Human Activity Recognition using Deep Learning Approach

MSc Research Project
Masters in Data Analytics

Laiba Rehman
x20144032

School of Computing
National College of Ireland

Supervisor: Mohammed Hasanuzzaman

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Laiba Rehman……………………………………………………………………………

Student ID: 20144032…………………………………………………………………………….……

Programme: Masters in Data Analytics…………… **Year:** 2021-2022…..

Module: MSc Research Project…………………………………………………………….………

Supervisor: Mohammed Hasanuzzaman…………………………………………………….………

Submission Due Date: 31st January 2022………………………………………………………….………

Project Title: Human Activity Recognition using Deep Learning Approach…………………………….………

Word Count: 8238……………………… **Page Count:** 23…………….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Laiba……………………………………………………………………………………

Date: 31st January 2022…………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Human Activity Recognition using a Deep Learning Approach

Laiba Rehman
x20144032

Abstract

This research is performed to detect human activities from the triaxial body acceleration and angular velocity recorded by accelerometer and gyroscope sensors from 30 volunteers. Throughout the research, we have performed excessive data exploration, data preprocessing, visualization and build deep learning models where we performed single and multi-layer LSTM model, CNN model, and divide-and-conquer based CNN model. t-distributed Stochastic Neighbour Embedding (t-SNE) is performed to separate the data belonging to 6 activities (walking, walking upstairs, walking downstairs, sitting, laying and standing) using different perplexity values. On the test data, we have applied t-SNE of 20, 50 and 90 perplexity values and observed that the laying activity is distinct from all other activities whereas standing and sitting activities overlap each other. Throughout the research, we also performed hyperparameter tuning where we picked the best parameters from all the models to give the best accuracy. The combined divide and conquer-based model where data sharpening is applied to combine the static and dynamic activities and fed into a final pipeline with the different number of epochs give a training accuracy of 98.3% and test accuracy of 96.8%. These indicate that the combined approach can be used in a better way compared to other deep learning methods. Also, the higher accuracy indicates that this can be also improved in future research where we can see that Static and dynamic activities are highly predicted with the help of divide and conquer-based approaches.

Keyword: LSTM, CNN, Divide and conquer, hyper parameter, etc.

1 Introduction

1.1 Background

Globally, there were 703 million individuals over the age of 65 in 2019, and this figure is expected to increase to 1.5 billion by 2050. As a result, their health issue poses a significant concern to the globe. To address this issue, an increasing number of mobile gadgets for monitoring people's health have been created. Human Activity Recognition (HAR), which analyzes and extracts important features from human physidcal activities such as walking upstairs, jumping, standing, laying, cooking, walking downstairs, laying, sitting, and eating various foods, etc plays a critical role in monitoring people's behaviors and assisting them in maintaining a healthy lifestyle. Furthermore, they are extensively employed in a variety of fields, including computer interaction, human-robot interaction, autonomous driving, smart homes, and smart education. Wearable devices and smartphones equipped with inertial sensors such as magnetometers, accelerometers, and gyroscopes may be used to efficiently gather time-series data and infer facts about human activities. As a result of their obvious benefits over other sensor modalities, sensor data collected for HAR tasks by mobile devices have dominated this study sector.

1.2 Motivation

HAR has been one of the most active research areas in the world of computers in recent years because of its broad application in a variety of fields including as eldercare, video surveillance, health, security applications, AR/VR and human-computer interaction. With the use of video monitoring, a computer's ability to recognize human behavior might be used to identify possible security threats over large geographic areas. The activities of older adults can be monitored, and with the correct feedback system in place, unexpected negative consequences such as sudden health issues may be identified early and effectively treated before they worsen. Patients suffering from mental disorders or diseases like Parkinson's Disease Dementia may also utilize HAR to monitor their actions daily and notice any abnormalities. Since the inception of HAR research, the performance of activity recognition has increased dramatically, such as the types of activities performed by human subjects, the algorithms used to train the classifiers, feature transformation, feature selection and extraction approaches, the length of time series data segments, the sensors utilized, and the signal sampling rates, and so on. Due to these choices, comparing multiple HAR techniques becomes difficult. The study of human behaviors captured by sensors and other devices has become a difficult task, which is why machine learning and deep learning models are increasingly utilized to find useful patterns within human activities and anticipate various health risks.

1.3 Deep learning Overview

Deep learning models are often advantageous compared to different machine learning models as machine learning models fail to extract complex patterns that can only be possible by Deep learning models. Deep learning models apply neural networks where different layers are constructed to feed the input and the weights and find a relationship between the features existing in the data to give the desired output.

A neural network works just like a human brain that has a different number of inputs that have a relationship with the weights. The weights which are associated with each input undergo computation with different neurons after which aggregation of sum takes place by an activation function to provide the desired outcome (Halvaei Niasar and Rahimi Khoei, 2015).

1.4 Rationale

For the implementation of the data into the system to perform various neural network models, we have used Jupyter notebook to perform deep learning and Python to perform all the analysis throughout the research. We used Jupyter Notebook in the Anaconda environment where we pre-installed all the Python libraries and other deep learning libraries like TensorFlow and Keras.

We have used the following tools, libraries and hardware system throughout the research

1. Python Language
2. Anaconda Navigator
3. Intel i5 processor with 8GB RAM
4. Tensorflow and Keras
5. Scikit learn library

We have also collected various resources such as journals and articles which are necessary to carry out the critical review of the approaches used by previous researchers. We have taken the help of various research articles and journals where only top peer-reviewed journals were collected from different sources such as Researchgate, Scopus, Semantic Scholar, Google Scholar, etc.

1.5 Research Question

"With the use of Deep Learning models, how well can we predict the accuracy of Human Activities using sensor data acquired by the accelerometer and gyroscope features accessible on smart phones?"

1.6 Aims and Objectives of the Research

The project aims to predict human activities from the body acceleration and velocities recorded by the smart sensors using deep learning methods. The objectives are
1. This project will utilize the UCI Human Activity Recognition dataset to develop Deep Learning models and then choose the optimum one depending on metrics. (Recall, precision, F1-score, and accuracy).
2. Additionally, a novel divide and conquer strategy for developing two-stage HAR is suggested, that, during the prediction phase, integrates sharpening test data to enhance HAR's performance.
3. To show that the proposed approach is capable of producing more complicated and meaningful characteristics.

2 Related Work

Several academics have already recognized human activities using a variety of algorithms and we will try to list some of the researches and approaches used by them along with their results and drawbacks of their approaches.

2.1 HAR on Machine learning based Approaches

Human activity recognition is also applied in a study conducted by Bhatt and Bhatt (2019) in which they focus exclusively on the EEG dataset. They classified the EEG SEED dataset using an SVM classifier with external library LibSVM (3.23), achieving a significant increase in performance and accuracy. Additionally, many current techniques utilizing various classifiers are examined on the dataset, including SVM with KNN, ELM SVM with DEAP dataset, and SVM with SEED dataset. They were able to boost the performance of each run by 4% utilizing the library LibSVM (3.23), resulting in 79.38 percent accuracy in a tensor flow environment. They have emphasized the importance of their study in determining electrical patterns in human brains, diagnosing epilepsy and seizure, and confirming brain death. They advised that for future study, to further enhance accuracy, they may utilize a winner take all technique to categorize the dataset into three segments: neutral, negative, and positive. Another possibility can be to employ additional classifiers, such as ELM, that are compatible with the EEG dataset.

