ACTIVITY MONITORING IN ASSISTED LIVING ENVIRONMENTS

MS (RESEARCH) THESIS

By

ANIL

DISCIPLINE OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JULY 2021

ACTIVITY MONITORING IN ASSISTED LIVING ENVIRONMENTS

A THESIS

Submitted in the fulfilment of the requirements for the award of the degree
Of

Master of Science (Research)

By

ANIL

DISCIPLINE OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JULY 2021

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECALARATION

I hereby certify that the work which is being presented in the thesis entitled **Activity Monitoring in Assisted Living Environments** in the fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE (RESEARCH)** and submitted in the **DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**, **Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from July 2019 to June 2021 under the supervision of Dr. Abhishek Srivastava, Associate Professor, Indian Institute of Technology Indore, India.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Anil
05-07-2021

Signature of the Student with Date

(ANIL)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Abhishek Srivastava
5/7/2021

Signature of Thesis Supervisor with Date

(Dr. Abhishek Srivastava)

Anil has successfully given her MS (Research) Oral Examination held on October 25, 2021

Signature of Chairperson, OEB
Date: 25/10/2021

Abhishek Srivastava
Signature of Thesis Supervisor
Date: 25/10/2021

Signature of Convener, DPGC
Date: 25-Oct-2021

Somnath Dey
Signature of Head of Department
Date: 25/10/2021

ACKNOWLEDGEMENTS

It gives me immense pleasure to take the opportunity to express my heartful gratitude towards my esteemed supervisor **Dr. Abhishek Srivastava**, who made this time learnable, enjoyable while being a constant source of inspiration during my work. Without his constant guidance and research directions, support, patience, immense knowledge and encouragement, this research work could not be completed. I am grateful to have worked under his supervision and guidance during my post grad studies.

I am sincerely thankful to **Dr. Somanth Dey** and **Dr. Bodhisatwa Mazumdar** my research progress committee members for taking out some valuable time to evaluate my progress all these years. Their valuable comments and suggestions helped me to improve my work at various stages.

I wish to express my sincere gratitude and thanks to **Dr. Somnath Dey**, HOD of Computer Science and **Dr. Neminath Hubballi**, DPGC Convener for his extended support and guidance.

My sincere acknowledgement and respect to **Prof. Neelesh Kumar Jain**, Director, Indian Institute of Technology Indore for providing me the opportunity to explore my research capabilities at Indian Institute of Technology Indore.

I would like to express my heartfelt respect to my parents, my elder sisters and my younger brother for their love, care and support they have provided to me throughout my life. I'd want to express my gratitude for the wonderful company of my best colleagues and friends, Uttkarsh, Saumya, Anup, Fyse, Shatendra, Ajay, Namrata.

Anil

To my family and readers

Abstract

Deep learning has recently emerged as a promising alternative in Human Activity Recognition (HAR) scenarios in ubiquitous and wearable sensors. The HAR enables us to employ a number of applications such as forecasting activity, health care systems by monitoring their behaviour, assisting the elderly or dementia patients, and so on. Therefore, there are numerous machine learning algorithms which are used in the Human Activity Recognition(HAR) such as, support vector machine, random forests, hidden markov model and so on. The problem with the traditional machine learning model is that we need to extract features manually. It requires expert domain knowledge and a very tedious task—The primary concern with adopting deep learning is that we do not need to extract features explicitly. Problems with real-world datasets, most notably imbalanced datasets and poor data quality, continue to limit the accuracy of activity recognition.

The processed data is separated into overlapping, fixed-size panes. This window gives input to the 1D CNN and extracts the features from each window, which are then provided to the LSTM network. This thesis proposed a one-dimensional Convolution neural network (or CNN) long-short term memory or LSTM (or 1dCNNLSTM) model that recognises the activity from sensor data input. The 1dCNN is used as a feature learning from processed input data sequence. And LSTM maintains the information regarding or maintains the temporal features between sequences. Following that, the output features are fed into a series of fully connected layers for final activity recognition. The model's performance has been tested on the Opportunity dataset, a very complex dataset recorded in a rich sensors environment. The proposed model performance is tremendous in Human Activity Recognition and verified in the experiment chapter.

Keywords: Assisted living, CNN, LSTM, Confusion matrix, ROC curves, Activity Recognition

CONTENTS

List of Figures

List of Tables

Chapter 1

Introduction

Technology has made our lives increasingly convenient over the years. A very good example of this is Smart Homes , an approach to home automation that provides us with security, comfort, ease, energy conservation. A specialised form of Smart Homes are Assisted Living Environments (ALE). These are living spaces for senior citizens and the differently abled that are endowed with various types of sensors (wearable, camera, motion sensor, pressure sensor, ambient sensor, and so on) most of which are integrated to normal every day use objects and appliances. ALEs most importantly monitor the condition and well-being of residents. In addition to this, ALEs contribute significantly towards enhancing the quality and discipline of life by reminding residents of activities to be performed, assisting in more effectively conducting daily chores, and also providing appropriate suggestions for getting this done. ALEs are especially useful for people suffering from dementia, Alzheimer's disease, or any other memory disorder through regular reminders on tasks such as taking medications, having breakfast, daily exercise, and others.

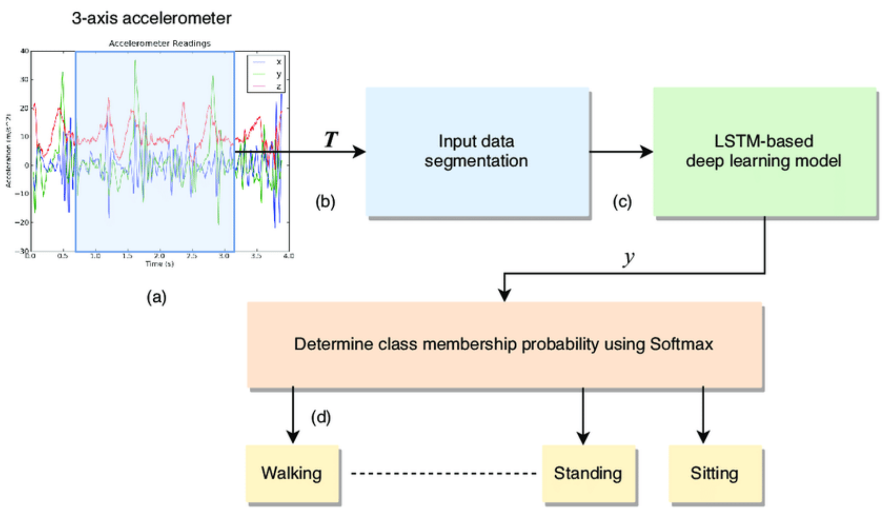

Fig 1.1: HAR deep learning model for activity recognition

Figure 1.1 depicts the HAR deep learning model and its many phases. The first step is to examine data preprocessing and segmenting the data samples into a fixed window size. The input is fed to the HAR model, which extracts characteristics from sensor readings. Finally, the output is predicted using the softmax activation function to identify the activity based on the class with the highest likelihood.

One approach in ALEs is the use of vision sensors to keep track of residents. While this is effective and numerous studies have been conducted to test the efficacy of vision sensors in ALEs, the big drawback with these is that the privacy of residents is compromised. , This is why sensors play such an important role in effective monitoring of ALEs.

Human Activity Recognition (or HAR) is the category of study that deals with recognising or detecting residents' activities in ALEs. Human activity comprises prehistoric activities such as standing, sitting, walking, and daily living activities such as eating, cooking, grooming, sleeping to name a few. . In this work, we study the activities of a single resident or subject. We utilise machine learning techniques on sensor data with the intent of effectively monitoring and recognising activities of a resident.

We collect data from a collection of heterogeneous sensors and employ various Machine Learning techniques such as
Bayesian Networks, Conditional Random Fields, Hidden Markov Models, Convolutions Neural Networks (CNN), Long Short Term Memory (LSTM) networks to effectively monitor and recognise Activities of Daily Living (ADL).

The main challenge with sensor datasets is the inherently noisy nature of the input. It is rather difficult to work upon sensor data sets than image or video data sets because it is difficult to analyse and conclude by looking at the numerical values. However, it is easy to understand the activity performed in videos or images. Several models, including Bayesian networks, conditional random fields, Hidden Markov Models, CNN, LSTM, and others, have been used to recognise ADLs from heterogeneous sensors.

