so resizes the inputs into a 2D virtual image [10]. The advantage of this approach is that it can incorporate the temporal dependencies between sensors [10]. This conversion from time-series to image data is, however, a non-trivial task and requires domain knowledge [10].

### 2.4.2 Pooling

Pooling is often paired with convolution in CNNs. The most common types of pooling are max or average pooling. Convolution is so often combined with pooling operations to avoid overfitting [10]. Each time an input image is pooled, the image size is reduced (usually to half the size). Therefore, pooling helps to reduce the training time on large training sets as it reduces the number of parameters to be tuned [10] (bengio, 2013).

### 2.4.3 Weight Sharing

## 2.6 Common Approaches to HAR

(Few take a machine vision approach – only found one paper)

Many different approaches have been taken to the HAR problem and many works tackle HAR from different angles. [16] explores, compares and contrasts the state-of-the-art deep learning methods for HAR using wearable sensors regarding performance. [16] gives great insights into the suitability of each model with respect to specific tasks in HAR. The three main deep learning architectures assessed in [16] are Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs) and regular Deep Feed-Forward Neural Networks (DNNs). It is observed from [16] RNNs that outperform CNNs significantly on activities that are short in duration but have a natural ordering. It is clear that the performance of a recurrent approach benefits largely from its ability to contextualise observations across long periods of time. It is, however, recommended that CNNs are used for more prolonged and repetitive activities such as walking/running. Although some RNNs were found to perform similarly to or even better than CNNs in this environment, the *average* performance of CNNs in this domain suggest a higher likelihood that the practitioner finds a suitable configuration using a CNN architecture. Furthermore, through the experiments in [16] it is recommended to explore different learning rates prior to optimizing the network architecture. This is due to the experiments in [16] demonstrating that a change in learning parameters has the largest impact on performance. The work with DNNs in [16] indicate that they require a significant investment in the exploration of parameters, showing a notable spread between peak and median performance. It is concluded from [16] that the use of CNNs or RNNs are a more sophisticated approach to DNNs, showing a smaller spread in performance and a higher likelihood of finding a configuration that works sufficiently, particularly in the case of CNNs.

[2] focuses on the HAR problem where the inputs are multi-channel time-series measurements gathered from inertial measurement units worn on different parts of the body, and the outputs are predefined human activities, e.g. walking, running, sitting etc. The architecture used by [2] is a deep CNN, automating feature learning from the raw inputs. As so often used in HAR, labelling the inputs via supervised learning produces learned features with more discriminative power [2]. [2] uses benchmark datasets like the Opportunity Activity Recognition dataset and Hand Gesture dataset to verify that the unique advantages to CNNs propel their CNN architecture to outperform other state-of-the-art machine learning solutions to HAR such as SVM, KNN, Means and Variance, and Deep Belief Network.

Similar to [2], [11] focuses on the HAR problem where the inputs are multi-channel time-series measurements. A novel CNN is proposed by [11] in which the network used processes the sequence measurements from different body-worn devices separately. This follows the idea of a wider rather than a deeper network, where the architecture consists of multiple branches, each processing the data from one specific inertial device only. This is done to provide some robustness against the inertial devices being asynchronous or having slightly different characteristics. With each branch providing an IMU-specific, intermediate representation of the data, a global representation is found via fully connected layers. 1D temporal convolutions and pooling operations are performed on the input’s sequences. [11] evaluates the proposed architecture on three benchmark datasets including the Opportunity Activity Recognition and Pamap datasets, outperforming the state-of-the-art. [11] advises that the capabilities of CNNs are improved by applying convolutions per sensor and per body worn device.

[17] proposes the use of 2D kernels in both convolutional and pooling layers to capture spatial dependency over sensors in addition to local dependency over time (already provided by the temporal convolution). [17] demonstrates a high performance of this multi-modal CNN, compared to state-of-the-art methods by experimenting on benchmark datasets.

As in [17], [18] applies 2D convolution and pooling to capture both spatial and local dependency. However, it is not the raw time-series data which is passed into the network. [18] carries out pre-processing steps to adapt the time-series input sequences into a 2D virtual image. It first uses an algorithm to create an image of sensor channels where each sensor channel is placed either above or below all other sensor channels at least once. This is done to allow the deep CNN to extract hidden correlations between neighbouring signals. This generates a new signal image. The DFT of each channel in this signal image is found to generate a new activity image. The activity image is the input which is passed to the architecture. According to [18], this approach outperforms the state-of-the-art in terms of recognition accuracy and computational cost.

[19] presents CNNs for multi-modal data (multi-sensor, e.g. accelerometers and gyroscopes) where it introduces both partial weight sharing and full weight sharing to the CNN models in a manner that both modality-specific characteristics and common characteristics across modalities are learned from multi-modal data. The results in [19] indicate that weight sharing can improve the performance of a CNN.