In this research, the authors Ziaeefard and Bergevin (2015) have highlighted a failure that recent action recognition methods that depend on mid and low-level features like trajectories and Spatio-temporal interest points, although they produce decent results, they fall short when dealing with complicated data due to the absence of semantics represented by them. They concentrated on current frameworks for action recognition based on semantic information in their research. They established a semantic space composed mostly of attributes, object/scene context, poselet and pose. Their Experiments demonstrate that semantic approaches outperform non-semantic methods in the majority of circumstances, especially when the poses/attributes of distinct activities are identical. Combining numerous features enhances performance in some circumstances. Additionally, they argued that in order to fully leverage the potential of semantic techniques, many issues require additional exploration. One potential path for semantic techniques could be to detect novel activities that have not been encountered in training instances (termed as zero-shot learning) and other innovative subjects could be activity analysis. and activity forecasting.

To develop a commercially viable HAR device, this research conducted by Nguyen et al., (2021) reviews state-of-the-art classification and power consumption approach like convolutional neural networks (CNN), data compression, and other new techniques. The study of the present literature establishes a foundation for HAR by recommendation of power reduction and classification techniques, analyzing the lack of available HAR datasets, current drawbacks, and their associated remedies and future developments in HAR. According to the authors, the absence of publicly available datasets makes it hard for new users to learn about HAR. This article devotes a part to freely accessible datasets.
This article discusses typical power reduction approaches that can aid in the acceleration of the process of incorporating HAR into wearable devices, which are a popular interest for self-care. Numerous branches and applications of HAR may be explored with the usage of wearable devices.

Human activity recognition was also observed in a paper produced by Mittal et al., (2021) where they studied various machine learning algorithms. They used algorithms such as Logistic regression, AdaBoost, Support Vector Machine, K-NN, Random Forest, Gaussian Naïve Bayes, and Discriminant Analysis. They have done the model selection technique with the help of K-fold cross-validation. They suggested a notion if they separated people according to their walking styles and ascertain any further judgments, these insights may be used to establish real-time monitoring of human assets in highly secure areas, to monitor elderly individuals with movement disorders or diseases for any difficulties based on their movement patterns. They determined that their work has an accuracy of above 94 percent in terms of recognition.

2.2 HAR on Deep Learning based approaches

Various deep learning approaches were also used to detect human activities on the data recorded by wearable sensors. The purpose of this study, proposed by Zheng (2021), is to present a unique deep learning model (termed as LGSTNet) capable of analyzing and extracting Local and Global Spatial-Temporal characteristics from sensory input for HAR. LGSTNet's concept is to divide an activity window into many sub-windows and then connect the Attention mechanism to a 2D CNN in order to accurately learn local spatial-temporal properties from those sub-windows. Similarly, a three-dimensional Convolutional Neural Network (CNN) is meant to learn global spatial-temporal properties from the whole window. The ablation and comparison are carried out on two publicly available datasets, UCI-HAR and WISDM. LGSTNet outperforms all the other models on both datasets. The only drawback of the study is the high computational requirement that takes a lot of training time which is not feasible for different types of data. The author also suggested that future work will focus on further improving the accuracy of recognizing fine-grained activities such as brushing teeth, drinking, eating various foods, and performing various daily routines. This will be accomplished by analyzing additional features obtained from multiple sensors installed in various locations on the body.

Suto (2021) have used Artificial Neural Network where they have taken component vector from the time window which they built as a contribution to their approach. The information of the location is not collected from their approach, it had been gathered from various human activities. They have used Rectified Linear Unit and Xavier in order to reduce the learning rate in their classification model. They set the learning rate to 0.01 with Adam optimizer since it can learn the outcomes quickly. With the combination of all these parameters, they have gained 90% accuracy on the data set accorded by acceleration sensor and 95% accuracy is gained on the data set which includes information of the location.

This research conducted by Cheng, Huang and Zong (2021) discusses a machine learning technique called Gaussian Mixture Hidden Markov Model (GMM-HMM) that is used to recognize device-free activities using WiFi Channel State Information (CSI). The authors proposed that human activity may be detected via the use of a unique mapping between signal and action fluctuations. On self-collected datasets, the suggested method gives an average accuracy of better than 97 percent. Thus, by achieving great results authors concluded that the detected activities comprise two distinct kinds of falling actions (walk-fall and stand-fall), indicating that the system has significant potential to be a non-intrusive, practical solution for fall detection and activity recognition. The answers to the difficulties of identifying the actions of several persons, identifying equally fine-grained human activities, and improving the system's resilience in complicated contexts are critical in engineering applications. These mentioned challenges are left for future work by the authors.

The purpose of this research proposed by Triwiyanto, Pawana and Purnomo (2020) is to optimize performance, simplify the deep learning architecture, and establish the optimal hyperparameters for categorizing 10 hand movements from two raw EMG signals using the convolution neural network (CNN) technique. The research establishes that the suggested technique outperforms existing classifiers in terms of accuracy. The prediction with the greatest average accuracy was 0.93. Authors suggested another approach that the proposed method could also be implemented on the Raspberry Pi system for developing prosthetics hand or on embedded system which depends on Digital Signal Processor (DSP). Additionally, they suggested that future research should concentrate on developing a new architecture for deep learning and compensating the system for higher performance.

2.3 Sensor placed on Positions

The optimal location of sensors on different body postures has been widely explored for a variety of activity identification applications, ranging from fall detection and gait tracking in older adults to daily life activities in healthy participants with no impairments. The authors Wang et al. (2017) analyzed the cross-location activity identification problem using a single IMU on-body sensor and suggested a transfer learning model based on the HAR model in which a new sensor position is adapted utilizing previously recognized sensor placements.

In this paper given by Atallah et al. (2011), it divided activities into four categories: very low level, low level, medium level, high level, ranging from static to dynamic, based on the combination of physical activities that raise the rate of energy consumption. Apart from the ear-worn sensor, the wrist, chest, waist, knee, arm, and ankle all include accelerometer sensors. According to the findings of the trials, the knee sensor is most effective at detecting high-intensity action such as running. The locations of the ear and arm sensors were next best. However, the authors note that feasibility should be examined since sensors placed on the knees or chest might disrupt everyday activities. As a consequence, a constant but realistic sensor position is recommended, depending on the application.

The authors of this work Hassan et al., (2018) developed a smartphone-based approach for detecting human activity using inertial sensors. Initially, raw data is utilized to extract efficient characteristics. The characteristics include mean, median and autoregressive coefficients. Finally, activity recognition is carried out using Deep Belief Networks (DBNs) and compared to more standard techniques such as SVMs and ANNs, where DBNs surpass them.

Conclusion

Based on all the approaches, we have seen a quick overview of different techniques used to determine activities recorded by different types of sensors. We also discussed various machine learning approaches used by other researchers in predicting human activities. Since our research is based on Deep learning models on detecting human activities. We have seen that most of the researchers have used concepts such as TCN network, LSTM and CNN that were proven effective in different types of data such as channel state information data as well as data recorded by different sensors. This is why we will develop divide and conquer based CNN model where we will divide the activities based on Static and dynamic activities from the raw data recorded by accelerometer and gyroscope sensors. We will also apply LSTM and CNN network and will compare with the divide and conquer based approach.

3 Research Methodology

The methodological component of every study proposal is critical, because it entails an examination of the tactics and procedures that will be used in the study. We may achieve high accuracy, low cost and time management by using a well-defined procedure. The technique used in this work is Knowledge Discovery in Databases (KDD). KDD is selected due of its frequent interactive and iterative and participatory use. At any time throughout this procedure, looping and iterations back to previous phases are also conceivable (Sharma and Gaur, 2016). These are the steps that we will be following for our research evaluation.