1.1 Background

Sensors are in extensive use nowadays, from little gadgets such as in mobile phones to colossal planes. Various sensors are available such as accelerometers, gyroscopes, ambient sensors, torque sensors, temperature sensors, pressure sensors, to name just a few. ALEs constitute an important application are where sensors are actively employed. In an ALE, the primary purpose of sensors is to track the movement of occupants and recognise the activities performed. Human Activity Recognition (HAR) is accomplished in two stages.

The first stage involves

- handling raw sensor data to execute various preprocessing procedures,
- dividing the dataset into sliding window sizes of a defined length, and
- splitting each time series into equal-length segments. A window size of 10 seconds is used in most works in literature [11], [12],[13], [14], [15].

The second stage involves

- extracting features from raw segments and,
- running the activity classification algorithm.

The second stage is critical since the quality of features extracted from the segments determines the overall accuracy of the HAR system. One approach that is widely used in literature is to rely on hand-crafted features, that require expert domain knowledge. Hand-crafted features include spectral entropy, minimum value, maximum value, auto-regressive, FFT coefficients. The other preferred technique is deep learning-based that automatically learns the required features from the data. The key benefit of this approach is that there is no need for manual feature engineering as deep learning learns and trains in an end-to-end manner.

In this work, we propose to recognise the current activity of a person in an ALE and also need to analyse his/her behaviour over a period of time and draw conclusions on his/her condition. To do this, we require not just a model that is effective in identifying activities but also one that is able to keep track of the identified activities over a period of time. Thus, in terms of sensor readings, the time series patterns of the readings need to be retained. CNN is mostly unable to retain such time series patterns in memory. Sequential models are good candidates for maintaining such time-series patterns. Recurrent Neural Networks (RNN) is one such sequential model that can efficiently perform such tasks. Long Short Term Memory (LSTM) is a special type of RNN architecture that is widely used in time-series processing. In our work, therefore, we use a hybrid model, comprising both a CNN and sequential LSTM model [16].

1.2 Thesis contribution

The main contributions of the thesis are as follows:

The thesis uses a combined form of two neural networks CNN and LSTM, creating a hybrid model that is able to effectively monitor and identify the activities of occupants of an Assisted Living Environment.

The hybrid nature of the model allows it to follow activities over time and form inferences about a person's state based on continuous monitoring and observation. The proposed model is validated using a publicly available real-world dataset with a wide range of sensors.

The hybrid techniques reduces the computational complexity of traditional HAR approaches while retaining high accuracy. When compared to 2D-CNN, 1D-CNN works more effectively in resource-constrained contexts with limited resources.

1.3 Organisation of the Thesis

The thesis is organised into five chapters. A summary of each chapter is provided below:

Chapter 1 (Introduction)

This chapter describes the background details of activity recognition on sensor datasets, and contributions of the thesis.

Chapter 2 (Literature Survey)

This chapter provides the detailed literature survey and various activity recognition techniques.

Chapter 3 (Monitoring Of Well Being Occupants)

This chapter explains convolution neural network, 1-dimensional convolution neural network, and Recurrent neural network and its variant. In later part of this chapter explains the combine version of hybrid model.

Chapter 4 (Experiment Result and Analysis)

In this chapter, we describe the dataset that is used to evaluate the model's performance. The Confusion matrices and ROC curves are plotted in the second part of the chapter to evaluated the efficacy of the proposed model.

Chapter 5 (Conclusion and Future Work)

This chapter concludes the thesis with pointers to future work.

Chapter 2

Literature Survey

Human Activity Recognition (HAR) encompasses a wide range of activities including behaviour analysis, monitoring, and gesture recognition. HAR, as shown in Figure 2.1, is done primarily using camera and sensor data

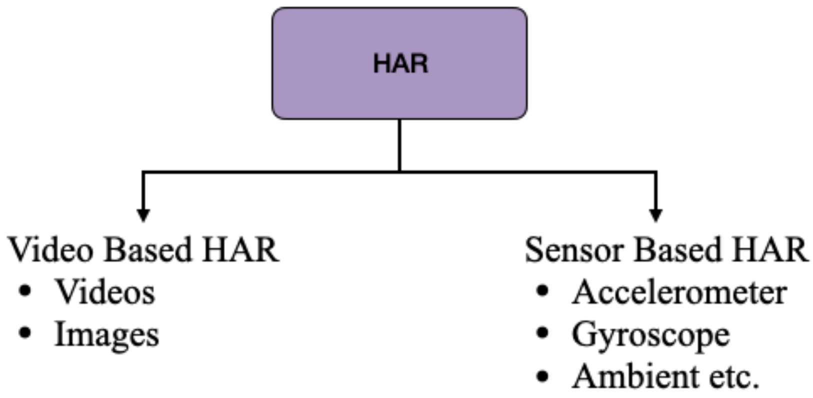

Figure 2.1: Types of HAR based on the Datasets

Camera datasets comprise sequences of images or videos that are used to recognise the activity performed [17][18][19]. On the other hand, sensor data sets have several types of values, such as binary-valued sensors and real-valued sensors. Therefore we have to track the motion based on a rather heterogeneous dataset [6][7][8][9]. In this thesis, our primary emphasis is on sensor-based HAR because of their privacy-respecting working mechanism.

The sensors are not restricted to HAR, and may be expanded to monitor health conditions using sensor data such as EEG, ECG, and so on [16].

2.1 HAR Methods:

Different approaches have been devised to tackle sensor data in HAR systems. The conventional machine learning approaches used for this purpose include Decision Tree, SVM, HMM [20][21] are shown in Figure 2.2.The main issues with machine learning approaches is that they are heavily dependent on handcrafted features. To extract feature from the dataset is a tedious task and requires expert domain knowledge.

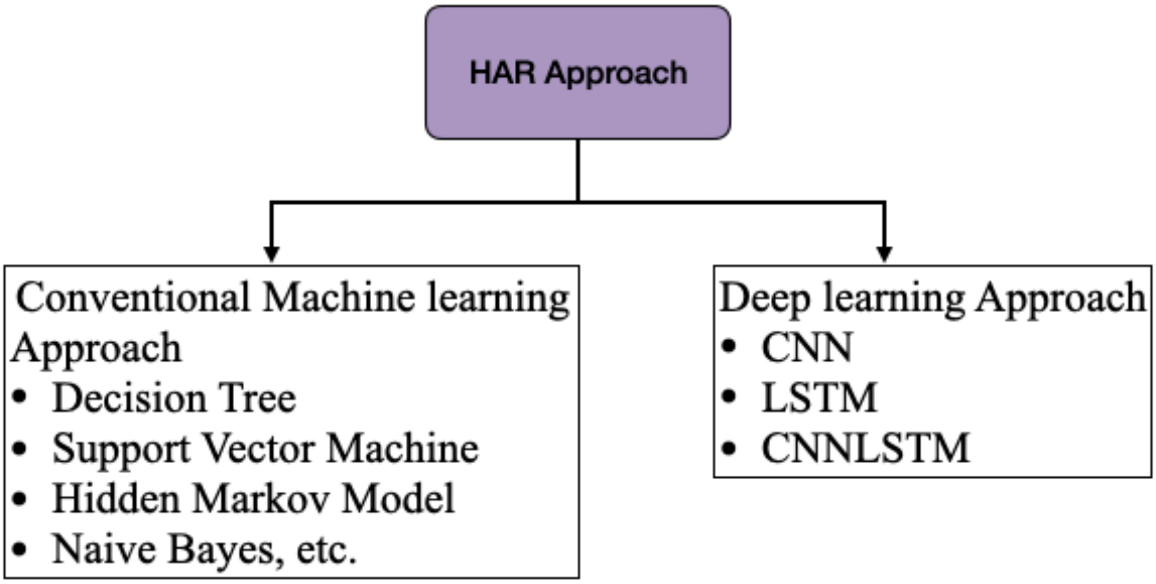

Figure 2.2: HAR approach

On the other hand, deep learning learns the high-level and meaningful features by training deep learning models.

M. Humayun Kabir *et al.* [20] propose a two-layer Hidden Markov Model for activity recognition. The activities are categorised depending on their locations, such as the restroom, kitchen, living room, and so on. Furthermore, the activities are categorised based on the object that the locations are placed in. Sensor data is segmented into fixed 60-second windows. The method is suitable only for binary sensors in a home environment.

Asghari *et al.* [21] describe a new online application of the Hierarchical Hidden Markov Model for detecting current activity on live sensor event streaming. The data stream is split in the first phase depending on the beginning and end of the activity patterns. The authors do not use a window technique to handle streaming sensor data as determining the appropriate window size adds another parameter. Instead of using a window, they offer a novel approach for determining the sequential pattern of sensor data at the start and end of each activity. Detecting the boundary of each activity is also a kind of segment to recognise the activity.