[15] takes an alternative approach to HAR with the use of photoplethysmography sensor data gathered from wearable devices. [15] looks to simplify the wearable approach to HAR by experimenting with wrist-mounted optical sensing (used usually for heart rate determination) to see if it can provide data useful for activity recognition. In contrast to many of the papers discussed, [15] takes a machine vision approach to the HAR problem, using the plots of the optical signals to produce activity classifications. More specifically, [15] implements transfer learning to retrain the penultimate layer of a pretrained CNN, using time-series *images* of the photoplethysmography signals as the inputs to the network. [15] achieves an average accuracy of 75.8%, which is a competitive result, suggesting that the overall design of activity monitoring and classification systems could be simplified to wearables based on optical measurements only. However, implementing a design of optical sensor only in an activity classification system does lead to a trade off in classification performance.

As can be seen, there are many different approaches to the HAR problem, with many variations in architectures and how the data is represented. After a careful examination of the literature on HAR, it can be concluded that the use of convolutional neural networks is the most effective network used for HAR to date.

## 2.7 Modality Transformation

Modality transformation refers to the conversion of data from a source mode to a target mode [12].

# Chapter 3 - Design of …

# Chapter 4- Implementation and Testing of …

# Chapter 5 - Results and Discussion

# Chapter 6 – Ethics

# Chapter 7 - Conclusions and Further ResearchReferences

|  |  |
| --- | --- |
| [1] | O. C. Ann and L. B. Theng, “Human Activity Recognition: A Review,” in *2014 IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2014)*, Batu Ferringhi, Malaysia, 28-30 Nov. 2014. |
| [2] | J. Bo Yang, M. N. Nguyen, P. P. San and X. K. S. Li Li, “Deep Convolutional Neural Networks On Multichannel Time Series For Human Activity Recognition,” in *Twenty-Fourth Joint Conference on Artificial Intelligence 2015*, 2015. |
| [3] | M. S. Ryoo, “Human activity prediction: Early recognition of ongoing activities from streaming videos,” in *IEEE International Conference on Computer Vision (ICCV)*, Barcelona, Spain, 2011. |
| [4] | W. S. Lasecki, Y. C. Song, H. Kautz and J. P. Bigham, “Real-time crowd labeling for deployable activity recognition,” in *2013 conference on Computer supported cooperative work-CSCW 13*, San Antonio, Texas, 2013. |
| [5] | E. S. Sazonov, G. Fulk, J. Hill, Y. Schutz and R. Browning, “Monitoring of posture allocations and activities by a shoe-based wearable sensor,” *IEEE Transactions on Biomedical Engineering,* vol. 58, no. 4, pp. 983-990, 2011. |
| [6] | G. Paragliola and A. Coronato, “Intelligent Monitoring of Stereotyped Motion Disorders in Case of Children with Autism,” in *2013 9th International Conference on Intelligent Environments - IEEE*, Athens, 2013. |
| [7] | E. Kańtoch and P. Augustyniak, “Human activity surveillance based on wearable body sensor network,” in *2012 Computing in Cardiology*, Krakow, 2012. |
| [8] | E. E. Stone and M. . Skubic, “Passive, in-home gait measurement using an inexpensive depth camera: Initial results,” , 2012. [Online]. Available: http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.ieee-000006240383. [Accessed 21 3 2019]. |
| [9] | A. T. Nghiem, E. . Auvinet and J. . Meunier, “Head detection using Kinect camera and its application to fall detection,” , 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6310538. [Accessed 21 3 2019]. |
| [10] | J. Wang, Y. Chen, S. Hao, X. Peng and L. Hu, “Deep Learning for Sensor-based Activity Recognition: A Survey,” *Pattern Recognition Letters, 2019 - Elsevier,* no. 119, pp. 3-11, 2019. |
| [11] | F. Moya Rueda, R. G. Grzeszick, G. A. Fink, S. Feldhorst and M. ten Hompel, “Convolutional Neural Networks for Human Activity Recognition Using Body-Worn Sensors,” *Informatics,* Vols. Informatics 2018, 5, 26, 2018. |
| [12] | M. Singh, V. Pondenkandath, B. Zhou, P. Lukowicz and M. Liwickit, “Transforming sensor data to the image domain for deep learning—An application to footstep detection,” in *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, 2017. |
| [13] | J. Yang, M. Nguyen, P. San, X. Li and S. Krishnaswamy, “Deep convolutional neural networks on multichannel time series for human activity recognition,” in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015. |
| [14] | M. Zeng, L. Nguyen, B. Yu, O. Mengshoel, J. Zhu, P. Wu and J. Zhang, “Convolutional neural networks for human activity recognition using mobile sensors,” in *6th International Conference on Mobile Computing, Applications and Services*, 2014. |
| [15] | E. Brophy, J. Dominguez, Z. Wang and T. Ward, “A machine vision approach to human activity recognition using photoplethysmograph sensor data,” in *2018 29th Irish Signals and Systems Conference (ISSC)*, 2018. |
| [16] | N. Hammerla, S. Halloran and T. Plötz, “Deep, convolutional, and recurrent models for human activity recognition using wearables,” in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2016. |
| [17] | S. Ha, J. Yun and S. Choi, “Multi-Modal Convolutional Neural Networks for Activity Recognition,” in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 2015. |
| [18] | W. Jiang and Z. Yin, “Human activity recognition using wearable sensors by deep convolutional neural networks,” in *23rd ACM international conference on Multimedia*, 2015. |

# Appendix 1

# Glossary