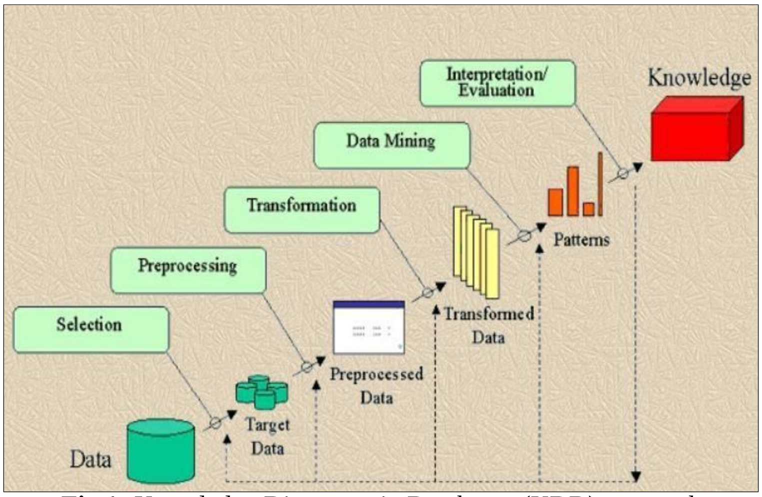

Fig 1: Knowledge Discovery in Databases (KDD) approach

4 Design Specification

4.1 t-SNE

t-SNE also known as t-distributed stochastic neighbour embedding is considered as a machine learning and statistical technique that helps to reduce the dimension of the data which further helps to identify patterns that are relevant in nature. The main advantage of applying t-SNE in the data is to preserve local structure which means that data points which are closer to one another in high dimensional space will also be closer to low dimensional space. It is very well suited compared to other techniques as it reduces the data in high dimensional space which are nonlinear in nature to a low dimensional space to produce beautiful visualizations that can help us find important patterns. The algorithm tries to compute the probabilities of the neighbors around each data point. Neighbors are considered as that group of points that are closer to each other (R.Banupriya and Karthik, 2018).

- **Parameter and Assumptions**

The main parameter considered in this algorithm is perplexity. Perplexity is roughly estimated as the number of neighbors calculated by finding the difference between the Gaussian distribution and the t-distribution data points. A very low perplexity indicates focusing on the data point on a local scale and high perplexity considers the data point on a large scale. (R. Banupriya and Karthik, 2018).

4.2 LSTM

LSTM or Long ShortTerm Memory Network is one of the expanding networks of the Recurrent Neural Network family with provides better performance in various temporal schemes. It gives a remarkable performance owing to the issue of vanishing gradient problems and is used in various classifications in time-series data (Abbasi et al., 2019).

In a Recurrent Neural Network, RNN can predict the output based on previous information but can recognize the input only for a short period of time due to a vanishing gradient issue. To tackle this program, LSTM is used that can handle the problem of long-term dependency. It performs better by recognizing the data through an input gate and overriding the new data by comparing the inner memory through a forget Gate (Abbasi et al., 2019).
It consists of an input gate, a forget gate and an output get and arranged in such a manner to manipulate the data that should be recognized as well as disremembered.
In LSTM network, gating technique is used to carry required data and input and forget gate is used to keep the gradients to flow for a long period of time.

4.3 CNN model

Convolutional neural network or CNN are used in image processing techniques where features of all the images are represented as matrix vector data are fed into a neural network model as input. Various filters act as a feature map that is used to extract features from all the images. CNN is also used in text and numeric data where the features should be fed into the input as a matrix of a particular dimension. There are different layers present in a convolution neural network and these layers have particular functions to pass the input to an activation function to give the desired output (Kamundala and Kim, 2018).

4.4 Divide and Conquer based Approach

The human activities recognition system that is used in our research contains data of six human activities. These activities are divided into Static and dynamic activities to implement divide and conquer approaches. The static approaches contain activity such a sitting, standing and laying where a human body does not undergo any physical movement. Dynamic activities contain activities such as walking, walking downstairs, and walking upstairs where human body moves and contain different acceleration. In this approach a one-dimensional convolutional neural network will be applied in classifying both Static and dynamic activities using the concept of test data sharpening. The difference in this model is to look into the performance after splitting the data based on Static and dynamic activities and applying an individual neural networks into them. The architecture of the divide and conquer based model is represented below.

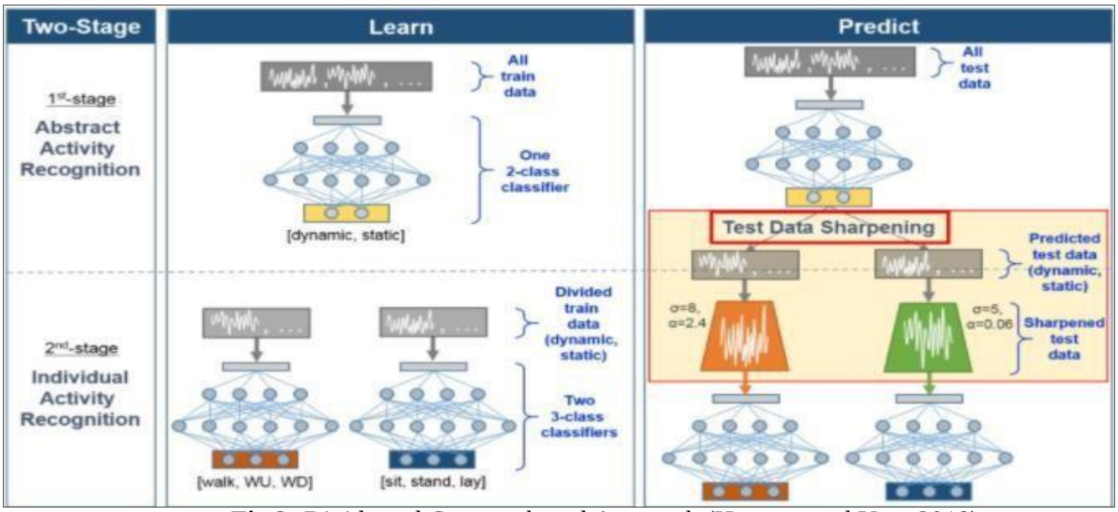

Fig 2: Divide and Conquer based Approach (Heeryon and Yun, 2018)

4.5 Dataset Description

The data is collected from the UCI repository that contains human activities collected from 30 persons performing different activities with smartphones worn to the waist. The data is recorded with the help of an accelerometer and gyroscope sensor in the smartphone and contains different activities such as walking, standing, laying, sitting, walking upstairs, and walking downstairs. The accelerometer recorded the linear acceleration of the human body and the gyroscope recorded the angular velocity with several variations. The size of the data set is around 10,000 rows with 561 features (M Bani Amer, 2021).

The acceleration signal recorded by accelerometer was separated into body in gravity acceleration signals with the help of low pass filter with frequency of 0.3 Hz. The several jerk signals were obtained with the help of body linear acceleration and angular velocity and was separated with the help of Euclidean nom that gave three dimensional signals as features. For training the models, we have split the data into 70% and 30% where we took 70% of the data for training and 30% of the data for testing.

All the acceleration readings and gyroscopic readings have X, Y and Z components and jerk signals are calculated for body acceleration. Altogether the data constituted 561 features in 6 classes which are split into training and testing (M Bani Amer, 2021).

5 Implementation

The implementation process includes different design stages that are used in implementing neural network models on our data. Such stages include collection of data, exploration of the data, preprocessing of data, visualizing different class as well as building the models on the data.

5.1 Collection of data

The collection of data includes downloading a data from UCI repository and storing it in the local system in a separate folder. Also, the steps include importing the data into the Jupyter notebook with the help of Pandas library provided by Python. All of the data is stored into different variables for easy implementation in our model and explained in the data description which is imported with the help of Pandas in our notebook.