Bryan David Minor *et al.* [22] propose a novel technique for predicting future activity occurrence times from sensor data utilising multiple imitation learning algorithms and converting the task to a regression problem. The approach is based on an activity recognition algorithm devised by Krishnan and Cook [23] on streaming sensor data. The activity recognition algorithm extracts features from a sliding window containing the sensor values. The optimal window size used in the experiments is 30.

Francisco Javier Ordóñez *et al*. [9] introduce a deep neural network for activity recognition. The term deep neural network implies a combination of CNN and RNN architectures, which are named DeepConvLSTM. The aim of the convolutional layers is to extract the features of the sensor data. The recurrent layers are used to learn the temporal dynamics of the sensor signal. The method was evaluated on two different datasets, and based on the dataset, the window size is varied. The window size is 500 ms, and the step size is 250 ms, which implies that the step size is overlapping and is utilised for evaluation.

CNN is also widely used in human activity recognition systems [7],[13]. In [13], CNN is applied to the mobile sensor dataset to capture scale-invariance and local dependency. Furthermore, they propose a weight sharing method called partial weight sharing. In [7], CNN is used to extract local features and calculate certain statistical features to retain information about global features. The window size interval is between 1 second and 2.56 seconds and the segment size is 50 to 128 for the UCI dataset. The WISDM dataset window is 20 to 200 or between 1 second and 10 seconds with a step size is 20.

A supervised classification algorithm such as a support vector machine is adopted for human activity recognition in [26]. The advantage of using SVM is that it requires a small set of data to train and test compared to CNN or other neural networks. SVM kernel is used for the non-linear problem and the authors test two types of kernel and compare them. For multi-class classification, three methods exist: M-SVM, one-versus-one, and one-versus-all. In this paper, the one-versus-one method is used because it minimises the indeterminate zone compared to one-versus-all. The feature extraction is done manually and the features are chosen for conducting experiments on two subjects and identifying the essential features.

One more hybrid deep learning technique in human activity recognition is by Jing Zhao *et al*. [25]. This work acknowledges certain specific activities as well as the transition between two different activities. The transition between two tasks has a short length and occurs infrequently. The hybrid deep learning approach defines the collection of LSTM layers placed on top of the 1-dimensional CNN layers. The proposed model is evaluated on a dataset to determine all parameters that impact accuracy and the optimum parameters for the model.

In [28] the authors discuss an unsupervised machine learning approach to feature learning from sensor data. The approach includes PCA, auto-encoder, de-noising auto-encoder. Instead of using a window technique, the authors adopt a channel-wise way that learns the features of each channel and concatenates crucial features from the channels. There are two-channels in this accelerometer and gyroscope.

Chapter 3

Monitoring the Well-Being of Occupants

We propose a hybrid neural network model for activity recognition from sensor data in an assisted living environment. An assisted living environment is equipped with a plethora of various types of sensors. The sensors are wired and wireless, and the possibility of errors in wireless sensors is usually high. To overcome this, we perform data preprocessing steps to remove corrupt values and/or noisy values, missing values, and outliers from the dataset. Subsequently, the sensor dataset is partitioned into overlapping window sizes. The significance of partitioning is that the model can learn the dependency on the sequences we provide in the window, and can use it for subsequent sequences. Finally, the window is given to the hybrid model as input.

A myriad of machine learning algorithms such as Support Vector Machine, Decision Tree, Neural Network, and other probabilistic graphical methods have been used for activity recognition. In this work, we propose a hybrid model for activity recognition that incorporates two machine learning algorithms: the first stage machine learning algorithm is a 1D-CNN and it extracts features from the sensor's input data; subsequently, the output of the 1D-CNN is fed into an LSTM model which does the needful to recognise the activity.

The proposed human activity recognition system consists of the following steps:

- Sensor preprocessing and partitioning the data into windows or segments: Before performing any operation, the raw sensor data is preprocessed and normalised. The sensor dataset is split into windows each with a specified number of sequences.
- The preprocessed data is fed into the 1D-CNN model to learn the relevant features required to recognise various activities.
- The output of the 1D-CNN acts as the input of the LSTM network. The LSTM layers are updated after each time step. The input of the LSTM layer at time t are the elements of all the feature maps at the 1D-CNN layer at time t, with t=1......T.

The architecture of the hybrid model is shown in Figure 3.1.

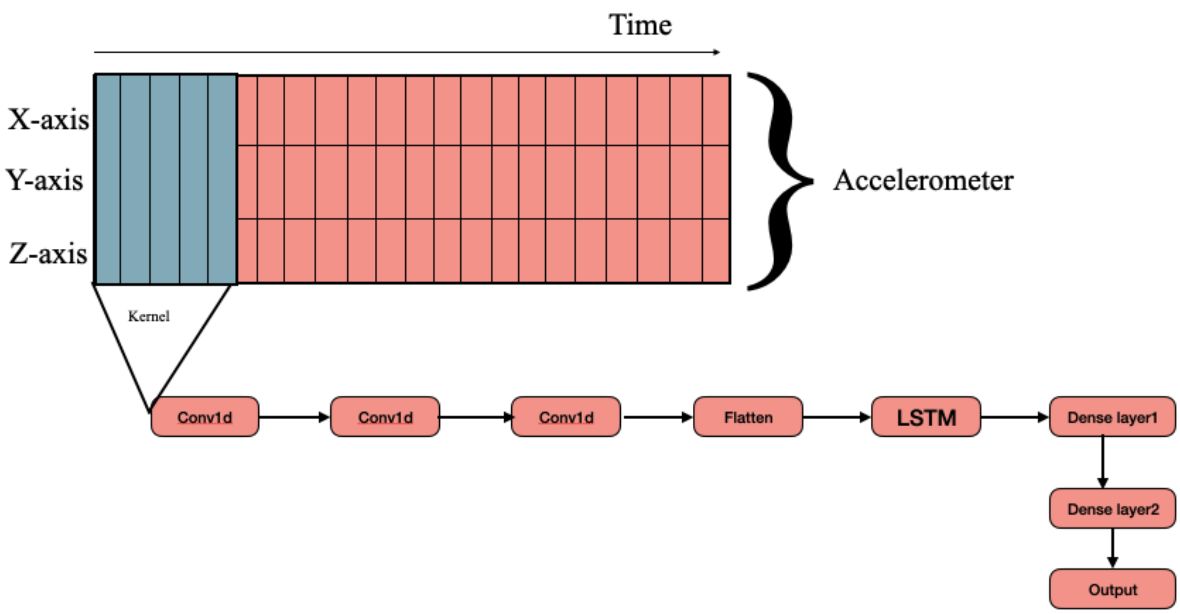

Figure 3.1: Architecture of Hybrid Model

A detailed description of the 1D-CNN and LSTM neural network are included in the following sections.

3.1 1D-CNN

The conventional CNN is intended to operate on 2D data such as images and videos. This is why they are referred to as "2D-CNNs". A similar feature learning method is used for one-dimensional data sequences, such as acceleration and gyroscopic data for human activity recognition. The model learns to extract features from observation sequences and to map internal features to distinct activity categories. The benefit of using CNNs for sequence classification is that they can learn from the raw time series data directly and do not require domain expertise to extract features manually. The model can learn the pattern associated with the time-series dataset of the sensors. It outperforms models fitted on a version of the dataset with extracted features. In dealing with one dimensional signals, 1D-CNN has certain advantages over 2D-CNN for the following reasons:

- While training a 1D-CNN, basic array operations will be performed instead of matrix operations in a 2D-CNN. The computing complexity, therefore, is considerably reduced.
- Training a 2D-CNN model requires special hardware such as GPU. On the other hand, 1D-CNN can be trained using a standard CPU implementation.
- Owing to their low computation requirements, 1D-CNNs are well suited for real-time and low-cost applications. Lightweight devices, in general, imply devices whose computation requirements are limited such as sensor nodes, mobile phones, and other handhelds devices.

