**CHAPTER - 1**

**INTRODUCTION**

**1.1 Motivation and Goal:**

With the advancement of technology, machine learning and deep learning methods are being applied in numerous fields. Among these Human activity recognition (HAR) has been one of the most significant topics in the active research field. Human activity recognition intends to monitor, recognize the activity of a person based on a series of observations and surrounding environment. In easy words, HAR technology is used in recognizing human daily activities, from simple activities like standing, sitting, walking upstairs, running to complex activities like cooking while standing, watching TV while sitting or lying [1].

Human activity recognition, due to availability of devices and wearable sensors at low cost has become an integral part of people’s daily lives and is being applied broadly in common domains like health management - elderly monitoring, disease prevention, rehabilitation, in the form of smart cities – domestic activity monitoring. Furthermore, HAR is applied in security concerns like through individual activity monitoring solutions, crowd anomaly detection. Besides, wearable and inertial sensors combined with embedded systems are being used in sports activities [2,3].

In this project our goal is to develop a model to recognize daily and simple human activities in real time. For simplicity our goal is to recognize 6 activities and we used 2 types of sensor data for this (3-axis accelerometer and 3-axis gyroscope).

**1.2 Approaches to human activity recognition**

In order to recognize human activities, different methods may be applied. Methods and ways of approaching may vary from type of collecting data and data processing to detection approaches. Based on collecting activity data, HAR can be divided into two types:

1. Computer Vision based
2. Sensor based

**1.2.1 Computer Vision based approach**

In computer vision based approach, images and videos are collected by means of optical sensors like cameras, CCTVs and then captured images or videos are analyzed. Vision based data being affordable and collectible hardly with any trouble, vast majority of research have been conducted on vision based HAR [4]. Besides, contact based like wearable sensor based approach sometimes requires sophisticated equipment and also it has to be affordable in cost, correct size and also the user’s acceptability. Moreover, vision based systems will not require the user to wear devices uncomfortable to them in different parts of the body. In this type of approach mainly three stages are followed which includes – 1. detection (first stage), which determines the part of the body to recognize or follow (methods like using skin color, shape, pixel values etc. are used), 2. Tracking (second stage), where links between successive images are provided (methods like feature tracking, contour tracking, optimal estimation etc. are used) and 3. Classification, which is the final stage, and where different machine learning and deep learning algorithms are used to finally recognize the activity [5].

**1.2.2 Sensor based approach**

As mentioned, human activity recognition can be carried out in a sensor based approach. Sensors of different kinds are used in data acquisition for the recognition to be performed. Sensors can be integrated in a device or can be used separately. Based on the platforms used, sensors can be:

1. Wearable Sensors
2. Smartphone Sensors

**(a) Wearable Sensors:**

These types of sensors are used for only one particular purpose/function. These can’t take any other measurements other than the ones these are made for. Usually these sensors are integrated into a device used for the given task only. Wearable sensors can further be classified into more types – *Inertial sensors* include accelerometer, gyroscope, magnetometer, *Physical health sensors* includes heart rate sensors(HR), skin temperature, oxygen saturation (SPO2) etc., *Environmental sensors* include, temperature, barometer, humidity, light sensors etc. Sometimes these wearable sensors are not used in applications, for example in human activity recognition due to size, price and acceptability to carry by the user [6].

**(b) Smartphone Sensors:**

Since smartphones, as well as smartwatches, are easy to carry and use, they are used widely to collect data. Today smartphones and smartwatches are embedded with sensors like accelerometer, gyroscope, barometer, GPS, temperature etc. Smartphones and watches are now also used in other fields alongside human activity recognition, like, health monitoring, monitoring sports activities etc. [3,6].

**1.3 Challenges in sensor based human activity recognition:**

In conducting human activity recognition, over the years, a variety of challenges have been faced. Challenges vary from type of sensors to use datasets to be used to methods or models and many more.

**1.3.1 Appropriate Sensors and Selecting Sensor Types:**

First of all, the problem arises and much challenge is faced in choosing a suitable sensor or set or sensors as well as the type of sensors that will be used to collect raw data.

**(i) Challenges in Wearable Sensors:**

Sensor based human activity recognition eliminates the drawback of vision based systems, where data has to be collected by means of external sensing methods like cameras. When collecting data by means of camera or CCTV the user or the subject has to stay within a particular range, which limits the reliability. Furthermore, data taken by cameras are influenced by the background, daylight, weather etc. And thus, sensor based systems are preferable which eliminates all of these drawbacks [14,15].

Now the challenge arises in choosing the suitable type of sensors. Sensors can be a separate wearable sensor or sensors used in smartphones. Each has their own trade off. A wearable sensor is used for only a particular purpose whereas smartphones have multiple sensors embedded in it. The first challenge in wearable sensors is to select which sensor to use, that is whether to use only one sensor or multiple sensors at a time. If only one sensor is used the data processing part becomes trouble free. But the recognition accuracy and thus the performance might not be satisfactory. On the other hand, with the use of multiple sensors, the accuracy and performance of the system will be more acceptable and better. Moreover, a single sensor based system might be good for recognizing simple activities like walking, sitting, standing to sit etc., though not ensured to good accuracy, but will face much challenge in detecting complex activities and will not give high accuracy. Hence, a multi-sensor based system is preferable over a single sensor. Multiple sensors like accelerometer, gyroscope, magnetometer etc. are used to collect data and recognize complex activities [13,16].

Though multi-sensor based wearable sensors have shown high accuracy and good performance, challenges are also faced in this method. First, if to be mentioned, users may feel uncomfortable in wearing or carrying these sensors while data is being collected. And it is problematic to carry multiple sensors and data becomes difficult to process. Besides, in unfavorable environments hindrances are faced while calibrating and collecting data [3,17]. To overcome these challenges smartphones and smartwatches have been preferred in recent years. Especially with the advancement of technology, powerful smartphones and smartwatches have been introduced, which have high computation capacities along with powerful sensors. Smartphones are used in our daily lives and it is not only a communicating device but also a good sensor platform and sensors are calibrated easily as well as data can be transferred easily by means of Bluetooth, Wi-Fi etc. [12,18].

**(ii) Challenges in Smartphone sensors:**

However, challenges are faced in using smartphones, more precisely in collecting sensor data. Moreover, one another problem with smartphones is the rapid loss of battery energy. The main challenge that is assumed is the variation of location of smartphones. In quite a number of papers like [8,18-21] this problem has been discussed and as well as solutions have been proposed. Unlike wearable sensors, smartphones aren’t kept in a fixed position and it is not reasonable to do so either. Smartphones are often kept in shirt or pant pockets, bags and in different pockets in bags etc. places and also these are kept in either right hand or left hand along with various orientations. Signals read will vary in these conditions which makes it arduous to recognize the particular activity. If activity recognition is done using smartphone data while keeping it in only a particular position, orientation or place, low accuracy and wrong recognition will result. And, it is the greatest challenge to make the recognition system with smartphones a position/location independent system. And also data collected from smartphones may contain a lot of noise. Hence, research has been conducted and is still being conducted focusing on such problems with location of smartphones [8,12,18].

