Behavioural Context Recognition and Analysis from Mobile Sensors

Activity and behaviour detection with data obtained from mobile and wearable device sensors

Abhijit S Iyer
Department of Computer Science and Engineering
IIT Jodhpur
Jodhpur, Rajasthan, India
M21CS001

Shruti Sureshan
Department of Computer Science and Engineering
IIT Jodhpur
Jodhpur, Rajasthan, India
M21CS015

***Abstract*—This document is the final report of the project that we undertook for the course 'Mobile and Pervasive Computing', which was based on human behaviour detection and analysis from data was available in the 'Extrasensory' dataset. Augmented data was taken from the extrasensory dataset and was used to experiment with and train ML models to detect the activity that a person was doing. The models were able to detect whether a person was sitting down, lying, running, standing, etc.**

I. Introduction

Data is the new gold. It is available everywhere around us in all sorts of forms, waiting to be exploited and made sense of. To serve our cause, we made use of sensor data available from smartphones and other wearable devices, which has the potential to be used to detect a human being individually if they can be used well. Readings from sensors such as smartphone accelerometer, watch accelerometer, gyroscope, etc. produce different tuples of readings results for different human beings, thereby allowing us to research to see if those readings can be exploited to find out unique behaviour patters in human beings. To facilitate our project, we made use of the 'Extrasensory' dataset, which is a dataset that was obtained by collecting smartphone and smartwatch sensor data available from around 60 iPhone users, over a period of 7 days by collecting data at every single minute. The collected data was then labelled manually based on the timestamps, i.e. based on memory whether the person was doing a specific activity at that time. This dataset was then made use of to perform data analysis by training ML models to recognize activities that a person is doing, since the data was labelled. Results showed that with available data, activity recognition was possible with extremely high accuracies but behaviour detection needed better data with labelling and such data was not available for training purposes.

II. Related Work

In the past, several studies have revealed that sensor processing for context recognition has improved significantly. Much work has been published to date on context recognition. Context recognition is concerned with inferring a person's physical state, environment and current activity at any given time. Traditional context recognition techniques often use handcrafted features from heuristic processes from single sensing modality [1]. In the last few decades, there has been a tremendous change in the way data is stored, perceived and processed. As enourmous amount of data is generated every second and if this data is processed well, it can offer highly valuable insights. In order to analyse massive amounts of data, a variety of data mining techniques have emerged. Thus, many attempts have been made to properly detect the user's context using data mining and machine learning techniques. We strongly believe that when analyzing data and suggesting methods, researchers should consider models that are appropriate for working in-the-wild [3]. In the literature [5], authors have provided a systematic analysis using several machine learning algorithms for activity recognition. Literature [6] presents the classification performance summary for recognizing body activities with wearable sensors using three different machine learning algorithms: SVM, HMM and ANN. In literature [2], the machine learning algorithm, KNN is adopted. It recognizes six types of human actions, including walking, walking upstairs, walking downstairs,sitting, standing etc. and achieved an accuracy of 96.7% [2].

Literature [3] focuses on building a unified Model for Multi-Modal Sensors and Multi-Label Classification. It uses a Multilayer Perceptron for analyzing the advantages of the model's hidden layers, which are shared among all sensors and all labels, and provide insight to the behavioral patterns that these hidden layers may capture [3].

III. Challenges identified

1) In real life, natural behavior is complex, multifaceted, and highly variable. Context recognition is difficult due to this high variability.
2) The dataset also indicates a practical challenge that is unbalanced and missing sensor data. The smartwatch was not always worn by the user; users sometimes switched off the location services to conserve battery, and audio was not available during a phone call to preserve privacy.

3) To perform behaviour detection, a dataset where each user's data would be tagged with their device sensor readings for a larger number of users was needed. However, such data was not publicly available and needed to be generated, which also included manual annotation and labelling of data. This was a task that is highly prone to human errors.

IV. Proposed System Methodology

A. Proposed solution

In this project, we analysed data obtained from various sensors present in smartphones and smart watches,and found how a unique combination of these data points can be used to identify a user uniquely. The workflow of the system is as shown in Fig. 1 which illustrates the process of taking the input, preprocessing, feature extraction and classification using various Machine Learning algorithms.