5.2 Exploration of data

Data exploration is an important step which is used to look into the nature of the data, presence of any missing variables and duplicate values present in the data. Also, the exploration helps us to understand how badly or good the data is and it leaves us a choice to treat the data if it requires cleaning (Chen and Wu, 2017). In this step we are going to look into the static and dynamic activities with the help of visualization and into the presence of missing and duplicate values inside the data. We will also look into the body and gravity acceleration with the help of box plot in order to understand acceleration during Static and dynamic activities. The flowchart of the entire research process is given below

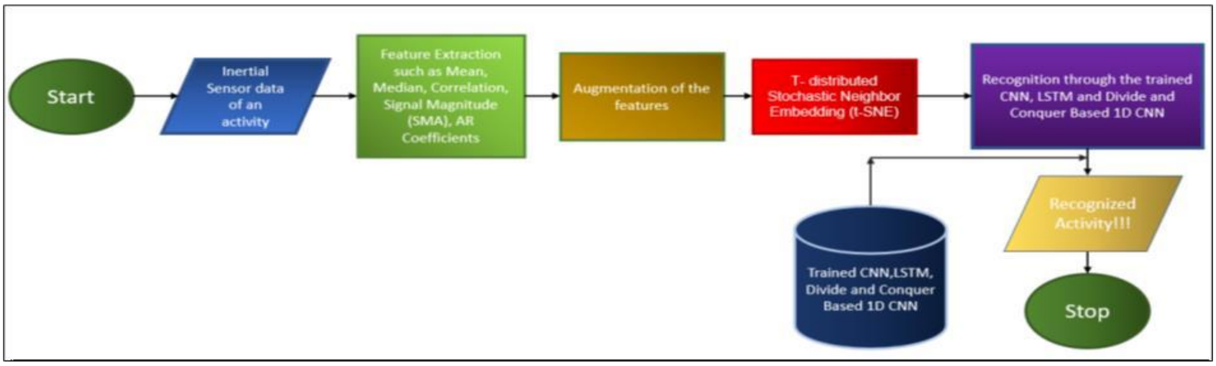

Fig 3: Proposed Architecture

5.3 Data Preprocessing

This step includes prepossessing of the features which are needed to be treated before feeding into the model (Chaudhary and Roy Chowdhury, 2019). In this step, we have performed scaling on the data where we scale all the features to bring into a different range before feeding into any deep learning models. There is no other transformation needed in preprocessing the data as the data is found to be very clean and balanced in nature and no other preprocessing steps such as categorical encoding, treating of outliers and missing values is done except scaling the data.

5.4 Data visualization

Data visualization helps us to understand the distribution of the data that includes the distribution of various features present in the data. In this step we are going to visualize all the activities with the help of the t-Distributed Stochastic Neighbour Embedding that will convert all the data from high dimensional to 2-dimensional space. We will use different perplexity values in order to separate the data in order to understand how distinct the classes are from each other. Also, during the exploration of the data, we are going to visualize the acceleration of the body during different activities with the help of box plot and also look into the distribution of the acceleration during different activities with the help of a line plot.

5.5 Building models

The building of the model includes training the model with the help of train data and predicting the accuracy with the help of test data. In this step we are going to build 3 models which are:

a) LSTM
In LSTM model, we will build a single layer and multilayer LSTM model and also a model with regularization technique in order to see if there is an improvement in accuracy compared to the baseline model.

b) Convolution Neural Network
In CNN model, we will build a baseline model and a model using L2 regularized parameter and compare the accuracy of the model.

c) Divide-And-Conquer based CNN model

In divide and conquer based approach we will build two models using three classes which will divide the model into Static and dynamic activities and build CNN model on each of them. We will also perform base line model and baseline model with regularized parameter to see if there is an improving accuracy in both the models.

6 Results and Findings

The Results and findings discuss the exploration of the data such as the data provided by a user, data points per Activity, the body acceleration and the angular velocity of moving and stationary activities, the classes of the data separated from each other with the help of t-SNE and the results obtained from different deep learning models. Let us discuss one by one obtained from the model starting from the exploration of the data to the evaluation of all the models.

6.1 Data Exploration

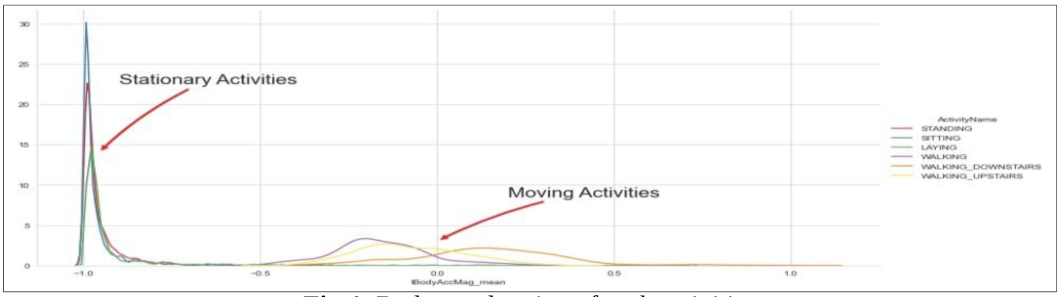

Fig 4: Body acceleration of each activities

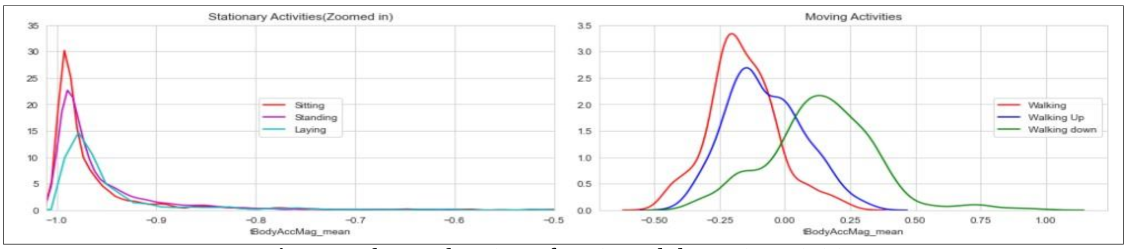

Fig 5: Body acceleration of static and dynamic activities

The following plots compare the mean body acceleration between stationary and moving activities. We can see that the mean body acceleration ranges between -1 to -0.5 in case of stationary activities and -0.5 to 1 in case of moving activities. This tells that the mean body acceleration increases during moving activities as the acceleration increases when human body starts to move.

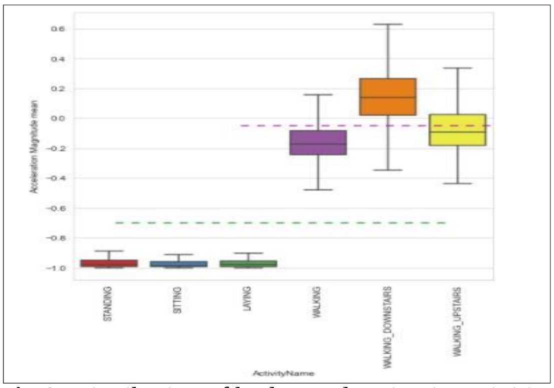

Fig 6: Distribution of body acceleration in activities

The box plot demonstrates the distribution of mean body acceleration during static and dynamic activities of a human body.

Observation

- If tAccMean is < -0.8 then the Activities indicate static activities i.e, Standing, Sitting or Laying.
- If tAccMean is > -0.6 then the Activities indicate dynamic activities i.e, Walking, WalkingDownstairs or WalkingUpstairs.
- If tAccMean > 0.0 then the Activity is WalkingDownstairs.