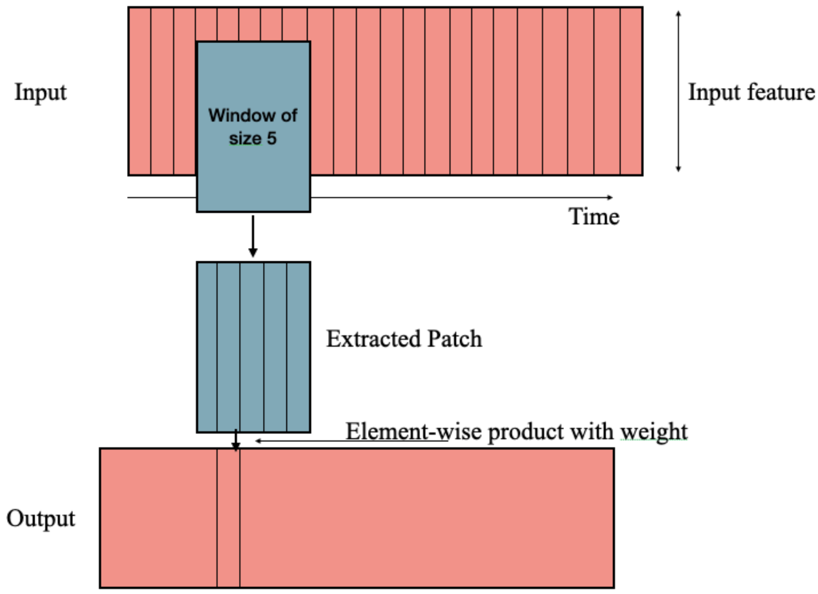

Figure 3.2: How 1D CNN works

Figure 3.2 shows the convolution operation in 1D-CNN and explains how it is differs from the conventional CNN convolution operation. An important difference is that we can afford to use larger convolution windows with 1D-CNNs. In a 2D convolution layer, a 3 × 3 convolution window contains 3 × 3 = 9 feature vectors; whereas with a 1D convolution layer, a convolution window of size 3 contains only 3 feature vectors. We can thus easily afford 1D convolution windows of size 7 or 9. If we compare with 2D-CNN, it needs significantly less computing power and less memory to store the parameters.

3.2 Long Short-Term Memory (LSTM)

Long short-term memory recurrent neural network is one of the most popular machine learning methods employed for handling sequential data. It keeps track of long-term dependencies and maintains information about the order. Following the CNN and pooling, the learned features are flattened to a single long vector and sent to the LSTM layer. We have added two fully connected layers or dense layers of LSTMs connected after the superficial LSTM layer. The fully connected layer ideally acts as a buffer between the learned features and the output, allowing the learned features to be interpreted before recognition of the activity.

The LSTM cell structure is shown in Figure 3.3. There are four gates that include: the cell state gate, forget gate, input data, and the output gate. These can collaborate with each other to preserve previous information and further improve the ability of learning useful information from accelerometer and gyroscopic time-series data. There are 128 neurons in both the Dense Layer1 and the Dense Layer2.

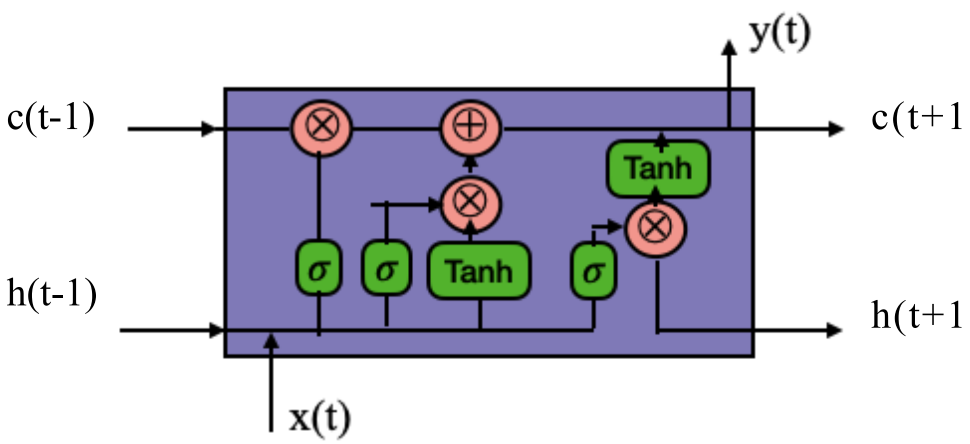

Figure 3.3: LSTM cell structure and computational flow

3.3 Data Preparation For Hybrid Model

The preliminary step in activity recognition is to prepare the dataset for training the model. The dataset that we validate our model with is the OPPORTUNITY dataset [1]. The dataset needs to be preprocessed before being fed into the model. Preprocessing involves eliminating non-numerical (NaN) values and missing values, and to replace these values with the column's mean value. The 1D-CNN works directly with the preprocessed dataset. There are a total of 116 features or variables at each time step. Further, each data series is partitioned into overlapping windows of two-seconds of data or 60-time steps. The 1D-CNN requires a three-dimensional input, that can be described as:

input-shape = (sample, steps, features)

Where 'sample' represents the total number of windows, 'steps' represents each time step in the window, and 'features' refers to the total number of axes from all sensors used to collect data.

In Equation 1, *Xij* represents the sensors values of different types of embedded sensors; *Yi* denotes the output that the model predicts. Four types of activities were recognised, and one more category belonging to null activity.

Equation 1 is the standard structure of the dataset of the sensors.

$$dataset = \begin{bmatrix} X11 & X12 & X13 & \dots & X1n & Y1 \\ X21 & X22 & X23 & \dots & X2n & Y2 \\ \vdots & & & & & \\ Xm1 & Xm2 & Xm3 & \dots & Xmn & Ym \end{bmatrix} \quad \textbf{(1)}$$

We need to prepare the dimensions of this matrix dataset as per the model requirements. The dataset in Equation 1 is split into three dimensions based on the step size. The new input shape is as shown earlier: *input-shape = (sample, steps, features)*. The splitting up of the data into three dimensional datasets increases the size of the data set in term of bytes. For example, the size of subject 1 before splitting is 149,676,640 bytes and it increases significantly to 9,492,329,472 bytes (excluding the output attribute).

The input matrix shown in Equation 2 is a 3D matrix and later reshapes into a four-dimensional matrix because of splitting the dataset into sequences, and the further splitting of the sequences into subsequences. For example, the 60-time steps in each window can be split into two subsequences of 30-time steps each.

$$Input = \begin{bmatrix} X11 & X12 & X13 & \dots & X1n \\ X21 & X22 & X23 & \dots & X2n \\ \vdots & & & & \\ Xm1 & Xm2 & Xm3 & \dots & Xmn \end{bmatrix} \quad \textbf{(2)}$$

The output is shown in Equation 3. There are a total of five activities in the dataset, including the null activity. We applied the 'one-hot-encoding' for dealing with categorical data in the output attribute. The given matrix is a binary matrix, and only one value in each row has a value of 1, the remaining values are 0.

$$output = \begin{bmatrix} y11 & y12 & y13 & y14 & y15 \\ y21 & y22 & y23 & y24 & y25 \\ \vdots & & & & \\ ym1 & ym2 & ym3 & ym4 & ym5 \end{bmatrix} \quad \textbf{(3)}$$

These are the standard operations that we performed on the dataset before it was fed to the model. Some of the operations such as splitting the data into windows and further split up into subsequences were based on the size of the window.

3.4 Hybrid Model Training

In the hybrid model being proposed, we combine the 1D-CNN and LSTM nets. First, the sensor data is pre-processed and normalised. Next, the 1D-CNN is designed to extract features from the sequential data. Finally, the LSTM layer further extracts some temporal features and predicts the activity. 1D-CNN can extract the effective features of time-series sequence data by performing 1D convolution operations using filters or kernels.

The hybrid model is trained in a supervised manner, back-propagating the gradients from the soft-max layer to the convolution layers of the 1D-CNN. The parameters are optimised based on the cross-entropy loss function using the Adam optimiser. Initially, the weights of the networks are assigned random values. At the beginning, the output is far from the actual values and the loss score is high. With every example that the network processes, however, the weights are adjusted in the correct direction and the loss score begins to decrease.

The gradients with respect to each weight and bias were computed backwards from the last layer to the first layer using the back-propagation method after the loss was known. The optimisation procedure was used to update the new weights and bias in the opposite direction of the gradients, with the goal of minimising loss.