**1.3.2 Choosing suitable methods:**

**(i) Using Hand Crafted Features:**

Apart from the above mentioned challenges, one of the biggest challenges is the method to choose for recognition i.e. whether to use a hand-crafted featured model or a deep learning model. Whether a hand-crafted featured model or deep learning model is used, selecting and finding appropriate features for the recognition system is the most significant task in human activity recognition. Feature extraction is more challenging as different activities can have similar characteristics and hence it becomes strenuous to obtain unique features for each activity [7,11]. Hand-crafted featured models require features to be selected manually, knowledge equivalent to an expert is required. Moreover, hand-crafted features might not work for other similar applications and also will face many difficulties while recognizing complex activities [22,23]. Furthermore, hand-crafted featured human activity recognition systems and most importantly labelled data from sensors are vital for good performance. Otherwise, expected accuracy and performance will not be obtained. Additionally, in order to collect such data and hence prepare proper datasets sophisticated infrastructure might be required leading to more cost and time consumption. Also, many datasets have labelled data indeed but those labels might be assigned to data having information from secondary activities. And, training errors might occur in machine learning algorithms due to such data and degrade the performance, especially while a class or activity is being assigned to a data segment [9,24].

**(ii) Deep Learning:**

In order to tackle the challenges and drawbacks of hand-crafted features, deep learning models are used in recent years. In deep learning models, necessary features are automatically extracted from raw sensor data. Unlike popular machine learning algorithms such as KNN (K nearest neighbor), SVM (Support Vector Machine) etc. deep learning models don’t need carefully engineered hand crafted features as they can develop the most efficient features from raw data [10]. Deep learning models contain multiple layers and the path deepens and hence the name deep learning [2]. Deep learning methods are being used in human activity recognition(HAR) to automatically extract useful features from raw sensor data using multiple layers of abstraction. And it is applied in HAR concerning the movements of human beings hierarchically [24]. Numerous deep learning models using deep learning networks like CNN (Convolutional Neural Network), LSTM (Long Short Term Memory), Bi-LSTM (Bidirectional Long Short Term Memory), RNN (Recurrent Neural Network) etc. have been introduced in large number of papers. For example, in CHIH-TA YEN *et al.* a deep learning model using CNN has been proposed [2], in Xiaokang Zhou *et al.* LSTM based model [13] and in [24] RNN based model has been proposed to overcome previous challenges and obtain better performances.

**1.3.3 Challenges in Deep learning method:**

However, implementation of deep learning models doesn’t mean that the system will not have a lack of performance and efficiency. Challenges are faced even when using deep learning models. As we know deep learning models have numerous layers, these require a large number of initializations and parameter tuning. All of these escalates computational costs, time and requires powerful processing units for fast computations. Moreover, due to such large computations, low processing devices and low energy mobile devices aren’t suitable [10]. Hence, pre-processing and dimensionality reduction are significant in the recognition system. Dimensionality reduction, can lessen the computational complexity mentioned above. Now, one of the greatest challenges is the pre-processing of the data to achieve optimum performance. For this, various processing techniques such as normalization, standardization etc. have been used and need further experimentation to find suitable accuracies and performances which is the challenge. Moreover, challenges remain in issues with hyper-parameters like learning rate optimization, kernel filter size, reduction of data size etc. [26]. In order to improve the performance data augmentation method is used to generate more training examples from existing small dataset and in order to reduce and sometimes prevent overfitting. For data augmentation techniques like permutation of location with sensor events, arbitrary rotation etc. are used. This method is still a challenge for implementing motion sensors’ (accelerometer, gyroscope) data for improving performance [27].

**1.3.4 Appropriate Datasets:**

Apart from all the challenges that have been mentioned above, challenges that cannot be denied, remain in using proper datasets. Custom datasets can be prepared collecting data using wearable sensors, smartwatches or smartphones. But preparing custom datasets is challenging because the raw data needs to be pre-processed properly for the deep learning layers to extract useful features and perform recognition. Since sometimes many challenges are faced in preparing custom datasets, publicly available datasets can be found and used for human activity recognition. In KAIXUAN CHEN *et al.* information and also challenges about publicly available datasets have been given. For example, the most common three datasets used are UCI HAR, WISDM Activity Prediction, OPPORTUNITY and their challenges that for the first and third one have been mentioned to multimodality of the data whereas for the second dataset the challenge is said to class imbalance [7].

**1.4 Research objectives:**

**1.5 Project Overview:**

In Chapter 1, a brief introduction of Human Activity Recognition has been provided along with concepts and motivation of this project. A brief description of various methods in this field has also been given. Furthermore, different challenges that are faced while performing research in this field have been discussed. Lastly, objectives of our research are reviewed in short.

In Chapter 2, we have discussed other works related to this field. Various methods in sensor based HAR of previous works along with their accuracies have been overviewed. Moreover, we have also made a comparison between methods of previous works along with their accuracy along with problems in respective works.

In Chapter 3, Theoretical Overview.

In Chapter 4, Our Methodology

In Chapter 5, Result Analysis

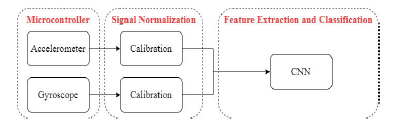
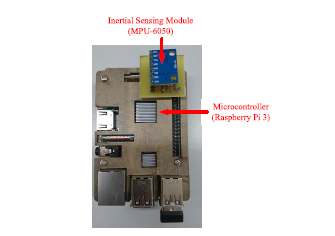
In Chapter 6, conclusion has been drawn and scope of improvement in the future in this research has been discussed.

**CHAPTER - 2**

**RELATED WORKS**

Over the years numerous researches have been conducted in sensor based Human Activity Recognition. In the researches new methods, models, and even devices have been proposed in order to improve the performance, flexibility across varying positions, environments and devices.

In Yen *et al.* [2], a micro-controller based device, positioned in the waist, is used to recognize six basic human activities (walking, sitting, standing, lying, going upstairs and going downstairs). The wearable device consisted of an inertial sensor including a microcontroller, a three-axis accelerometer, and a three-axis gyroscope. In this research they have used a recognition algorithm consisting of signal acquisition, signal normalization, and a feature learning method where the feature learning method was based on a 1D Convolutional Network that can extract feature and preform classification automatically from the raw data. They used both public dataset (UCI-HAR) and also collected their own raw data. The method and model used in the paper resulted in accuracies of 98.93% and 97.19% for UCI-HAR and recorded data respectively on training samples and 95.99% and 93.77% on test samples respectively.



a

b

Figure 2.1: (a) Structure of the hardware used in the paper; (b) The proposed algorithm