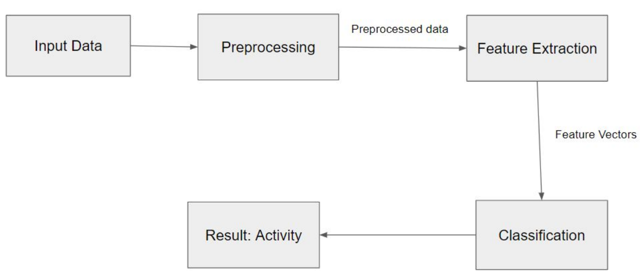

Fig. 1. Block diagram

B. Dataset

The extrasensory dataset contains sensory data and labels collected from over 60 smartphone users in a span of 7 days. Sensor measurements are included in each case. Every minute, the sensor data were automatically recorded for a 20-second frame. For each user, there are around a thousand data instances, each of which comprises data obtained from various sensors. The information was gathered from people who were going about their daily routines. There are 51 different context-label columns, and an example-label pair's value is either 1 (the label is appropriate for the example), 0 (the label is not relevant), or 'NaN' (missing information). The class labels are Physical Activities (e.g., walking, running), Daily Activities (e.g., sleeping) , Locations (e.g., school, work, home) etc. The dataset includes following sensor data:

- Phone Accelerometer
- Gyroscope
- Magnetometer
- Watch Accelerometer
- Watch Compass
- Location
- Audio
- Phone State

The plot for the number of examples per label is as follows:

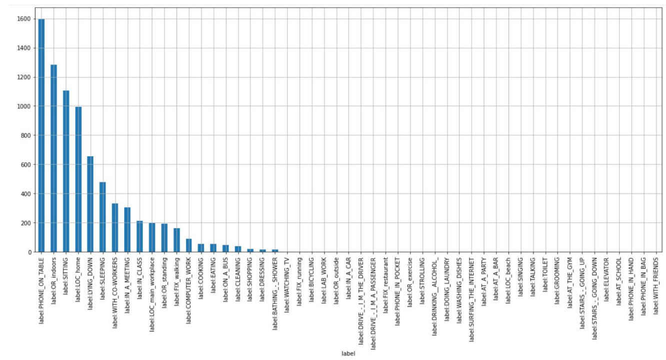

Fig. 2. No of examples per label

C. Preprocessing

We take the unstructured, raw data and transform it into a format that can be understood and analyzed. Data cleaning is the most important step here as it will correct all of the inconsistent data. To handle the missing values, we can ignore the tuples or fill it with the mean of that particular attribute. To handle the noisy data, we perform clustering so as to detect the outliers in the dataset as it falls outside the particular cluster. Also, before applying the machine learning algorithms, we standarize the features, for this we subtract the mean of training samples and scale to unit variance. Many machine learning techniques require dataset standardisation because they may perform poorly if individual characteristics do not more or less resemble standard normally distributed data.

D. Algorithms

1) Logistic Regression: Under the Supervised Learning approach, one of the most prominent Machine Learning algorithms is logistic regression. The easiest way to think of logistic regression is as a type of linear regression that is used to solve classification problems. It's a method for predicting a categorical dependent variable from a set of independent factors. A logistic function (sigmoid) is used to model a binary output variable in logistic regression. This sigmoid function is a mathematical function for converting predicted values into probabilities. Thus, it's a statistical analysis technique for predicting a binary outcome based on the previous data set observations. The main distinction between linear and logistic regression is that the range of logistic regression is limited to 0 and 1. Furthermore, logistic regression does not require a linear connection between input and output variables, unlike linear regression.