The summary of the proposed hybrid model is depicted in Figure 3.4. The figure shows the total number of trainable parameters for the hybrid model. It contains information related to layers such as number of layers, number of neurons on each layer, and other relevant information.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
time_distributed (TimeDistri (None, None, 28, 64)      8896
_________________________________________________________________
time_distributed_1 (TimeDist (None, None, 9, 64)       0
_________________________________________________________________
time_distributed_2 (TimeDist (None, None, 7, 64)       12352
_________________________________________________________________
time_distributed_3 (TimeDist (None, None, 7, 64)       0
_________________________________________________________________
time_distributed_4 (TimeDist (None, None, 5, 64)       12352
_________________________________________________________________
time_distributed_5 (TimeDist (None, None, 1, 64)       0
_________________________________________________________________
time_distributed_6 (TimeDist (None, None, 64)          0
_________________________________________________________________
lstm (LSTM)                  (None, 128)               98816
_________________________________________________________________
dense (Dense)                (None, 128)               16512
_________________________________________________________________
dense_1 (Dense)              (None, 128)               16512
_________________________________________________________________
dense_2 (Dense)              (None, 5)                 645
=================================================================
Total params: 166,085
Trainable params: 166,085
Non-trainable params: 0
```

Figure 3.4: Hybrid model summary

Chapter 4

Experimental Results and Analysis

We evaluate the thee efficacy of the hybrid model in recognising activity using the Opportunity Dataset [1]. We divide each user dataset into 'training' and 'testing' parts and train and test the model respectively for each user with the corresponding data. In each case, we investigate the accuracy further, we plot the respective confusion matrix.

4.1 Dataset

The Opportunity dataset is used to assess our model's performance. The Opportunity dataset [1] is made up of a collection of complex naturalistic activities gathered in a sensor-rich environment. Overall, it includes recording four subjects in a daily living scenario, conducting morning activities with sensors of various modalities embedded into the surroundings, the item, and on the body, as shown in Figure 4.1. During the recordings, each subject participated in five Activities of Daily Living (ADL) and one drill session. During each ADL session, subjects performed the activity without any constraints. The ADL was on working in the kitchen an included: checking the utensils and ingredients in the kitchen, preparing coffee and drinking, preparing sandwiches, eating sandwiches, cleaning the kitchen, and others. In the drill session, the subjects performed 20 repetitions of a predefined set of 17 activities such as opening the dishwasher, closing the dishwasher, opening the fridge, closing the fridge, and others.

The our primary focus on the locomotion activity such Walk, sit, stand, and lying. The dataset comprises of 242 attributes and more than 5 lakhs data samples. There are various type of sensor used 145 attributes of body worn sensor, 60 attributes of the sensor which are embed to objects, 37 attributes of ambient sensors which are mounted on switches and drawers, kitchen appliances and doors and reed switches are placed on dishwasher and drawers.

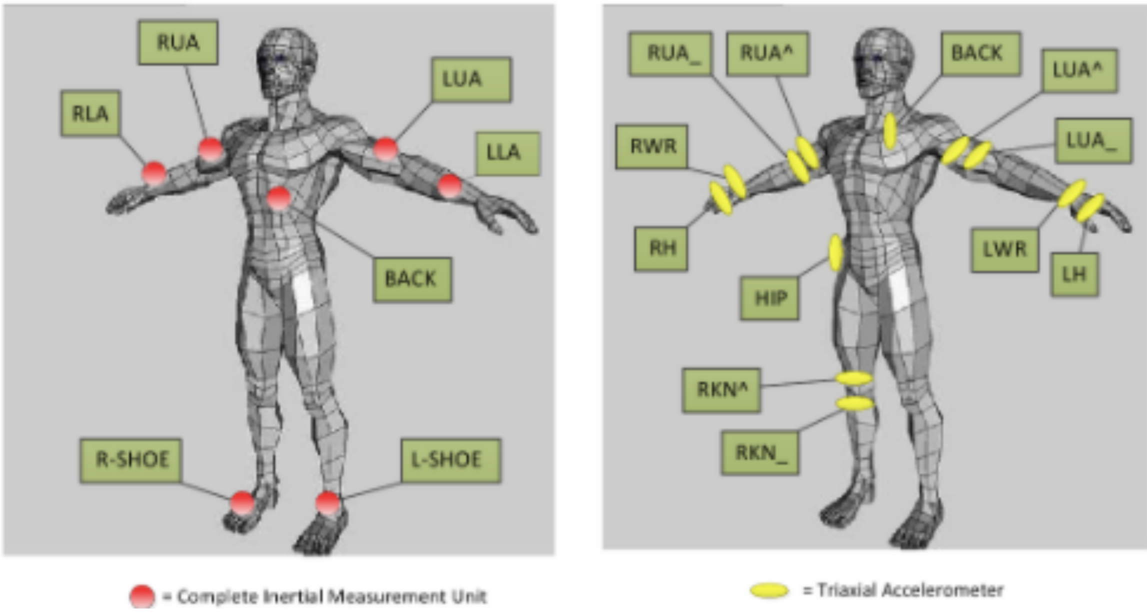

Figure 4.1: Placement of on body sensor in Opportunity dataset[1]
(left: Inertial measurement unit, right: 3-axial accelerometer [1]

Instead of using all the attribute we used those attribute which highly involved in the locomotion activities. We used body worn sensors dataset to accomplish our task and total attributes are 144. After the pre-processing step we further reduces the dimensions of the dataset by investigating the values of the particular sensor. The dataset we used consist of real valued attribute except the output attribute which is unique id given to each activity.

4.2 Results

To train and test our model for this thesis, we use the identical subset used in the Opportunity challenge dataset. We only consider the locomotion activity, and the dataset is divided into the 'training' and 'testing' set for each subject.

We measure the performance of our model by plotting the Confusion Matrix and the Receiver Operating Characteristics (ROC) curves. Confusion matrices contain information related to the actual label and the predicted label for locomotion classification. We identify the classification errors, as well as their quantities.

The ROC is a graph that depicts a classification model's performance across all classification levels. The ROC curve plots two parameters, i.e., the true-positive rate and the false-positive rate.

With the help confusion matrix, we can calculate the other performance metrics to see how well our model generalise.

1. Accuracy is one of the metric for evaluating classification model.

$$Accuracy = \frac{Number\ of\ correct\ prediction}{total\ number\ of\ predictions}$$

Accuracy alone doesn't tell the whole story when we are working with class imbalance classification. We have to look some other metrics for class imbalance classification problem.