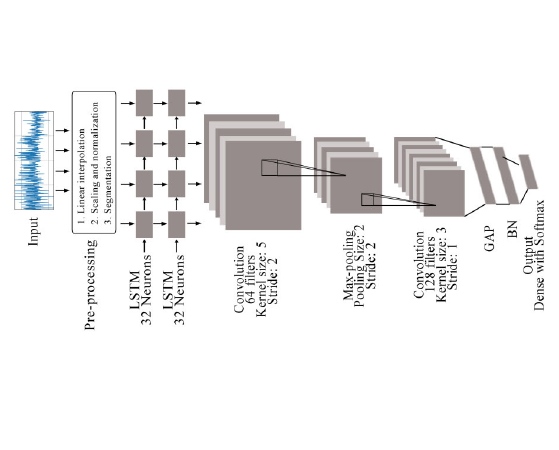
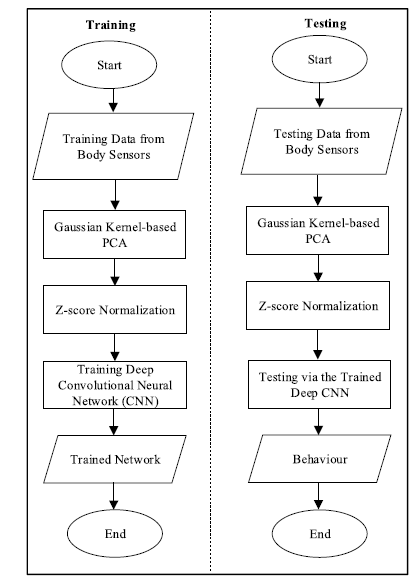
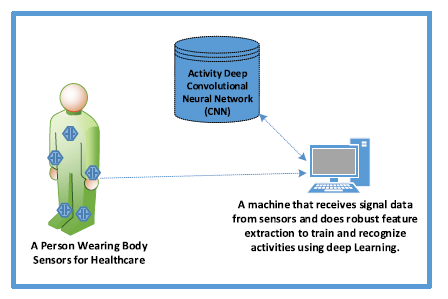
In K. Xia *et al.* [25], a deep learning model consisting of LSTM and CNN is proposed where raw data collected from mobile sensors were fed. The proposed architecture in the paper also contained GAP (Global Average Pooling) layer in order to reduce model parameters and a batch normalization layer after it to speed up the convergence and training process. The model was evaluated on three public datasets (UCI-HAR, WISDM, and OPPORTUNITY) where overall accuracy for UCI-HAR, WISDM and OPPORTUNITY were 95.78%, 95.85% and 92.63% respectively.

Figure 2.2: Frame diagram of the proposed model

One of the most popular fields in which HAR is applied is health care. With the advancement of body sensors, this technology is being applied in health care. Moreover, due to the ongoing popularity and advancement in Internet of Things (IoT), sensor based HAR has become a hot topic in smart healthcare which can escalate rehabilitation for elderly and weak people [13].

In Md. Zia Uddin [15], an approach using body sensor and CNN model (deep Convolutional Neural Network) was proposed for smart healthcare. In the research signals from body sensors like ECG, magnetometer, accelerometer and gyroscope were used and analyzed to extract suitable features using methods like Gaussian kernel based PCA (Principal Component Analysis) and Z-score normalization. Based on the extracted features the proposed CNN model was trained and the entire approach was applied not only on raw data collected but also on mHealth public dataset (which contained recordings of 10 subjects for 12 activities) and compared with other typical approaches. Their approach resulted an average accuracy of 93.90% whereas approaches that they compared with – ordinary ANN (Artificial Neural Network) with an average accuracy of 87.99%, DBN (Deep Belief Network) with an average accuracy of 90.01%.



a

b

Figure - 2.3: (a) Schematic setup for the recognition system; (b) Flow chart for the proposed approach

In Xiaokang Zhou *et al.* [13], a semi supervised deep learning framework for HAR in IoT environment was introduced. The deep learning model was developed in order to improve the efficiency of HAR performance and deal with feebly labeled sensor data. In order to apply auto labelling, a method or scheme which is a Deep Q Network (DQN) was designed and it was developed based on a distance - based reward rule and along with it a LSTM based architecture was implemented in order to train and perform the recognition. Evaluation was done on finding accuracies on different position of sensors on the body and results were compared with other approaches like typical DNN (Deep learning network), typical machine learning approaches like Random Forest (RF) and Support Vector Machine (SVM). The proposed method gave higher accuracies on different body positions than the other methods. Experimental results obtained from fusing different features from different sensors were evaluated based on performance metrics like Recall, Precision and F1-score, where the proposed method gave the best performance.

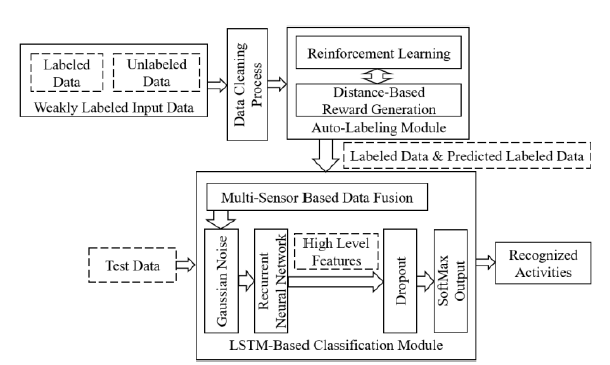


Figure - 2.4: The proposed semi – supervised deep learning framework.

Indeed, there has been much advancement in wearable sensors for which the overall performances of HAR have increased with the introduction of many frameworks, methods or models after conducting research, but the problem with wearable sensors, as mentioned many times, is that its requirement of additional sensing components along with user discomfort.

Hence, with the technological advancement of smartphones, smartwatches and thus sensors in these devices, smartphones are used in the field of HAR widely nowadays. Myriad of researches has been and is also being conducted using smartphone data due to their availability and being widely used by everyone in their daily lives.

In C.A. Ronao [23], a robust deep learning framework with deep convolutional neural network (convnet) has been proposed in order to improve the performances of HAR using smartphone data. The convnet automatically extract useful features and the temporal – dependency of local time series data is eliminated by the convolution layer while the small input translations are eliminated by pooling layer used in the model. Dataset was prepared collecting raw data with smartphones, from 30 subjects performing six daily activities with the smartphone in their pocket. In this research they used 7352 samples (21 subjects) for training and 2947 samples (9 subjects) for testing purposes. The model after training and validation results were compared with other state-of-the-art methods. The proposed model combined with a multilayer perceptron classifier (MLP) gave an accuracy of 94.79% on test samples and using temporal Fast Fourier transform (tFFT) on the data the proposed convnet gave a slightly increased accuracy by approximately 1% which was 95.75%.

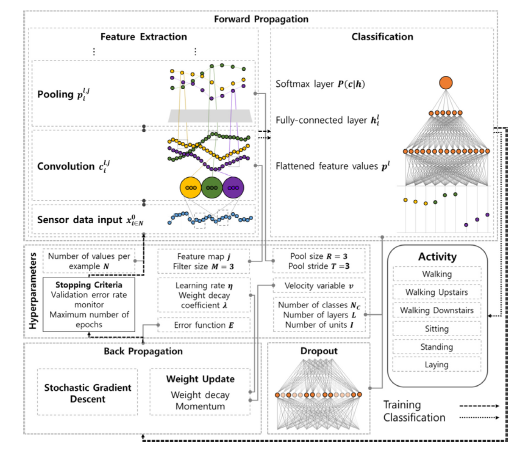


Figure - 2.5: The overview of the model proposed in the paper [23]

In the above mentioned paper (C.A. Ronao [23]), smartphone data were used for HAR. But smartphone has some issues to be dealt with, which includes the variation in orientation, positions/ locations and these will vary the signals for the same activity which will degrade the overall performance and the reliability of HAR.