2) K-Nearest Neighbours: The K-Nearest Neighbour method is based on the Supervised Learning approach and is one of the most basic Machine Learning algorithms. It is a non-parametric algorithm. It assumes that the new data and the existing cases are comparable and places the new case in the category that is most similar to the existing categories. Because it delivers very precise predictions, the KNN algorithm can compete with the most accurate models. But if our dataset

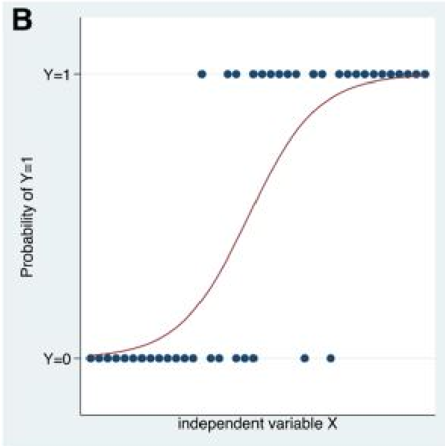

Fig. 3. Logistic Regression [7]

necessitates a large value of k, this will raise the algorithm's computing cost. The steps are as follows:

(a) Choose the value of k i.e. the nearest data points.
(b) Determine the Euclidean distance between K neighbours.
(c) Using the obtained Euclidean distance, find the K closest neighbours.
(d) Count the number of data points in each category among these k neighbours.
(e) Assign the new data points to the category with the greatest number of neighbours.
(f) End

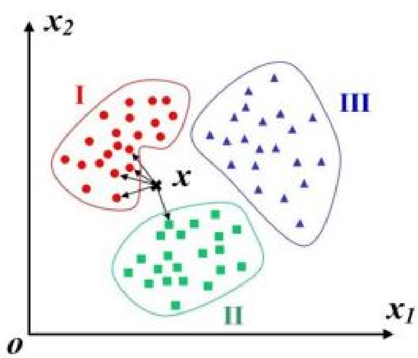

Fig. 4. Illustration of KNN algorithm (k = 3) [2].

3) Random Forest: Random Forests are a supervised tree based machine learning algorithm where we make use of decision trees to make splits at every depth. However, in Random Forests, such a split takes place at random at every depth i.e. the set of features based on which splitting happens is something that takes place out of the blue. Hence, the term 'random' in the algorithm name. As the name suggests, it is a forest, i.e. it consists of a collection of decision trees where feature splitting takes place at random. There is a direct relationship between the number of trees and the accuracy of the algorithm - the more the number of trees, the better will be the algorithm prediction. Having a large number of trees ensures that there is no overfitting that takes place in the model. Random forests can also be used for regression tasks but in our case, we have used Random Forests to try and classify user activitiy, i.e. whether the user is walking or not, etc.

4) K-means Clustering: The K Means clustering algorithm is an unsupervised machine learning algorithm that can be applied on an unlabelled dataset to cluster the various data points into different clusters, each of which may contain similarities since the data points in a cluster are close to each other and may have similar features or behaviour patterns. A hyperparamter to the training algorithm is the number of clusters 'k', that decides the number of clusters into which our datapoints will be classified into. This machine learning algorithm allows us to group datapoints into various dissimilar groups, allowing us to find out the different categories that the datapoints belong to without having labels in our dataset.

Algorithm 1 K-Means Clustering

```
Specify the number k of clusters to assign
Randomly initialize k centroids
repeat
    Assign each point to its closest centroid
    Compute the new centroid of each cluster
until The centroid positions do not change
```

5) Gaussian Mixture Model: The Gaussian Mixture Model is another machine learning model that comes under the category of unsupervised machine learning algorithms which allows us to train the model without having labels in our dataset.We assume that each cluster follows a Gaussian Distribution, or in other terms, we assue that each cluster comes from a different Gaussian distribution. All in all, the dataset is modelled to be coming from several Gaussian distributions. The parameters of the Gaussian distribution are however deemed to be unknown. This model is similar to the K means clustering algorithm, which clusters the given datapoints into the specified number of clusters based on their similarity and dissimilarity. GMMs however are at an advantage over K means in the fact that GMM can be used for those datapoints where clusters cannot be clearly defined.

6) Dimensionality reduction using Principal Component Analysis: Principal Component Analysis or PCA, is a method of reducing the dimensinality of the input dataset (or) remove attribtues from a given dataset which are highly correlated to another attribute, so that the model will be able to function better. In spite of the reduced dimensionality, we ensure that no crucial information is lost in this process of dimensionality reduction. The whole point of PCA is to reduce dimensionality to trade accuracy for simplicity a bit. In our case, the original dataset had data collected from 255 different sensors, and many of which were correlated. Hence, for the sake of simplicity, we used PCA to reduce the dimensionality. Another advantage of PCA is data visualization. With reduction in the number of dimenstions, data can be visualized easily.