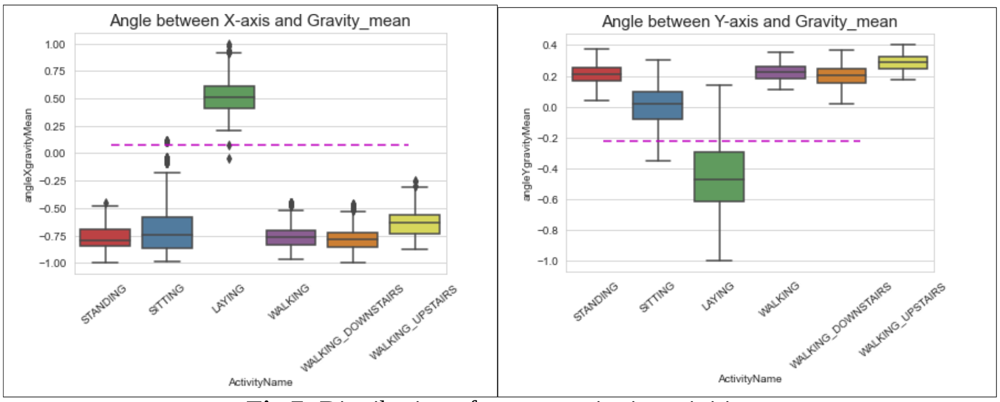

Fig 7: Distribution of mean gravity in activities

This plot tells about the mean gravity recorded by different sensors during different activities and the distribution of mean gravity during Static and dynamic activities.

Observation

- If the angle between X-axis and mean gravity > 0, then the activity indicates laying activity
- If the angle between y-axis and mean gravity < -0.2, then the activity indicates laying activity

6.2 t-SNE Findings

This visualization includes the conversion of training and testing data into 2-dimensional space using different perplexity values. The change in perplexity values will distinct all the activities in a much better way. In the train data, the perplexity value of 2,5,10 and 20 is applied to separate the activities in a much better way.

1. t-SNE on train data

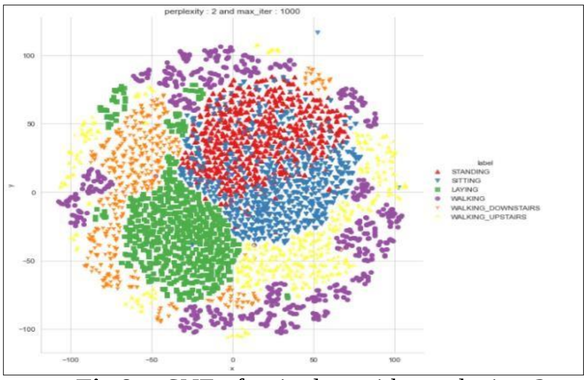

Fig 8: t-SNE of train data with perplexity=2

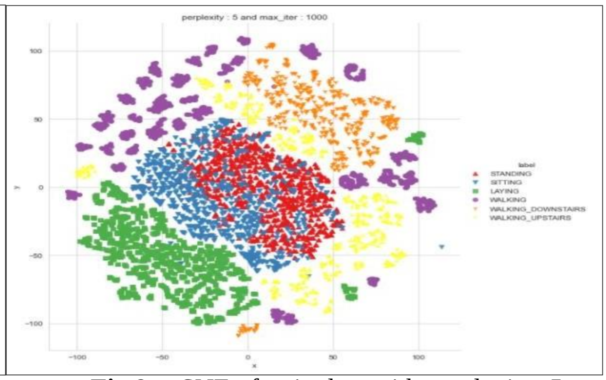

Fig 9: t-SNE of train data with perplexity=5

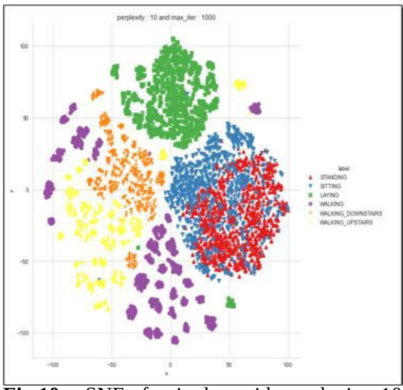

Fig 10: t-SNE of train data with perplexity=10

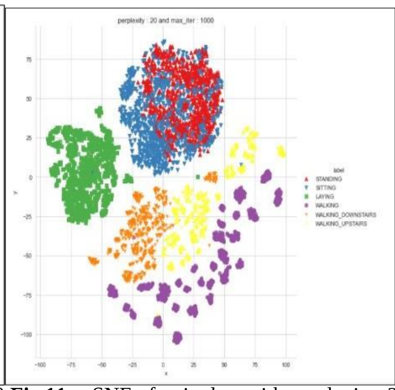

Fig 11: t-SNE of train data with perplexity=20

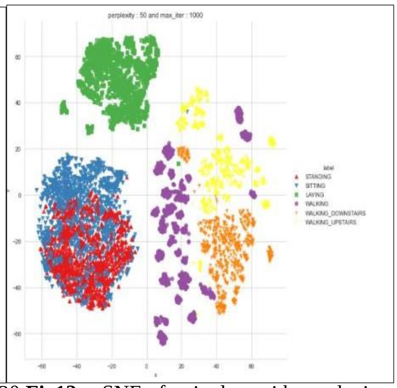

Fig12: t-SNE of train data with perplexity=50

With such a low perplexity 2, we can see that most of the data classes are overlapping with each other such as the laying activity is overlapping with standing and Sitting activities and also with walking downstairs activities. Let us look into the distinction of the classes after increasing the perplexity value.

After increasing the perplexity value to 5, we can see that the data is slightly separated compared to the perplexity value of 2. Even the sitting and standing activities are overlapping with each other because of the same features.

From the perplexity value of 10 and 20, we can see that the data is separated where the laying activities are separated from each other whereas Standing and Sitting activities are separated from all other activities. Also, the walking activity is separated from walking upstairs and downstairs activities.

Finally, after setting the perplexity value to 50, we can see that the laying activities are clearly distinct able from other dynamic activities. This experiment explains to us that the features of this activity are different compared to the dynamic activities. Moreover, the features on the Standing and Sitting activities are quite same as they are overlapping with each other in different perplexity values and the features of the walking activities are separable compared to the walking upstairs and downstairs activities.

2. t-SNE on test data

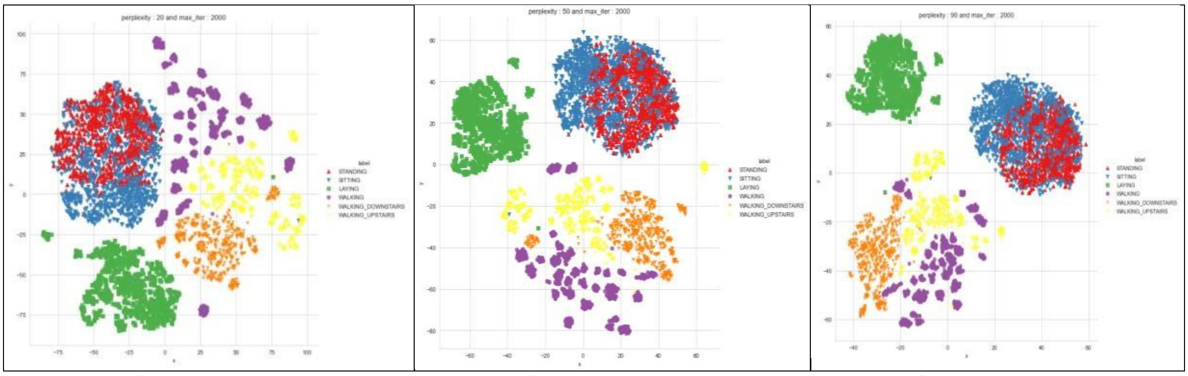

Fig13: t-SNE of test data with perplexity=20 **Fig14:** t-SNE of test data with perplexity=50 **Fig15:** t-SNE of test data with perplexity=90

After performing t-SNE on test data with perplexity value of 20, we can see that the activities are slightly separable but some of the features of the activities like walking upstairs and down stairs are slightly overlapping. Let us increase the perplexity value and look into the separation of the classes.