Zhenghua Chen *et al.* [18], has worked on the problem that occurs due to orientation issue in smartphones. A HAR system was introduced in the paper based on coordinate transformation and principal component analysis (CT-PCA) and online support vector machine (OSVM). PCA eliminates the problem due to orientation and improves the accuracy of the system and the model showed effective performance on different placements and subjects separately. Again, to overcome the degradation of performance over different placements in overall, an online independent SVM algorithm was used. Results were obtained for many orientations and compared with other state of arts methods, where the proposed method outperformed all of those in every orientation and subjects. For instance, for the orientations in Figure 2.6 (a), (b), and (c) accuracies achieved by the proposed method were 96.22%, 94.89%, 93.56% respectively and a total average of 94.89%.



a

b

c

Figure - 2.6: One set of orientation experimented in Zhenghua Chen *et al*. [18]

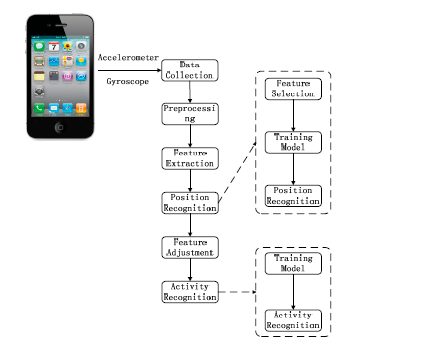
In Rong Yang [28], another method for dealing with smartphone location was introduced which is known as PACP (Parameters Adjustment Corresponding to smartphone Position). In this method, according to the paper, features were extracted from raw data of accelerometer and gyroscope and used to recognize the position of the smartphone first. Then the sensor data were adjusted and thus necessary features were extracted and train the model to recognize the activities. In the research, they collected data by building an iOS app and data were collected from 10 volunteers performing 5 activities (walking, standing, running, going upstairs, and going downstairs) keeping the phone in 4 different positions (bag, trouser pocket, coat pocket, and hand). Results obtained by the proposed PACP method were compared with another method described in Mario *et al.* [30]. The proposed PACP method gave better accuracy of 91.27% where in Mario *et al.* the accuracy was 87.24%.

Figure - 2.7: An Overview of the proposed PACP method in Rong Yang [28]

In Masud Ahmed *et al.* [29], a Position and Orientation Independent Deep Ensemble Network (POIDEN) was introduced to improve the performance of HAR system irrespective of smartphone orientation and position/location. The method was developed and tested for complex activities like sitting on a bike or train or subway. The proposed architecture uses a LSTM network to select an optimum classifier for the activity recognition. For the purpose of classifier selection i.e. for the LSTM network they used intermediate feature set (IFS), while for the classification purposes statistical classifier feature set (SCFS) were used. Moreover, in order to deal with the orientation characteristics, the rotational characteristics of accelerometer and magnetometer were transformed from local co-ordinate to earth co-ordinate and additionally jerk features, position independent features were used along with parameter adjustment. In the paper, SHL (Sussex – Huawei Locomotion) dataset was used where there were 59 days’ equivalent training data, 3 days of validation data and 20 days of test data, varying smartphone positions from 4 places (torso, hip, hand, and bag). The result achieved from the architecture was evaluated against the dataset. Since, the activities were very complex the overall F1 score came out to be 70.57% and accuracy to be 67.5%.

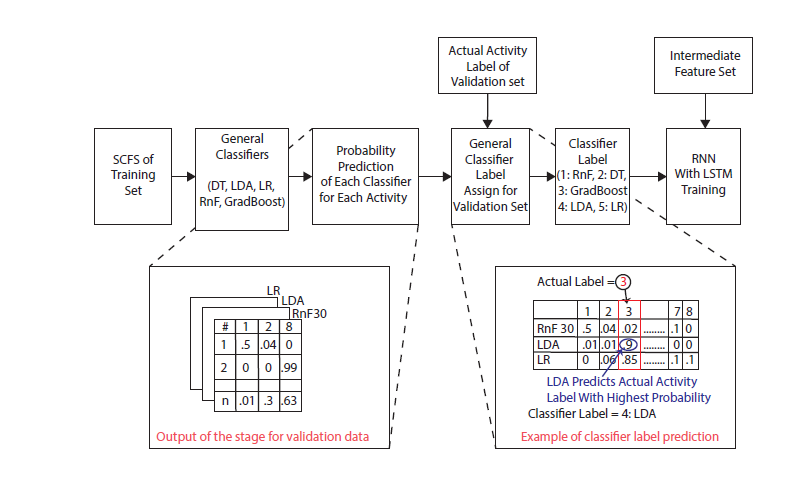


Figure - 2.8: The proposed POIDEN architecture.

**CHAPTER 3**

**THEORETICAL OVERVIEW**

The goal of the project is to be able to classify human activities from smartphone sensors. All smartphones nowadays have built in IMU (Inertial Measurement Unit). With the smartphone placed in the right position, some basic activities can be detected using the data from the IMU sensors. In this project, data from accelerometer and gyroscope were used to predict human activities like walking, sitting, standing, climbing etc. Both the accelerometer and gyroscope sensors provide data for 3 axes (X, Y, Z) with respect to the smartphone orientation.

**Sensor Description**

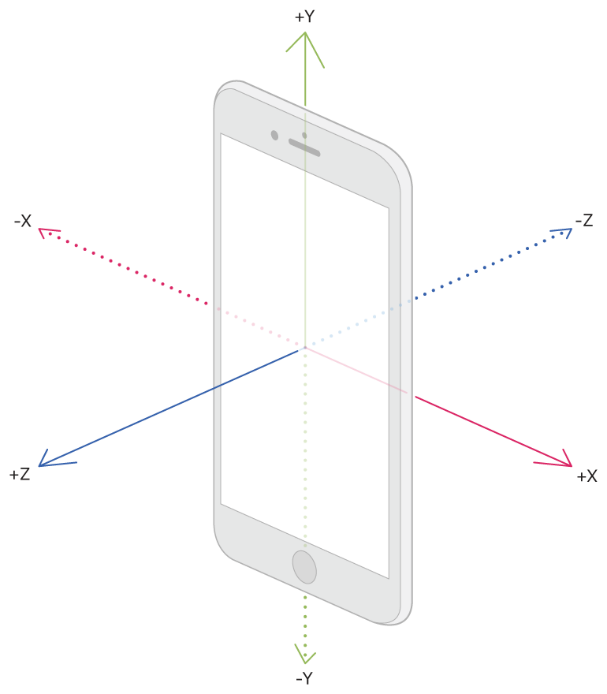
**Accelerometer:** The accelerometer measures acceleration of the phone with respect to its reference frame. The measurement is made for 3 axes namely X, Y, Z which are defined by the orientation of the smartphone IMU. The general reference frame is X-axis towards the right side of the front face, Y-axis towards up on the front side and Z-axis is outward from the front face. The measurement is given in unit ms-2. The sampling frequency and precision depends on the smartphone with the sampling frequency varying from 20-200 Hz [1]. For convenience, the data used in this project was resampled to 50 Hz.