7) Artificial Neural Network: The concept of Aritificial Neural Networks comes from the way our brain works. The brain contains several neurons, that are connected to each other with the help of synapses, through which information flows to produce results. Artificial Neural Networks also work on a similar philosophy, where we have artificial neurons that are connected by connectors, each of which have certain weights which decide how the information will flow. In our case, we have modelled the Artificial Neural Network to be a binary classifier, to classify between activities such as walking, sitting and lying.

8) Multilayer Perceptron: The Multilayer Perceptron Model is a feed forward neural network. They can be used for classification, regresion, recognition, prediction and approximation. In our case, we have used multilayer perceptron model to perform multiclass classification among various labels such as walking, sitting, standing, lying, etc. The mapping between inputs and outputs in a multilayer perceptron is non linear.

E. Performance Evaluation Parameters

The performance of the system was evaluated using the following performance evaluation parameters:

- Accuracy: It is defined as the ratio of the number of correct predictions to the total number of input samples. This is shown in Equation 1

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total} \tag{1}$$

- Precision: It's defined as the number of correct positive outcomes divided by the classifier's predicted number of positive results. This is shown in Equation 2

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{2}$$

- Recall: It is calculated by dividing the number of correct positive results by the total number of relevant samples. This is shown in Equation 3

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{3}$$

- F1 Score: It's the harmonic mean of recall and precision. It tells you the robustness of your classifier. This is shown in Equation 4

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

V. Results

A. Activity Detection Based on Accelerometers

In this experiment, we have predicted the label of whether the user using the smartphone is walking, based on data collected from smartphone accelerometers and smartwatch accelerometers only. Three different classification algorithms were used to solve this purpose. The results are as follows:

Sensor	Classifier	Accuracy
Acc, Wacc	Logistic Reg	0.947
Acc, Wacc	KNN	0.954
Acc, Wacc	Random Forest	0.959

B. Activity Detection based on Accelerometers, Gyroscope and Location Sensor

A similar experiment was conducted as the previous experiment, but the gyroscope sensor data and location data was included to predict if a user was walking. The results are as shown in the figure below:

Sensor	Classifier	Accuracy
Acc, Wacc, Gyr, Loc	Logistic Reg	0.946
Acc, Wacc, Gyr, Loc	KNN	0.946
Acc, Wacc, Gyr, Loc	Random Forest	0.950

C. Activity Detection based on Location Data and Audio Data

This is again another similar experiment that was performed, using location sensory data and audio sensory data. Data from these points were used to predict if a person is walking or not, according to the labels assigned in the Extrasensory dataset. In this case too, the experiment was performed on three different classifiers.

Sensor	Classifier	Accuracy
Loc, Audio	Logistic Reg	0.928
Loc, Audio	KNN	0.947
Loc, Audio	Random Forest	0.9475

D. Clustering all instances, based on all features available in dataset

In this experiment, we applied K-Means clustering and GMM on the dataset, without filtering out on any feature. The silhouette scores for different number of clusters were observed as follows:

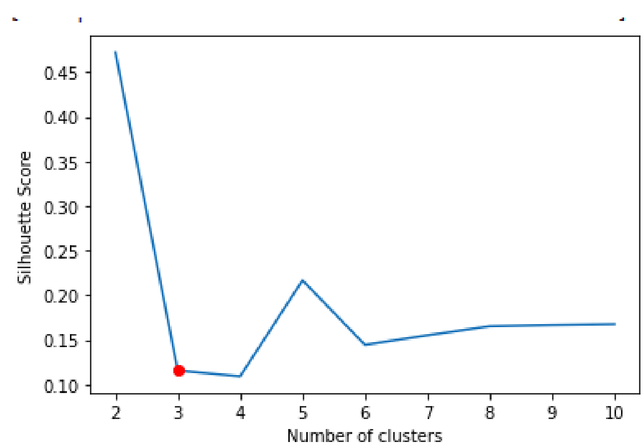

Fig. 5. Silhouette Scores for each k

E. Clustering all instances after applying PCA

Since the number of features are quite high, we used dimensionality reduction to remove highly correlated features to obtain better clarity on the data. Post dimensionality reduction using PCA, we applied clustering mechanisms to check on silhouette score.