2. Precision is tells the proportion of positive classification correctly.

$$Precision = \frac{True\ Positive}{False\ Positive\ +\ True\ Positive}$$

3. Recall calculates the actual positive classification correctly.

$$Recall = \frac{True\ Positive}{True\ Positive\ +\ FalseNegative}$$

4. F1-score is harmonic mean of precision and recall.

$$F1 - Score = \frac{2 *\ Precsion\ *\ Recall}{Precision\ +\ Recall}$$

All these metric are shown in Table 4.1 for all the subjects.

The performance of our model is calculated for each subject and a detailed discussion on the confusion matrix and ROC curves in included in the remainder of this chapter.

Subject 1:

The confusion matrix for subject 1 is shown in Figure 4.2. We observe that the null class is incorrectly classified more than 40% of the time. The model is unable to discriminate between the null-class and the standing type of locomotion, resulting in poor model performance. It is also worth noting that the static activity is generally accurately categorised. Static activities comprise locomotion types such as sitting, and lying down..

The ROC curves are mostly used for binary classification. In a multi-class classification scenario, the one-versus-all technique is utilised to make it work like a binary class classification. The same effect analysed by the confusion matrix is also analysed using the ROC curves. When compared with other classes, class 0 has the smallest area under the curve. In the ROC plot in Figure 4.3, class 0 is the same as the null class in the confusion matrix. This leads us to a similar conclusion as that drawn by the confusion matrix.

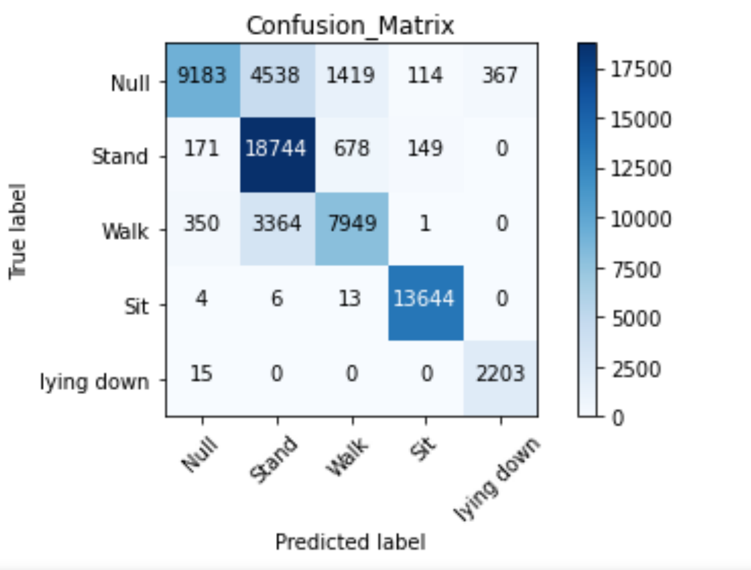

Figure 4.2 Confusion matrix of subject 1

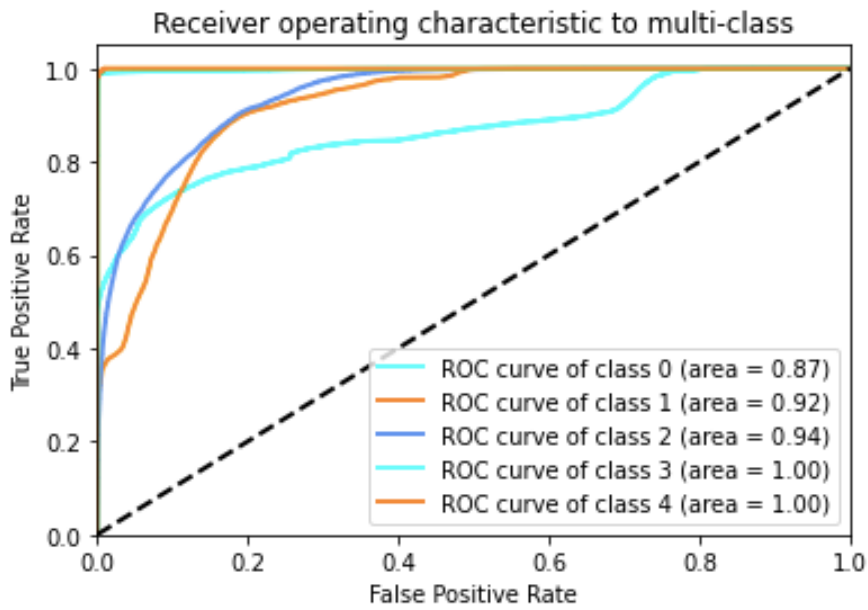

Figure 4.3 ROC curves of subject 1

Subject 2:

The confusion matrix of subject 2 is shown in Figure 4.4. The model behaves in a manner similar to that of subject 1. The null class is again wrongly classified more than 30% of the times. Almost 50% of the null values are wrongly classified as the 'walk' class. The walk class is the second most incorrectly classified class, with a rate of more than 32% of the incorrect labels being 'stand' instead of 'walk'. Lastly, the class sample 'stand' is predicted in almost all the classes and most of them are predicted correctly.

The same effect is observed in the ROC curve in Figure 4.5. Class 0 has the least area under the curve, which is the equivalent null class in the confusion matrix of Figure 4.6. Similarly, class 1 has the second least area under the curve and corresponds to the walk class. The ROC curve provides important insights on the variations in the true positive rate against the false-positive rate. Similar to the earlier observations, class 1 has the least but one area under the curve. Class 1 corresponds to the 'walk' class in subject 2's confusion matrix.

The activity recognition system seems unable to discriminate between the null class and other classes. The null class reflects an error in input or other issues while recording the activity. The null primarily affects the overall system performance and degrades its accuracy.

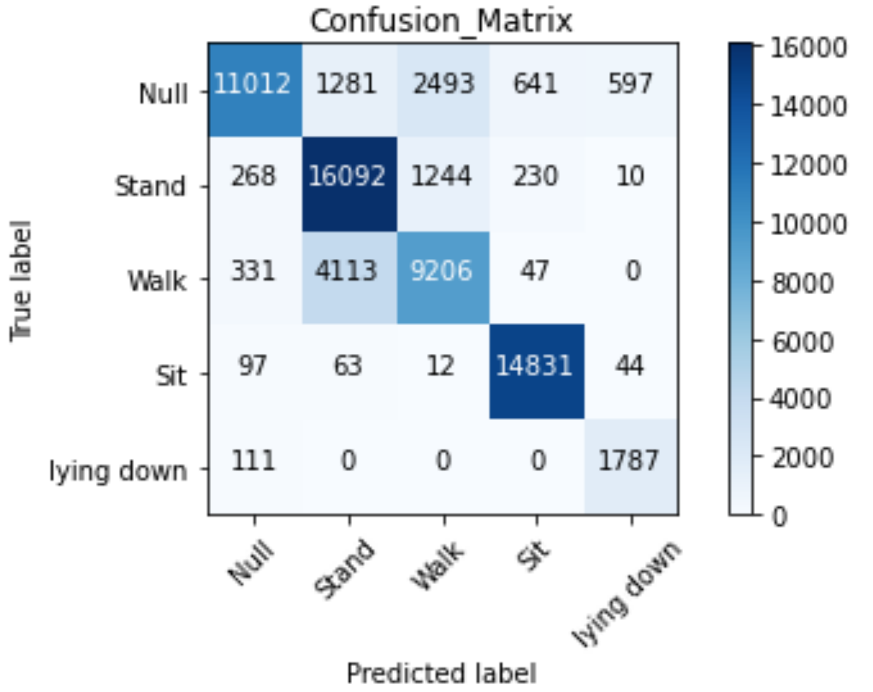

Figure 4.4 Confusion matrix of subject 2

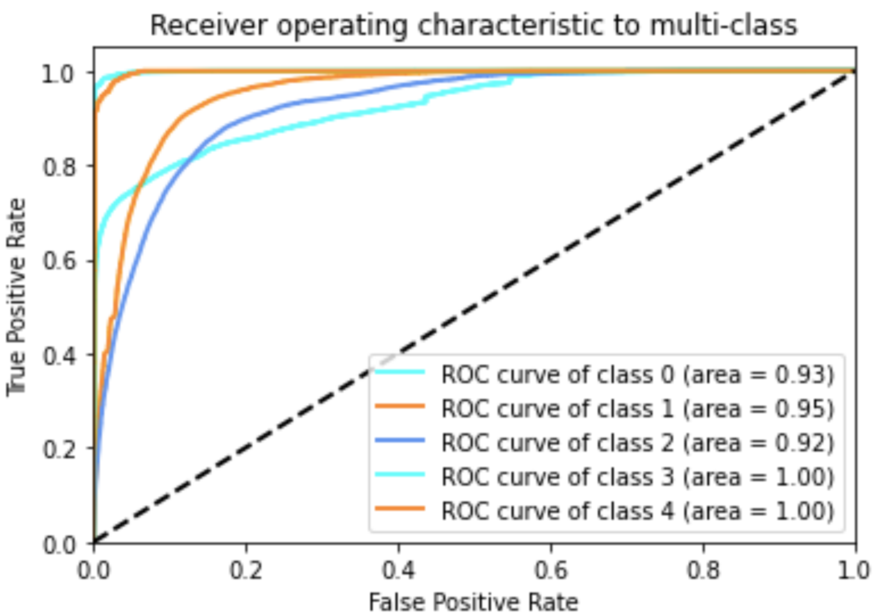

Figure 4.5 ROC curve of subject 2

Subject 3:

A confusion matrix is a standard tool for analysing classification model results. It quantifies the number of labels that are erroneously and adequately predicted. The confusion matrix of subject 3 is shown in Figure 4.6. The Human Activity Recognition (HAR) system suffers from significant errors while predicting the null class. If we analyse it further for subject 3, the null class is predicted wrongly more than 50% of the times. On most such occasions, the model predicts the null activity as a 'stand' activity. The 'walk' class is the second most undesirable class for the model for classification. We do observe, however, that the model can predict activities correctly that do not involve any mobility other than the null class.

Another way to analyse the model performance is the ROC curve shown in Figure 4.7. Class 0 has the least area under the curves and the null class in the confusion matrix. If we compare with other subjects' performance, the curve corresponding to subject 3 has the smallest area under the curve for class null.

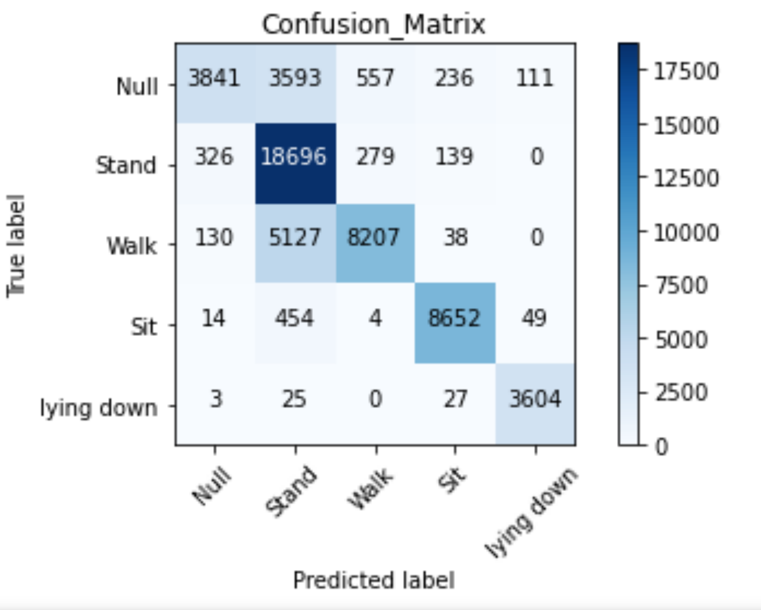

Figure 4.6 Confusion Matrix of subject 3

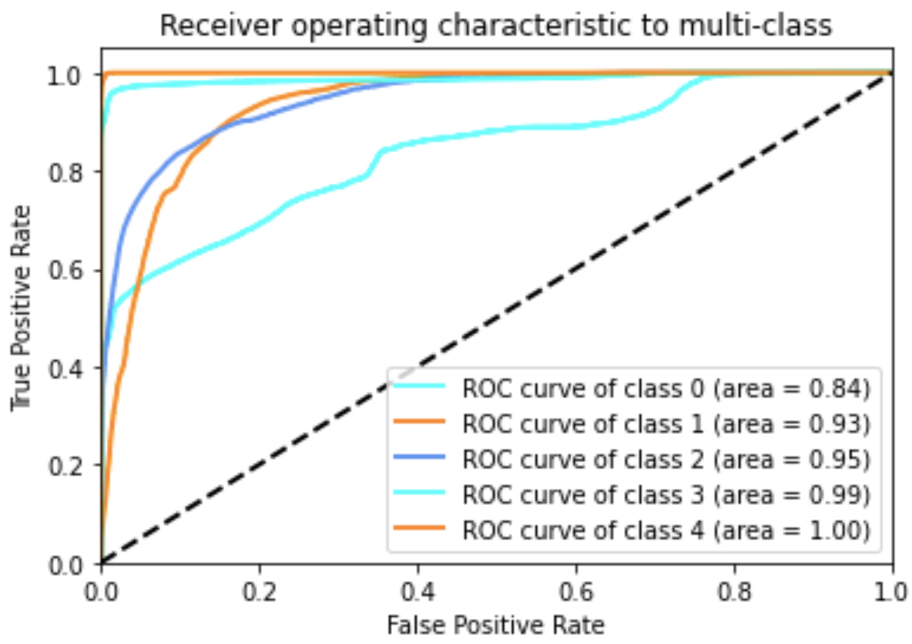

Figure 4.7 ROC curves of subject 3

Subject 4:

The confusion matrix of subject 4 is depicted in Figure 4.7. As previously discussed in the cases of subject 1 through subject 3, the model underperforms in predicting the null class. Nevertheless, the performance of the HAR system in the predictions for subject 4 are high as compared to the others. The incorrect prediction of the null class is only around 19% and of these, it predicts the null class as the 'walk' class 61.28% of the times. There remains, however, an issue with handling the 'walk' class by the model. It predicts 40% of the total walking instances wrongly. Furthermore, the model is unable to distinguish between the 'walk' class and the 'stand' class. The performance of the remaining classes, i.e., stand, sit, and lying down is good enough because there is no physical movement in these types of activity.