Figure – 3.1: Smartphone Coordinate axes

**Gyroscope**: This sensor returns the rotational velocity of the device around the 3 coordinate axes. The reference frame of gyroscope sensor is the same as accelerometer. The unit of data in this sensor is degree/second (°/sec). The sampling frequency of this sensor also depends on the smartphone used, varying from 20-200 Hz and data from this sensor was also resampled to 50 Hz for use in this project [3].

**Dataset Description**

For this project, datasets containing activities of daily life (ADL) were used. The datasets used are namely:

* UCI HAPT DATASET
* WISDM
* Dataset - RealWorld (HAR)

**WISDM**:

In this dataset, data were collected from 36 users at a sampling frequency of 20Hz. Activities performed by every user in this experiment are jogging, walking, walking upstairs, walking downstairs, sitting and standing. Data were collected from a smartphone placed at user’s waist. Only accelerometer sensor data was collected in this dataset. The dataset is quite unbalanced and had to be made balanced by various augmentation techniques [31].

**UCI HAPT Dataset**:

The experiments were carried out with a group of 30 volunteers within an age bracket of 19-48 years. They performed a protocol of activities composed of six basic activities: three static postures (standing, sitting, lying) and three dynamic activities (walking, walking downstairs and walking upstairs). The experiment also included postural transitions that occurred between the static postures. These are: stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, stand-to-lie, and lie-to-stand. All the participants were wearing a smartphone (Samsung Galaxy S II) on the waist during the experiment execution. We captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz using the embedded accelerometer and gyroscope of the device. The experiments were video-recorded to label the data manually.

The inertial sensor data part was used in this project. All types of transition data were re-labeled to a single class called transition. So, the total number of activity classes were 7: walking, walking downstairs, walking upstairs, standing, sitting, lying and transition. This dataset contains both accelerometer and gyroscope data [32].

**Dataset-RealWorld (HAR):**

The data set covers acceleration, GPS, gyroscope, light, magnetic field, and sound level data of the activities climbing stairs down and up, jumping, lying, standing, sitting, running/jogging, and walking of fifteen subjects (age 31.9±12.4, height 173.1±6.9, weight 74.1±13.8, eight males and seven females). For each activity, the acceleration of the body positions chest, forearm, head, shin, thigh, upper arm, and waist were recorded simultaneously. Each subject performed each activity roughly 10 minutes except for jumping due to the physical exertion (~1.7 minutes). Concerning male and female, the amount of data is equally distributed. Each movement was recorded by a video camera to facilitate the usage [33].

**Classification Models**

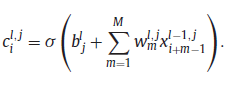
Human Activity Recognition models can be of two types, first in which features are predefined before training and secondly deep learning models where the models define features while training. The deep learning approach was implemented in this project. The models used in this project are mainly combination of two types of models:

* Convolutional Neural Network (CNN)
* Long Short-Term Memory (LSTM)

**Convolutional Neural Networks:**

Convolutional Neural Networks (CNN) are great at finding salient features in a signal. The lower the convolution layer in the model, the more it obtains local salience of signals that may characterize each signal class. The high-level layers form a high-level representation of the signals. [34]

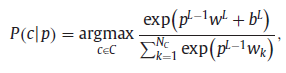
Let, xi0 = [x, x2, x3,…,xN] be the accelerometer and gyroscope sensor data input vector where N is the window length. Then the output of the lth layer is given by the equation



Pooling layers are commonly used with CNN layers. These layers can retrieve a summary of the output of the previous layer and represent it with smaller vector thus reducing computation complexities for next layers. Pooling layers can be of various kinds like average pooling or max-pooling. Considering Cil,j to be the output of the previous layer of the max-pooling layer, the output of the layer will be given by the equation



where R is the pooling size and T being the pooling stride. Adding softmax layer at the end of a stack of CNN and pooling layers can be used to classify human activity signals. The output of the softmax layer being given by equation,



where c is the activity class, L is the last layer index, and NC is the total number of activity classes. [23]

In [35] CNN model is used to detect transportation mode from smartphone sensor data. Fig shows the proposed model structure in [35].

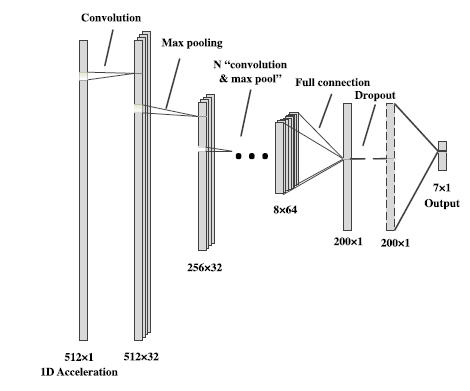
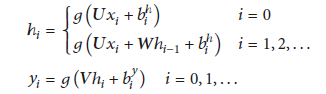


Figure – 3.2: CNN model use in [35]

As mentioned above, CNNs are good at identifying salience features from signals but lose temporal features in the process. But temporal information may be vital in separating certain activity classes. Recurrent Neural Networks (RNN) can obtain temporal features from sequential time series data like sensor data from smartphone. General RNN models usually suffer from vanishing and exploding gradient problem. This problem can be avoided by combining Long Short-Term Memory (LSTM) with RNN [36].

An RNN layer consists of 3 layers namely, input layer, hidden layer and output layer. Let, the input set be x = [x0, x1, x2, x3, …, xt, xt+1, ……], hidden set be h = [h0, h1, h2, h3, …, ht, ht+1, ……], output layer be y = [y0, y1, y2, y3, …, yt, yt+1, ……] and 𝑈, 𝑊, 𝑉 denote weight metrics from the input layer to the hidden layer, from the hidden layer to the hidden layer, and from the hidden layer to the output layer, respectively. Then the outputs of the hidden layer and output layer can be defined as,



where hi is the output of the hidden layer, yi is the output of the output layer and g(.) is the activation function [37].

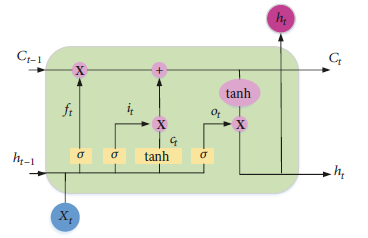
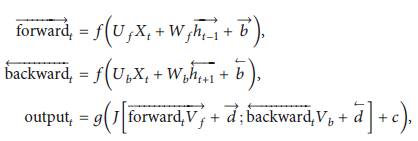


Figure – 3.3: Functional unit of an RNN

The transmission of state information in an RNN or one directional LSTM layer is one way, from front to back. Bidirectional LSTM is used to solve this problem and increase availability of information. To facilitate transmission of information in both ways two layers are implemented, forward and backward layer, which have no interaction between them until final output layer. Output for bidirectional LSTM can be defined by the following equation [38].