After setting the perplexity below to 50, we can see that the static activity is separable compared to the dynamic activities but the walking activities are not clearly separable compared to the walking upstairs and downstairs activities.

After setting the perplexity better to 90, we can see that the static and dynamic activities are clearly separable where the Standing and Sitting activities are same as the features of them are quite same. Also, the laying activity is easily distinct able from all other activities. Also the dynamic activities are separable from each other where the walking, walking downstairs and walking upstairs activities have different features that are separable from each other.

6.3 Evaluation of models

Different models are implemented with different combinations of layers such as one layered LSTM, two LSTM, 2 layered LSTM with L2 regularizer, CNN model, and CNN model with regularizer. A divide and conquer based model is implemented by dividing the activities as static and dynamic and converting the model from 6 classes to 2 classes.
The evaluation of the model is done on LSTM and CNN models having 6 classes based on metrics such as Precision, Recall and F1 score.

Let us look into the performance of the models on 6 classes given by each algorithm.

Model	Precision	Recall	F1-score	Regularizer	Epochs
LSTM(1 layer)	0.9	0.9	0.9	No	30
LSTM(2 layers)	0.9	0.9	0.9	No	30
LSTM(2 layers)	0.92	0.91	0.91	L2	30
CNN	0.92	0.92	0.92	No	30
CNN	0.91	0.91	0.91	L2	30

Table 1: Evaluation metrics of LSTM and CNN models

- **Observation**

From the evaluation metrics performed on all the models, we can see that the F1-score given by the CNN model without a regularizer gave the best accuracy compared to all other deep learning models. CNN gave a 92% of F1 score indicating that there are only 8% of false positive and false negative data predicted by the CNN model. The evaluation on divide and conquer based model is not done as this model was performed only on two classes that divided all the six classes into Static and dynamic activities. Although the evaluation metrics are not performed, we have analyzed the accuracy of all the models based on training and testing accuracy.

Model	Train Accuracy	Test Accuracy	Regularizer	Epochs	Classes
LSTM(1 layer)	95.4	89.5	-	30	6
LSTM(2 layer)	95.75	89.92	-	30	6
LSTM(2 layers)	94.54	90.8	L2	30	6
CNN	99.2	92.4	-	30	6
CNN	94.87	90.9	L2	30	6
Divide and Conquer	1.00	99.8	-	30	2
CNN(Static)	97.34	93.53	-	30	3
CNN(Dynamic)	95.28	90.69	-	30	3

Table 2: Accuracies given by different models

The accuracies of all the models have been analyzed with and without a combination of regularizers on the different number of classes. The single-layered and multi-layered LSTM models and the CNN model are performed in the original data. The data is later divided into Static and dynamic activities and classified into groups where divide and conquer based model is performed to determine the accuracy. Moreover, a separate CNN model on static activities of three classes and dynamic activities of three classes are individually performed to determine the training and testing accuracies.

- **Observation**

From all the models, we can see that CNN model without a regularizer gave the highest training and testing accuracy compared to all other models that are classified based on six classes. It also supports the evaluation result given by the CNN model as because it gave the highest F1-score compared to other models. From this result we can tell that CNN model can classify the human activities in a better way compared to the LSTM models.

Additionally, the divide and conquer based model classified based on Static and dynamic activities gave a very high accuracy compared to other models. The CNN model based on static activities gave better accuracy compared to the model based on dynamic activities. The combined Static and dynamic activities based on divide and conquer based CNN model will be evaluated based on best parameters picked during hyper parameter tuning of all the models.

6.4 Hyper-Parameter tuning

LSTM, CNN and Divide and conquer based CNN models are performed in order to pick best parameters that give the highest accuracy. The best parameters given by all the models are evaluated based on training and testing accuracies. Let us look into the best parameters that are selected from hyper parameter tuning where each model gave its best accuracy.

- **LSTM Best parameters**

'Dropout'	0.3802031741395868
Dropout_1	0.6903389204823146
Dropout_2	0.3654341425327902
LSTM	38
LSTM_1	36
LSTM_2	32
Optimizer	Adam
l2	0.00015208023802140732
l2_1	0.000643128044948208
l2_2	0.0007102309264917989
lr	0.016347608866364167
lr_1	0.024543333891182614
Epochs	30

Table 3: Best parameters of LSTM

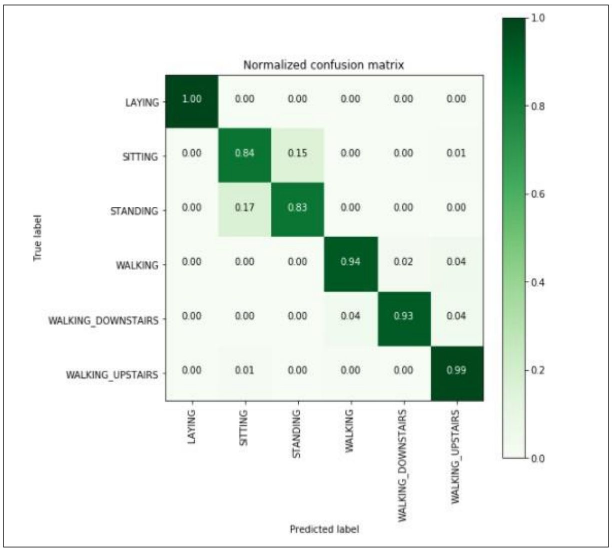

Fig 16: Confusion matrix by LSTM

From the combination of all the best parameters, the confusion matrix is made by the LSTM model.

- **Observation**

The LSTM model with best parameter perfectly classified the laying activities in the test data. Moreover, the activities such as walking upstairs, walking down stairs and walking activities are perfectly predicted with slight mis classification. There is a high misclassification observed in predicting sitting and standing activities by the LSTM model.

- **CNN Best parameters**

Dense	64
Dropout	0.6397045095598795
batch_size	64
optimizer	Adam
filters	32
filters_1	24
kernel_size	7
kernel_size_1	3
l2	0.07999281751224634
l2_1	0.0012673510937627475
lr	0.0011215010543928203
lr_1	0.0021517590741381726
nb_epoch	25
pool_size	3

Table 4: Best parameters of CNN

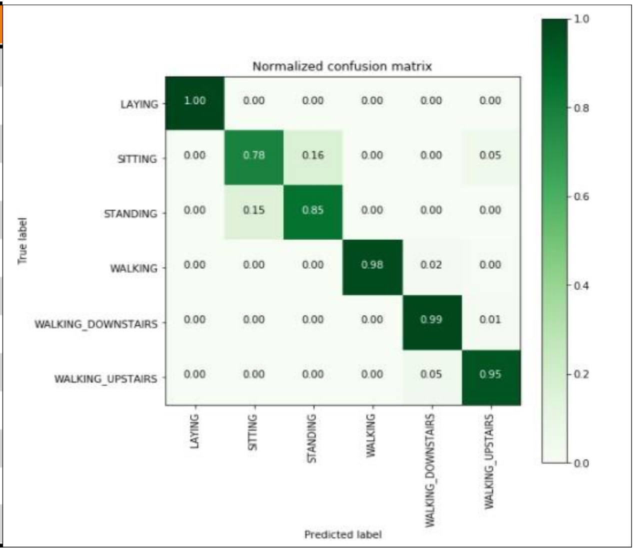

Fig 17: Confusion matrix by CNN

From the combination of all the best parameters, the confusion matrix is made by the CNN model as shown above.