Just as in the case of the earlier subjects, there is one more way to analyse the classification model and that is using the ROC curve. The ROC curves are used to address the one-versus-all approach. The area under the curve for class 0 is 0.96, which is higher than that for other subjects. However, it underperforms when predicting class 2, which is the 'walk' class, and the area under the curve is 0.90, which is the lowest. The activity 'lying down' scores the highest in terms of correct predictions in all the subjects compared to other activities.

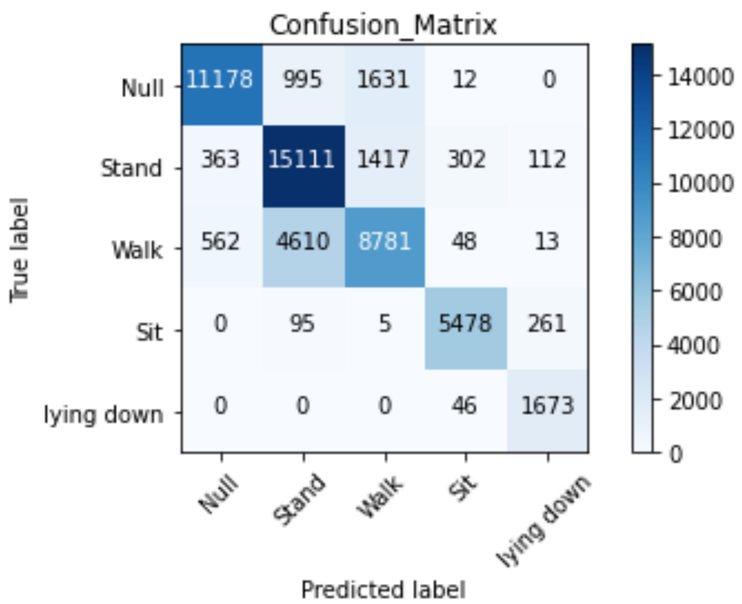

Figure 4.8 Confusion matrix of subject 4

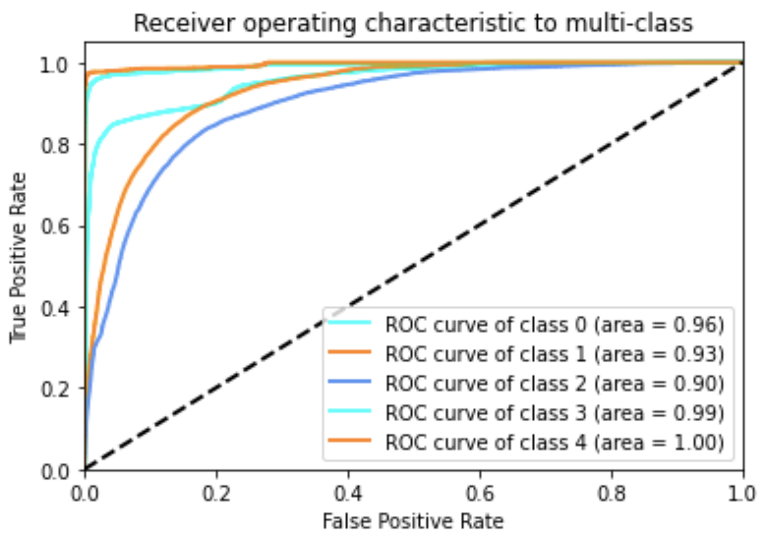

Figure 4.9 ROC curves of subject 4

There are some certain other measures in order to more deeply analyse the results provided by the model. Accuracy is quite straightforward to calculate. However, it has significant drawbacks because it does not consider anything related to the class imbalance dataset and tends to mislead the model. The other measures that help us gain better insight include: recall, precision, and the F1-score shown in Table 4.1. The formulations for these measures are discussed earlier in the chapter. In all these measurements, we calculate the True Positive, True Negative, and False Negative. The True Positive shows the number of accurately categorized predictions, whereas a certain class incorrectly rejects the false negative. True Negative implies the number of incorrectly identified predictions.

	Accuracy	**Precision**	**Recall**	**F1-Score**
Subject 1	0.822148	0.822148	0.822148	0.822148
Subject 2	0.820462	0.820462	0.820462	0.820462
Subject 3	0.794648	0.794648	0.794648	0.794648
Subject 4	0.801264	0.801264	0.801264	0.801264

Table 4.1: Other evaluation metrics for all the subjects

Chapter 5

Conclusions and Future Work

In this thesis, we propose an approach for Human Activity Recognition (HAR) that are especially useful in an Assisted Living Environment. The proposed method is a hybrid approach and combines the working of a 1D-CNN and an LSTM Neural Network. The approach potentially reduces the computational complexity of usual HAR approaches. As an HAR works with streaming sensor data, we use a sliding window technique to effectively handle the same and also to utilise the temporal features properly. The key advantages of the proposed method are: i) the feature extraction is carried out in a task-dependent and non-handcrafted manner; ii) when it comes to classification of human activities, the extracted characteristics have greater discriminative strength; iii) the model learns dependencies between sequences and keeps them in the same order as they are.

The performance of the proposed method is discussed in Chapter 4 of the thesis. The model is seen to perform well especially with activities like ‘lying’, ‘sit’, and ‘stand’. There is, however, considerable class imbalance with all subjects of the dataset. Of the remaining activities, the model's performance somewhat degrades because the model is unable to distinguish between the ‘walk’ and null classes. The null activity is rejected incorrectly and predicts either ‘walk’ or ‘stand’. ‘walk’ is the second most undesirable class from the model's perspective. To improve the model's performance, one approach could be removal of all instances of the null class because it does not belong to any activity. Overall, the critical point related to the model's performance is that it can correctly classify activities without any motion involved. We believe that the proposed method can compete as a feature learning and classification tool for the HAR problem.

In the future, we plan to further refine the model to identify more complex activities like preparing a sandwich, preparing coffee, having a drink, and so on. Such activities are termed complex because they are a combination of a large number of simple activities and a separate set and orientation and sensors would be required for each.

Bibliography

[1] Daniel Roggen, Alberto Calatroni, Mirco Rossi, Thomas Holleczek, Gerhard Tröster, Paul Lukowicz, Gerald Pirkl, David Bannach, Alois Ferscha, Jakob Doppler, Clemens Holzmann, Marc Kurz, Gerald Holl, Ricardo Chavarriaga, Hesam Sagha, Hamidreza Bayati, and José del R. Millán. "Collecting complex activity data sets in highly rich networked sensor environments" In Seventh International Conference on Networked Sensing Systems (INSS'10), Kassel, Germany, 6 2010.

[2] https://towardsdatascience.com/basics-of-the-classic-cnn-a3dce1225add

[3] Serkan Kiranyaz and Onur Avci and Osama Abdeljaber and Turker Ince and Moncef Gabbouj and Daniel J. Inman, 1D convolutional neural networks and applications: A survey, volume 151, pages 107398, year 2021, issn 0888-3270, doi :https://doi.org/10.1016/j.ymssp.2020.107398}

[4] Nan, Y.; Lovell, N.H.; Redmond, S.J.; Wang, K.; Delbaere, K.; van Schooten, K.S. Deep Learning for Activity Recognition in Older People Using a Pocket-Worn Smartphone. *Sensors* **2020**, *20*, 7195. https://doi.org/10.3390/s20247195

[5] Cook DJ, Krishnan NC, Rashidi P. Activity discovery and activity recognition: a new partnership. *IEEE Trans Cybern*. 2013;43(3):820-828. doi:10.1109/TSMCB.2012.2216873

[6] Andrey D. Ignatov and Vadim V. Strijov. Human activity recognition using quasi-periodic time series collected from a single tri-axial accelerometer. Multimedia Tools and Applications, 1:1–14, 2015.

[7] Andrey Ignatov, Real-time human activity recognition from accelerometer data using Convolutional Neural Networks, Applied Soft Computing, Volume 62, 2018, Pages 915-922, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2017.09.027.