LSTM models are used to classify human activities in [36], [37] and [39].

**CNN and LSTM Hybrid models:**

In detection of human activities from sensor signals both salient and temporal relations are required to be identified. Although CNN models are great at obtaining salient features from signals, temporal features are ignored. Again, LSTM models are great at identifying temporal features from signals. So, using a hybrid model of CNN and LSTM layers may be better at classifying activity signals.

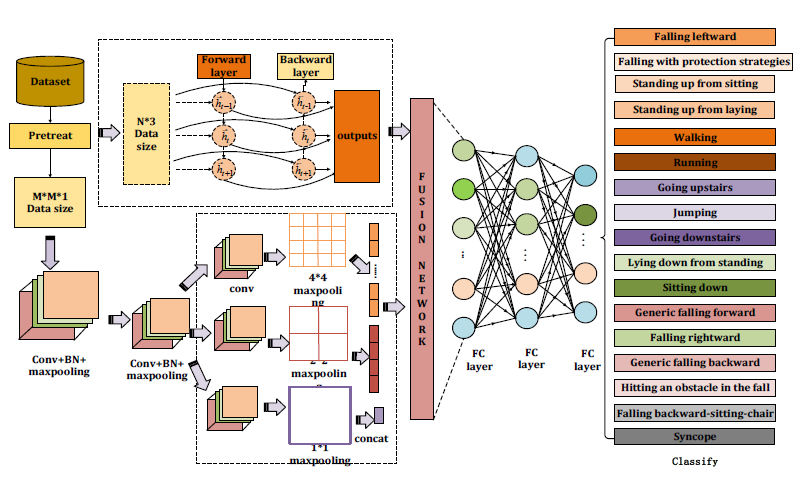


Figure – 3.4: Fusion Network model of CNN and LSTM in [40]

Different hybrid models using CNN and LSTM layers can be created. In [40] features detected from CNN model and LSTM model is combined in a single layer and classification is done based on that layer.

In [41] and [42] CNN layers were applied on input signal first followed by LSTM layer to define feature layer that is finally used for classification.

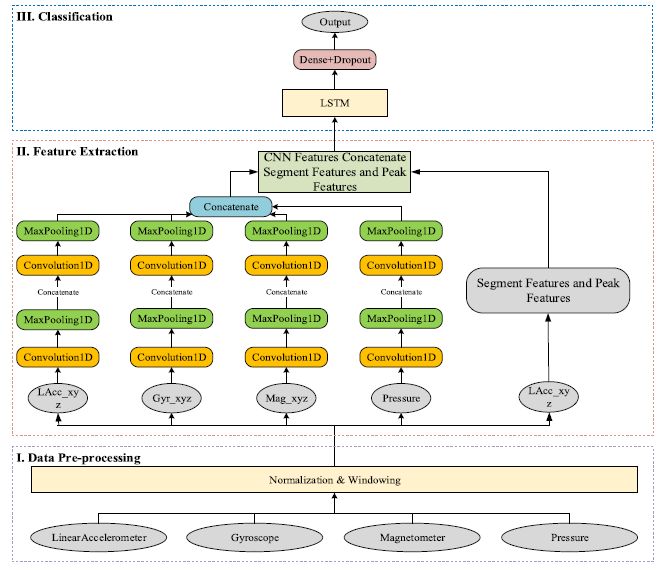


Figure – 3.5: A Hybrid model of CNN and LSTM for activity recognition.

In [25] and [43] LSTM layers were after the input layer which creates an abstract representation of the signal on which CNN layers were applied to finally form features used for classification.

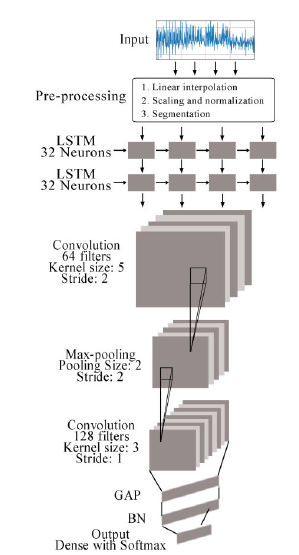


Figure – 3.6: A model where the input was fed into an LSTM layer.

**Data Augmentation:**

Deep learning techniques have achieved benchmark results and proved to be excellent in classification in many different fields. But in order to achieve such results deep learning models require lots of input training data which may not be available for many fields.

In case of human activity datasets based on sensor data, data scarcity is a problem for some datasets. Again, for some datasets class imbalance is a problem as some activities have more data as the activities are easy to perform [44].

Data Augmentation can solve these problems and improve the performance of Deep Learning models by enabling them to generalize better [27]. Demonstrates some data augmentation techniques for time-series sensor data for Parkinson’s disease detection using Convolutional Neural Networks.

Some methods used to augment data are namely permutation, time warping, scaling, magnitude warping, cropping and rotation.

Permutation is used to randomly change the temporal location of different incidents within the data window. Time warping slightly distorts time steps between data points and changes temporal location of events. Jittering introduces random noise in the data. Rotation is used to rotate signal around an axis to avoid bias towards a certain orientation [27].

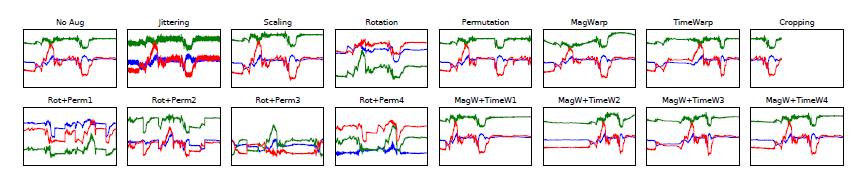


Figure – 3.7: Various data augmentation methods and corresponding outputs obtained in [27]

**Global Reference Frame:**

Sensor data measured by the Inertial Measurement Unit (IMU) of a smartphone is with respect to the reference frame of the IMU or the smartphone. As a result, changing the orientation of the smartphone will lead to different sensor data. But in most sensor dataset, data was collected for a specific orientation hence a specific reference frame of IMU. A model trained on such data will not perform on sensor data from other orientations and will have a bias to the specific orientation on which it was trained. This problem can be solved by converting the data from a specific reference frame to a global reference frame.

In [28] acceleration data from accelerometer was separated into acceleration towards the gravity and acceleration towards the horizontal direction. Then features were calculated from the separated data.