- **Observation**

The CNN model with best parameters predicted the laying activity on the test data with 100% accuracy. The prediction of sitting activity is less compared to the LSTM model. All the dynamic activities have more accuracy compared to the LSTM model except the prediction of the walking upstairs activity.

- **1. CNN (Static) Best parameters (Divide and Conquer Based Approach)**

Dense	64
Dense_1	64
Dropout	0.45377377480700615
choiceval	rmsprop
filters	32
filters_1	16
kernel_size	5
kernel_size_1'	3
l2	0.0019801221163149862
l2_1	0.8236255110533577
lr	0.003918784585237195
lr_1	0.002237071747066137
nb_epoch	30
pool_size	2

Table 5: Best parameters of CNN (static)

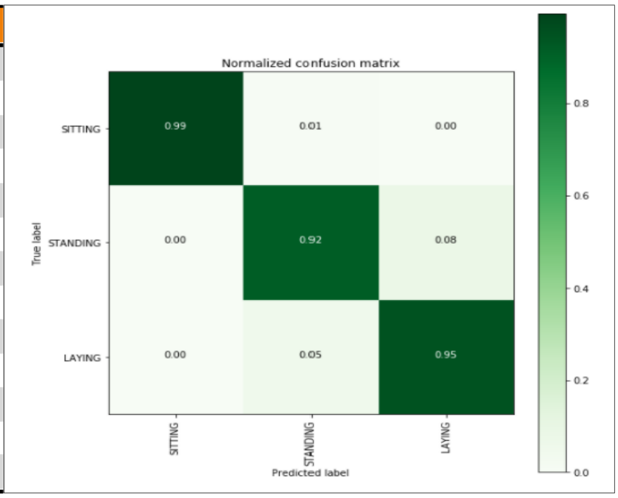

Fig 18: Confusion matrix by CNN (static)

From the combination of all the best parameters, the confusion matrix is made by the CNN model on static activities as shown above.

- **Observation**

The static activities predicted by the CNN model gives very high accuracy as all the activities such as sitting, standing and laying had very less misclassification by the CNN model with best parameters. All the static activities on the test data are predicted with more than 90% accuracy by the CNN model with best parameters.

2. CNN (dynamic) best parameters (Divide and Conquer Based Approach)

Dense	32
Dense_1	32
Dropout	0.48642317342570957
choiceval	Adam
filters	32
filters_1	32
kernel_size	7
kernel_size_1	7
l2	0.10401484931072974
l2_1	0.7228970346142163
lr	0.000772514731035696
lr_1	0.003074353392879209
nb_epoch	35
c pool_size	5

Table 6: Best parameters of CNN (dynamic)

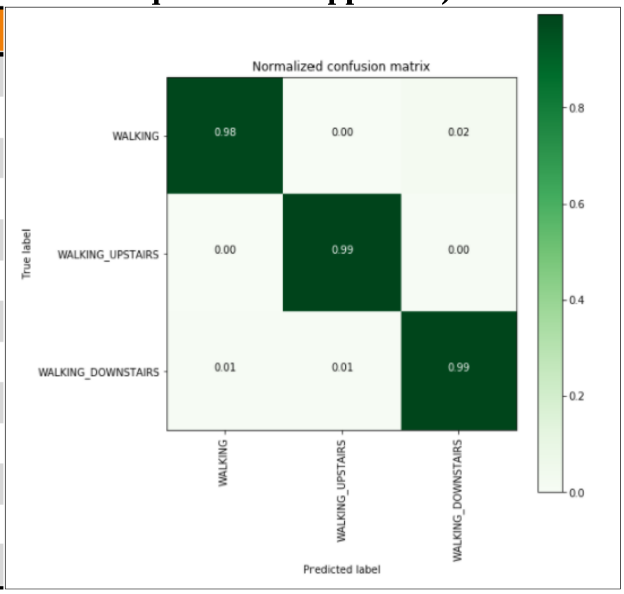

Fig 19: Confusion matrix by CNN (dynamic)

From the combination of all the best parameters, the confusion matrix given by the CNN model on dynamic activities is given above

- **Observation**

The dynamic activities such as walking, walking upstairs and walking down stairs are predicted with very high accuracy by the CNN model where a divide and conquer based approach is applied only on dynamic and static activities. The CNN model with best parameters can classify the dynamic activities with very high accuracy if static activities are not present in the data.

3. Combined CNN (Static and Dynamic)

The Classification accuracies given by the CNN model on classifying both Static and dynamic activities are fitted into a final pipeline in order to give the combined accuracy in classifying all the 6 human activities. The combined CNN model after building different CNN models based on divide and conquer based approach are evaluated based on confusion Matrix where the final pipeline model is classified as both Static and dynamic activities in a single model.

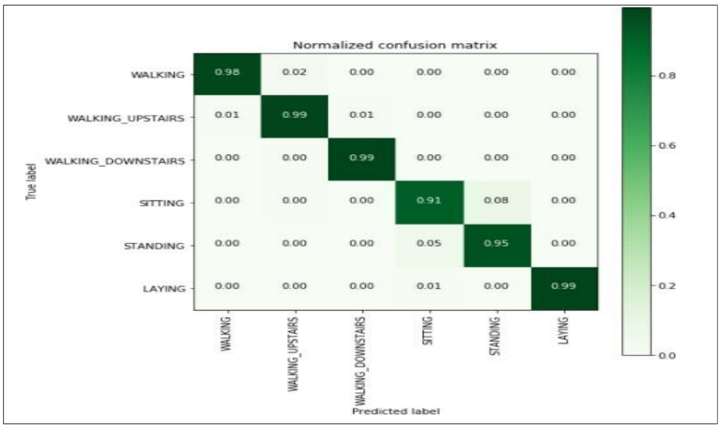

Fig 20: Confusion matrix by CNN (static + dynamic)

- **Observation**

The divide-and-conquer based approach gave the best accuracy as it gave more than 90% accuracy in predicting all the activities. LSTM and CNN model fails to detect sitting and standing activities with such accuracy compared to the divide-and-conquer based approach model.

5.5 Accuracies from Hyper parameter

Accuracy from Hyper parameter tuning			
Model	**Train Accuracy**	**Test Accuracy**	**Epochs**
LSTM	**94.56**	**91.99**	**30**
CNN	**96.3**	**92.29**	**25**
CNN(Static and Dynamic)	**98.32**	**96.84**	**Combined**

Table 7: Accuracies obtained from tuning the models

From all the models performed after training with various parameters, we can see that the CNN model that was trained differently on both Static and dynamic activity separately gave the highest accuracy compared to the CNN model with 6 classes and their LSTM model. The final pipeline CNN model combined with Static and dynamic activities gave 98. 32 % train accuracy and can predict the human activities with 96.84% accuracy. All these models had undergone hyper parameter tuning with different combination of parameters such as optimizers, dropouts, regularizers, etc. in order to see the best combination of parameters that give the highest accuracy.

7 Discussion

During the exploration of the data, we have observed that the mean acceleration is less than -0.8 during starting activities and greater than -0.6 during dynamic activities. Also, the angular velocity is counted through mean Gravity where the angle between mean gravity in X component indicates that the laying activity is greater than 0 and all other activities are less than 0. The angle between y-axis and mean gravity is less than -0.2 during laying activity and other activities are greater than -0.2.

We have also performed t-SNE which is t-distributed Stochastic Neighbour Embedding to separate the data belonging to 6 activities properly using different perplexity values. On the train data, we have performed with perplexity values of 2,5,10, 20 and 50.

With the highest value of perplexity to 50, we can see that the data are nicely separable except the standing and sitting activities which overlap each other. This indicates that the features related to standing and sitting activities are quite the same that cannot be separated with the help of t-SNE. The dynamic activities are well separable after applying high perplexity value.