[8] J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy. Deep convolutional neural networks on multichannel time series for human activity recognition. In Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15, pages 3995–4001, 2015.

[9] F. Ordonez and D. Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. Sensors, 16(1):115, 2016.

[10] Yousefi, M. R., Soheili, M. R., Breuel, T. M., & Stricker, D. (2015), A comparison of 1D and 2D LSTM architectures for the recognition of handwritten Arabic. Document Recognition and Retrieval XXII. doi:10.1117/12.2075930.

[11] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. Activity recognition using cellphone accelerometers. SIGKDD Explor. Newsl., 12(2):74–82, March 2011.

[12] Francois Chollet. 2017. <i>Deep Learning with Python</i> (1st. ed.). Manning Publications Co., USA.

[13] J. Lockhart. Mobile sensor for data mining. Fordham Undergraduate Research Journal, 1:67–68, 2011.

[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 Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on, pages 197–205, Nov 2014.

[15] B. Kolosnjaji and C. Eckert. Neural network-based user-independent physical activity recognition for mobile devices. In IDEAL 2015: 16th International Conference, pages 378–386, 2015.

[16] K. Kuspa and T. Pratkanis. Classification of mobile device accelerometer data for unique activity identificationunique activity identification. 2013.

[17] Xu, Gaowei and Ren, Tianhe and Chen, Yu and Che, Wenliang, A One-Dimensional CNN-LSTM Model for Epileptic Seizure Recognition Using EEG Signal Analysis, Frontiers in Neuroscience, 14, PAGES:1253, 2020, DOI=10.3389/fnins.2020.578126

[18] Coppola, C., Cosar, S., Faria, D.R. *et al.* Social Activity Recognition on Continuous RGB-D Video Sequences. *Int J of Soc Robotics* **12,** 201–215 (2020). https://doi.org/10.1007/s12369-019-00541-y

[19] Activity Prediction using a Space-Time CNN and Bayesian Framework
Hirokatsu Kataoka1 , Yoshimitsu Aoki 2 , Kenji Iwata 1 and Yutaka Satoh1
National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan
2 Keio University, Yokohama, Kanagawa, Japan

[20] Bruna Maria Vittoria Guerra, Stefano Ramat, Giorgio Beltrami and Micaela Schmid. Automatic Pose Recognition for Monitoring Dangerous Situations in Ambient-Assisted Living, Laboratory of Bioengineering, Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

[21] M. Humayun Kabir, M. Robiul Hoque, Keshav Thapa, and Sung-Hyun Yang. 2016. Two-layer hidden Markov model for human activity recognition in home environments. Int. J. Distrib. Sen. Netw. 2016, Article 15 (January 2016), 1 pages. DOI:https://doi.org/10.1155/2016/4560365

[22] Asghari, P., Soleimani, E. & Nazerfard, E. Online human activity recognition employing hierarchical hidden Markov models. *J Ambient Intell Human Comput* **11,** 1141–1152 (2020). https://doi.org/10.1007/s12652-019-01380-5

[23] B. D. Minor, J. R. Doppa and D. J. Cook, "Learning Activity Predictors from Sensor Data: Algorithms, Evaluation, and Applications," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 12, pp. 2744-2757, 1 Dec. 2017, doi: 10.1109/TKDE.2017.2750669

[24] N. Krishnan and D. J. Cook, “Activity recognition on streaming sensor data,” Pervasive Mobile Comput., vol. 10, pp. 138–154, 2014.

[25] Francisco Javier Ordóñez Morales and Daniel Roggen. 2016. Deep convolutional feature transfer across mobile activity recognition domains, sensor modalities and locations. In Proceedings of the 2016 ACM International Symposium on Wearable Computers. Association for Computing Machinery, New York, NY, USA, 92–99. DOI:https://doi.org/10.1145/2971763.2971764.

[26] Huaijun Wang, Jing Zhao, Junhuai Li, Ling Tian, Pengjia Tu, Ting Cao, Yang An, Kan Wang, Shancang Li, "Wearable Sensor-Based Human Activity Recognition Using Hybrid Deep Learning Techniques", Security and Communication Networks, vol. 2020, Article ID 2132138, 12 pages, 2020. https://doi.org/10.1155/2020/2132138

[27] Fleury A, Noury N, Vacher M. Supervised classification of Activities of Daily Living in Health Smart Homes using SVM. Annu Int Conf IEEE Eng Med Biol Soc. 2009;2009:6099-102. Doi: 10.1109/IEMBS.2009.5334931. PMID: 19965259.

[28] Shuvo, Md Maruf Hossain et al. “A Hybrid Approach for Human Activity Recognition with Support Vector Machine and 1D Convolutional Neural Network.” *2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)* (2020): 1-5

[29] Li Y., Shi D., Ding B., Liu D. (2014) Unsupervised Feature Learning for Human Activity Recognition Using Smartphone Sensors. In: Prasath R., O’Reilly P., Kathirvalavakumar T. (eds) Mining Intelligence and Knowledge Exploration. Lecture Notes in Computer Science, vol 8891. Springer, Cham. https://doi.org/10.1007/978-3-319-13817-6_11

[30] Ronao, Charissa Ann and S. Cho. “Human activity recognition with smartphone sensors using deep learning neural networks.” *Expert Syst. Appl.* 59 (2016): 235-244

[31] Rybalkin, V., Sudarshan, C., Weis, C. et al. Efficient Hardware Architectures for 1D- and MD-LSTM Networks. J Sign Process Syst 92, 1219–1245 (2020). https://doi.org/10.1007/s11265-020-01554-x

[32] Mohammad Reza Yousefi, Mohammad Reza Soheili, Thomas M. Breuel, Didier Stricker, "A comparison of 1D and 2D LSTM architectures for the recognition of handwritten Arabic," Proc. SPIE 9402, Document Recognition and Retrieval XXII, 94020H (8 February 2015); https://doi.org/10.1117/12.2075930

[33] Jianfeng Zhao, Xia Mao, Lijiang Chen, Speech emotion recognition using deep 1D & 2D CNN LSTM networks, Biomedical Signal Processing and Control, Volume 47, 2019, Pages 312-323, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2018.08.035.