Henpraserttae *et.al* [45] presents a transformation matrix to convert acceleration value from local reference frame of IMU to global reference frame. Here, at first the downward direction is found from averaging values of dynamic part of the acceleration data. Let w be the mean of the dynamic portion. The forward axis can be calculated from the projection of data onto the plane normal to w. The projected data on the plane can be found by subtracting the acceleration data along the vertical axis along w from the original data. It can be calculated using the equation



where x' is the removed acceleration signals along vertical axis and x is the original acceleration data. Next Eigen-decomposition was performed on the covariance matrix of the projected data using the following equation,



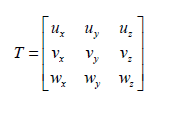
where μ' is the mean of the projected data, calculated as follows



Lastly, the sideward axis can be found by considering the cross product between the vertical and the forward axes



Then, the transformation matrix T is



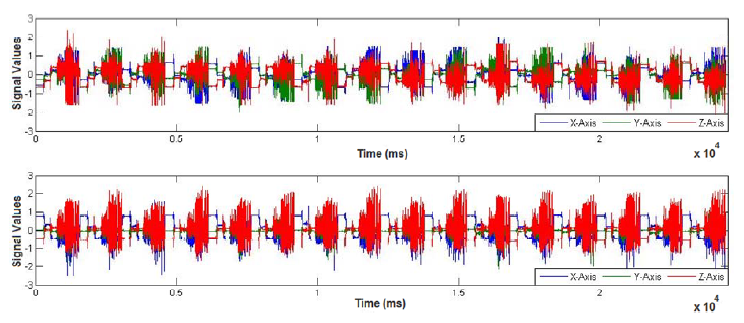


Figure – 3.8: Transformation output from local reference frame to global reference frame