On the test data, we have applied t-SNE of 20, 50 and 90 and observed that static and dynamic activities are distinct from each other. In the test data also, we have seen that the laying activity is distinct from all other activities and standing and sitting activities overlap each other.
We have also evaluated the model with the help of matrix such as F1-score, precision and recall to determine the misclassification given by the model with the help of regularizers and without regularizers. We have applied 30 epochs on each model where we have seen that the CNN model gave the highest score compared to all other LSTM with different layers and CNN models. This indicates that CNN is suitable to classify human activities based on 6 classes.

Throughout the research, we had also performed hyper parameter tuning where we have picked the best parameters from all the models to give the best accuracy. From the model we have seen that the training accuracy goes to 94.5% and the test accuracy goes to around 92% with 30 epochs. The LSTM model gives 96.3 percent train accuracy and test accuracy is 92.3 % with 25 epochs. The combined divide and conquer based model where data sharpening is applied to combined the static and dynamic activities and fed into a final pipeline with different number of epochs give a training accuracy of 98.3% and test accuracy of 96.8%. These indicate that the combined approach can be used in a better way compared to other deep learning methods. Also, the higher accuracy indicates that this can be also improved in future research where we can see that Static and dynamic activities are highly predicted with the help of divide and conquer based approaches.

8 Conclusion and Future Work

Human activity recognition system can be useful in various Healthcare domain in order to determine the psychological state of human being during different activities. The human behavior and different disease can be detected with the help of acceleration recorded during daily human activities. For instance, the system is set to detect that a person is seated in a wheelchair at a given time of day, and if any change is detected, such as the person falling from the wheelchair, the system is activated and an alert is generated.

Human activity recognition is performed by previous researchers based on machine learning and deep learning approaches and various accuracies were observed that can be applied in real life applications. We have done a critical review to study all those approaches and then we observed the accuracies obtained by them. We saw that most of them performed deep learning models like LSTM, CNN and TCN model. Moreover, different machine learning models were also performed in same human activity recognition data and they achieved different accuracy from it. In our research, we have performed deep learning models like single and multi-layered LSTM, CNN and divide-and-conquer based approaches.

The dynamic activities were detected with greater accuracy where we separated the static and dynamic activities and perform an individual Classifier on them. The only problem encountered in these models were that LSTM and CNN gave very low accuracy in detecting sitting and standing activities. This problem is solved by using divide and conquers based approach where data sharpening is used to combine the static and dynamic activities and we could see that the sitting and standing activities are also detected with more than 90% accuracy with the help of divide and conquer based approach.

The final model we obtained from picking the best parameters includes LSTM, CNN, and a combined CNN model with a different number of epochs that gave the highest accuracy compared to performing individual LSTM and CNN model. This model required a high amount of time during training. Thus, it was fed into a final pipeline to combine both separate CNN models classifying Static and dynamic activities.

The future research work includes building noble architecture on the same data to see if the accuracy improves or not. The data should be recorded from more participants in order to see the influence on the distribution of the data during different activities. The model can also be deployed in cloud-based services that can be easily accessible to the public in detecting human activities and also it should be updated time to time to mitigate the issues encountered during executing the model and increase the accuracy after adding the data.

In this research we have achieved a better accuracy with the help of divide and conquer based approach and we can also achieve more accuracy if we can add more number of features in the data from different participants. What we did not achieve is an open-end framework of the data that can be easily accessible to the public for future use. We also did not achieve to any model that can give the highest accuracy compared to our proposed architecture. This leaves us a chance to perform any future research on the similar data and topic to perform any other architecture that can give better performance compared to our present algorithm.

References

Abbasi, M.U., Rashad, A., Basalamah, A. and Tariq, M. (2019). Detection of Epilepsy Seizures in Neo-Natal EEG Using LSTM Architecture. *IEEE Access*, 7, pp.179074–179085.

Bhatt, A. and Bhatt, A. (2019). EEG based Emotion Recognition using SVM and LibSVM. *International Journal of Computer Applications*, 178(45), pp.1–3.

Chaudhary, N. and Roy Chowdhury, D. (2019). Data Preprocessing for Evaluation of

Recommendation Models in E-Commerce. *Data*, 4(1), p.23.

Chen, Y. and Wu, W. (2017). Application of one-class support vector machine to quickly identify multivariate anomalies from geochemical exploration data. *Geochemistry: Exploration, Environment, Analysis*, 17(3), pp.231–238.

Cheng, X., Huang, B. and Zong, J. (2021). Device-Free Human Activity Recognition Based on GMM-HMM Using Channel State Information. *IEEE Access*, 9, pp.76592–76601.

Halvaei Niasar, A. and Rahimi Khoei, H. (2015). Sensorless Direct Power Control of Induction Motor Drive Using Artificial Neural Network. Advances in Artificial Neural Systems, 2015, pp.1–9.
Heeryon, C. and Yun, S.M. (2018). Divide and Conquer-Based 1D CNN Human Activity Recognition Using Test Data Sharpening. *Sensors*, 18(4), p.1055.

Kamundala, E.K. and Kim, C.H. (2018). CNN Model to Classify Malware Using Image Feature. *KIISE Transactions on Computing Practices*, 24(5), pp.256–261.

M Bani Amer, M. (2021). Smartphone and Smartwatch for Human Activity Recognition. *Annals of Advanced Biomedical Sciences*, 4(1).

Mittal, V., Umrania, D., Bansal, P. and Ratnani, P. (2021). Modelling and Analysis of Various Machine Learning Algorithms for Human Activity Recognition. *SSRN Electronic Journal*.

Nguyen, B., Coelho, Y., Bastos, T. and Krishnan, S. (2021). Trends in human activity recognition with focus on machine learning and power requirements. *Machine Learning with Applications*, 5, p.100072.

R.Banupriya and Karthik, R.S. (2018). An Automatic Big Data Visualization Framework to Plot Chart Using T-SNE Algorithm. *International Journal of Computer Sciences and Engineering*, 6(11), pp.591–596.

Atallah, L., Lo, B., King, R. and Yang, G.-Z. (2011). Sensor Positioning for Activity Recognition Using Wearable Accelerometers. *IEEE Transactions on Biomedical Circuits and Systems*, 5(4), pp.320–329.

Hassan, M.M., Uddin, Md.Z., Mohamed, A. and Almogren, A. (2018). A robust human activity recognition system using smartphone sensors and deep learning. *Future Generation Computer Systems*, [online] 81, pp.307–313. Available at: https://www.sciencedirect.com/science/article/pii/S0167739X17317351 [Accessed 19 Feb. 2020].

Wang, Z., Wu, D., Gravina, R., Fortino, G., Jiang, Y. and Tang, K. (2017). Kernel fusion based extreme learning machine for cross-location activity recognition. *Information Fusion*, 37, pp.1–9.

Sharma, N. and Gaur, B. (2016). An approach for efficient intrusion detection for KDD dataset: A survey. *International Journal of Advanced Technology and Engineering Exploration*, 3(18), pp.72–76.

Suto, J. (2021). The effect of hyperparameter search on artificial neural network in human activity recognition. *Open Computer Science*, 11(1), pp.411–422.

Triwiyanto, T., Pawana, I.P.A. and Purnomo, M.H. (2020). An Improved Performance of Deep Learning Based on Convolution Neural Network to Classify the Hand Motion by Evaluating Hyper Parameter. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(7), pp.1678–1688.

Zheng, G. (2021). A Novel Attention-based Convolution Neural Network for Human Activity Recognition. *IEEE Sensors Journal*, pp.1–1.

Ziaeefard, M. and Bergevin, R. (2015). Semantic human activity recognition: A literature review. *Pattern Recognition*, 48(8), pp.2329–2345.