**BIBILOGRAPHY**

1. Razzaq, Muhammad A.; Cleland, Ian; Nugent, Chris; Lee, Sungyoung. 2020. "SemImput: Bridging Semantic Imputation with Deep Learning for Complex Human Activity Recognition" Sensors 20, no. 10:2771. <https://doi.org/10.3390/s20102771>
2. C. -T. Yen, J. -X. Liao and Y. -K. Huang, "Human Daily Activity Recognition Performed Using Wearable Inertial Sensors Combined With Deep Learning Algorithms," in IEEE Access, vol. 8, pp. 174105-174114, 2020, doi: 10.1109/ACCESS.2020.3025938.
3. Sousa Lima, Wesllen; Souto, Eduardo; El-Khatib, Khalil; Jalali, Roozbeh; Gama, Joao. 2019. "Human Activity Recognition Using Inertial Sensors in a Smartphone: An Overview" Sensors 19, no. 14: 3213. <https://doi.org/10.3390/s19143213>
4. L. Minh Dang, Kyungbok Min, Hanxiang Wang, Md. Jalil Piran, Cheol Hee Lee, Hyeonjoon Moon,Sensor-based and vision-based human activity recognition: A comprehensive survey,Pattern Recognition,Volume 108,2020,107561,ISSN 0031-3203. <https://doi.org/10.1016/j.patcog.2020.107561>.
5. Beddiar, D.R., Nini, B., Sabokrou, M. *et al.* Vision-based human activity recognition: a survey. *Multimed Tools Appl* **79,**30509–30555 (2020). <https://doi.org/10.1007/s11042-020-09004-3>
6. H. Yu, S. Cang and Y. Wang, "A review of sensor selection, sensor devices and sensor deployment for wearable sensor-based human activity recognition systems," 2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA), 2016, pp. 250-257, doi: 10.1109/SKIMA.2016.7916228.
7. Kaixuan Chen, Dalin Zhang, Lina Yao, Bin Guo, Zhiwen Yu, and Yunhao Liu. 2021. Deep Learning for Sensor-based Human Activity Recognition: Overview, Challenges, and Opportunities. ACM Comput. Surv. 54, 4, Article 77 (May 2021), 40 pages. DOI: <https://doi.org/10.1145/3447744>.
8. A. Das Antar, M. Ahmed and M. A. R. Ahad, "Challenges in Sensor-based Human Activity Recognition and a Comparative Analysis of Benchmark Datasets: A Review," 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), 2019, pp. 134-139, doi: 10.1109/ICIEV.2019.8858508.
9. Song-Mi Lee, Sang Min Yoon and Heeryon Cho, "Human activity recognition from accelerometer data using Convolutional Neural Network," 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), 2017, pp. 131-134, doi: 10.1109/BIGCOMP.2017.7881728.
10. Henry Friday Nweke, Ying Wah Teh, Mohammed Ali Al-garadi, 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,Volume 105,2018,Pages 233-261,ISSN 0957-4174. <https://doi.org/10.1016/j.eswa.2018.03.056>.
11. Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. ACM Comput. Surv. 46, 3, Article 33 (January 2014), 33 pages. DOI: <https://doi.org/10.1145/2499621>.
12. Wan, S., Qi, L., Xu, X. *et al.* Deep Learning Models for Real-time Human Activity Recognition with Smartphones. *Mobile Netw Appl* **25,**743–755 (2020). <https://doi.org/10.1007/s11036-019-01445-x>.
13. X. Zhou, W. Liang, K. I. Wang, H. Wang, L. T. Yang and Q. Jin, "Deep-Learning-Enhanced Human Activity Recognition for Internet of Healthcare Things," in IEEE Internet of Things Journal, vol. 7, no. 7, pp. 6429-6438, July 2020, doi: 10.1109/JIOT.2020.2985082.
14. Mukhopadhyay, S.C. & Ghayvat, Hemant & Liu, Jie & Gui, Xiang. (2015). Wellness Sensors Networks: A Proposal and Implementation for Smart Home to Assisted Living. IEEE Sensors Journal. 15. 1-1. 10.1109/JSEN.2015.2475626.
15. M. Z. Uddin and M. M. Hassan, "Activity Recognition for Cognitive Assistance Using Body Sensors Data and Deep Convolutional Neural Network," in IEEE Sensors Journal, vol. 19, no. 19, pp. 8413-8419, 1 Oct.1, 2019, doi: 10.1109/JSEN.2018.2871203.
16. J. Qi, P. Yang, M. Hanneghan, S. Tang and B. Zhou, "A Hybrid Hierarchical Framework for Gym Physical Activity Recognition and Measurement Using Wearable Sensors," in IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1384-1393, April 2019, doi: 10.1109/JIOT.2018.2846359.
17. L. Gou, D. Peng, X. Chen, L. Wu and Q. Tang, "A Self-Calibration Method for Angular Displacement Sensor Working in Harsh Environments," in IEEE Sensors Journal, vol. 19, no. 8, pp. 3033-3040, 15 April15, 2019, doi: 10.1109/JSEN.2018.2879099.
18. Chen, Zhenghua & Zhu, Qingchang & Yeng, Chai & Zhang, Le. (2017). Robust Human Activity Recognition Using Smartphone Sensors via CT-PCA and Online SVM. IEEE Transactions on Industrial Informatics. PP. 1-1. 10.1109/TII.2017.2712746.
19. J. Yang, H. Zou, H. Jiang and L. Xie, "Fine-grained adaptive location-independent activity recognition using commodity WiFi," 2018 IEEE Wireless Communications and Networking Conference (WCNC), 2018, pp. 1-6, doi: 10.1109/WCNC.2018.8377133.
20. Hur, Taeho; Bang, Jaehun; Kim, Dohyeong; Banos, Oresti; Lee, Sungyoung. 2017. "Smartphone Location-Independent Physical Activity Recognition Based on Transportation Natural Vibration Analysis" Sensors 17, no. 4: 931. <https://doi.org/10.3390/s17040931>
21. Almaslukh, Bandar; Artoli, Abdel M.; Al-Muhtadi, Jalal. 2018. "A Robust Deep Learning Approach for Position-Independent Smartphone-Based Human Activity Recognition" *Sensors* 18, no. 11: 3726. <https://doi.org/10.3390/s18113726>.
22. E. Zdravevski et al., "Improving Activity Recognition Accuracy in Ambient-Assisted Living Systems by Automated Feature Engineering," in IEEE Access, vol. 5, pp. 5262-5280, 2017, doi: 10.1109/ACCESS.2017.2684913.
23. Charissa Ann Ronao, Sung-Bae Cho,Human activity recognition with smartphone sensors using deep learning neural networks,Expert Systems with Applications,Volume 59,2016,Pages 235-244,ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2016.04.032>.
24. M. Munoz-Organero, "Outlier Detection in Wearable Sensor Data for Human Activity Recognition (HAR) Based on DRNNs," in IEEE Access, vol. 7, pp. 74422-74436, 2019, doi: 10.1109/ACCESS.2019.2921096.
25. K. Xia, J. Huang and H. Wang, "LSTM-CNN Architecture for Human Activity Recognition," in IEEE Access, vol. 8, pp. 56855-56866, 2020, doi: 10.1109/ACCESS.2020.2982225.
26. Ordóñez, Francisco J.; Roggen, Daniel. 2016. "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition" *Sensors* 16, no. 1: 115. <https://doi.org/10.3390/s16010115>.
27. Terry T. Um, Franz M. J. Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. 2017. 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 (ICMI '17). Association for Computing Machinery, New York, NY, USA, 216–220. DOI: <https://doi.org/10.1145/3136755.3136817>.
28. Yang, Rong, and Baowei Wang. 2016. "PACP: A Position-Independent Activity Recognition Method Using Smartphone Sensors" *Information* 7, no. 4: 72. <https://doi.org/10.3390/info7040072>.
29. Masud Ahmed, Anindya Das Antar, Tahera Hossain, Sozo Inoue, and Md Atiqur Rahman Ahad. 2019. POIDEN: position and orientation independent deep ensemble network for the classification of locomotion and transportation modes. 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 (UbiComp/ISWC '19 Adjunct). Association for Computing Machinery, New York, NY, USA, 674–679. DOI: <https://doi.org/10.1145/3341162.3345570>.
30. Miao, F., He, Y., Liu, J. *et al.* Identifying typical physical activity on smartphone with varying positions and orientations. *BioMed Eng OnLine* **14,**32 (2015). <https://doi.org/10.1186/s12938-015-0026-4>.
31. Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. 2011. Activity recognition using cell phone accelerometers. SIGKDD Explor. Newsl. 12, 2 (December 2010), 74–82. DOI: <https://doi.org/10.1145/1964897.1964918>.
32. Jorge-L. Reyes-Ortiz, Luca Oneto, Albert Samà, Xavier Parra, Davide Anguita,Transition-Aware Human Activity Recognition Using Smartphones,Neurocomputing,Volume 171,2016,Pages 754-767,ISSN 0925-2312. <https://doi.org/10.1016/j.neucom.2015.07.085>.
33. M. Radhakrishnan, S. Eswaran, A. Misra, D. Chander and K. Dasgupta, "IRIS: Tapping wearable sensing to capture in-store retail insights on shoppers," *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2016, pp. 1-8, doi: 10.1109/PERCOM.2016.7456526.
34. Yang, Jian-Bo & Nhut, Nguyen & San, Phyo & li, Xiaoli & Shonali, Priyadarsini. (2015). Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition. IJCAI. [https://doi.org/10.1155/2015/258619](https://doi.org/10.1155/2015/258619" \t "_blank).
35. X. Liang, Y. Zhang, G. Wang and S. Xu, "A Deep Learning Model for Transportation Mode Detection Based on Smartphone Sensing Data," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 12, pp. 5223-5235, Dec. 2020, doi: 10.1109/TITS.2019.2951165.
36. Preeti Agarwal, Mansaf Alam,A Lightweight Deep Learning Model for Human Activity Recognition on Edge Devices,Procedia Computer Science,Volume 167,2020,Pages 2364-2373,ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2020.03.289>.
37. Yu Zhao, Rennong Yang, Guillaume Chevalier, Ximeng Xu, Zhenxing Zhang, "Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors", *Mathematical Problems in Engineering*, vol. 2018, Article ID 7316954, 13 pages, 2018. <https://doi.org/10.1155/2018/7316954>.
38. Hong Zhao, Chunning Hou, Hala Alrobassy, Xiangyan Zeng, "Recognition of Transportation State by Smartphone Sensors Using Deep Bi-LSTM Neural Network", *Journal of Computer Networks and Communications*, vol. 2019, Article ID 4967261, 11 pages, 2019. <https://doi.org/10.1155/2019/4967261>.
39. Hui Xing Tan, Nway Nway Aung, Jing Tian, Matthew Chin Heng Chua, Youheng Ou Yang,Time series classification using a modified LSTM approach from accelerometer-based data: A comparative study for gait cycle detection,Gait & Posture,Volume 74,2019,Pages 128-134,ISSN 0966-6362, <https://doi.org/10.1016/j.gaitpost.2019.09.007>.
40. J. Wang, Q. Long, P. Rahc, K. Liu, and Y. Xie, “Human action recognition on cellphone using compositional bidir-LSTM-CNN networks,” in *Proceedings of the 2019 International Conference on Computer, Network, Communication and Information Systems (CNCI 2019)*, 2019.   
    <https://doi.org/10.2991/cnci-19.2019.95>.
41. Y. Qin, H. Luo, F. Zhao, C. Wang, J. Wang and Y. Zhang, "Toward Transportation Mode Recognition Using Deep Convolutional and Long Short-Term Memory Recurrent Neural Networks," in *IEEE Access*, vol. 7, pp. 142353-142367, 2019, doi: 10.1109/ACCESS.2019.2944686.
42. S. Deep and X. Zheng, "Hybrid Model Featuring CNN and LSTM Architecture for Human Activity Recognition on Smartphone Sensor Data," *2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, 2019, pp. 259-264, doi: 10.1109/PDCAT46702.2019.00055.
43. T. Su, H. Sun, C. Ma, L. Jiang and T. Xu, "HDL: Hierarchical Deep Learning Model based Human Activity Recognition using Smartphone Sensors," *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1-8, doi: 10.1109/IJCNN.2019.8851889.
44. Sansano, E, Montoliu, R, Belmonte Fernández, Ó. A study of deep neural networks for human activity recognition. *Computational Intelligence*. 2020; 36: 1113– 1139. <https://doi.org/10.1111/coin.12318>.
45. A. Henpraserttae, S. Thiemjarus and S. Marukatat, "Accurate Activity Recognition Using a Mobile Phone Regardless of Device Orientation and Location," *2011 International Conference on Body Sensor Networks*, 2011, pp. 41-46, doi: 10.1109/BSN.2011.8.