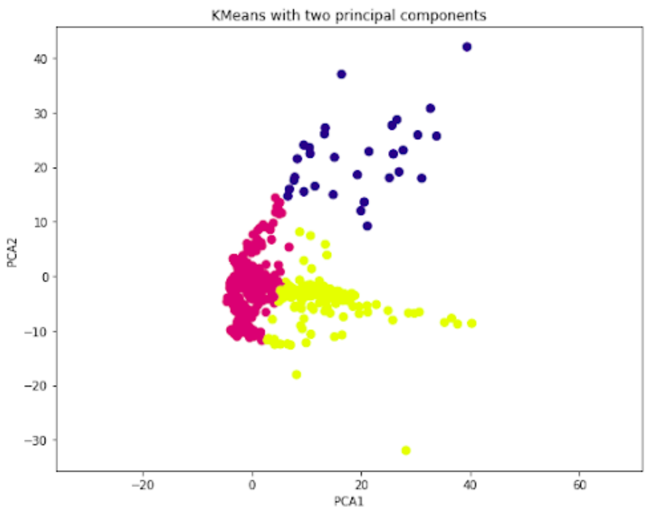

Fig. 6. Three clusters formed

F. Clustering Based on Accelerometer, Gyroscope, Location and Audio Applying PCA

Instead of choosing all features available in the input dataset, the clustering experiment was carried out only on the above mentioned features of data points. The silhouette score graphs for different values of number of clusters can be seen as follows: (Fig 7 and Fig 8)

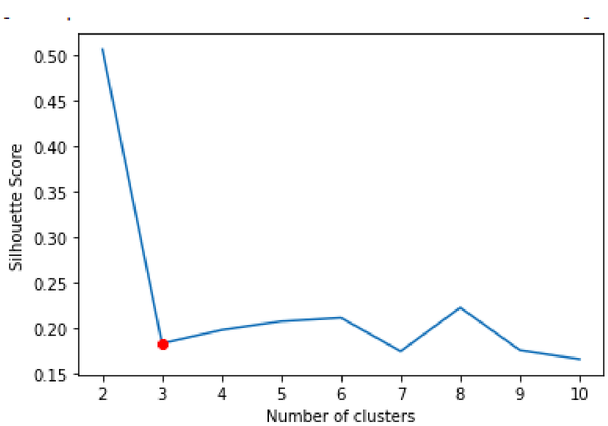

Fig. 7. Number of clusters

G. Clustering Based on Accelerometer and Gyroscope

In another experiment, clustering was done only on accelerometer data and gyroscope data. This was done to see if there were significant changes in results when in comparison to all sensor data. The silhouette score chart can be shown as follows: (Fig 9 and Fig 10)

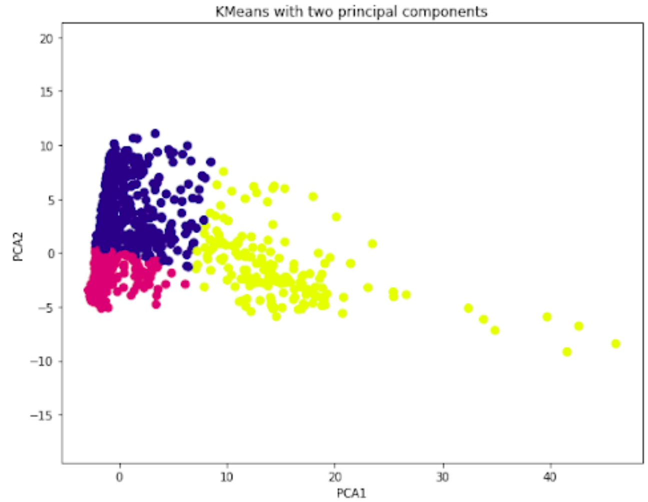

Fig. 8. Output for three clusters

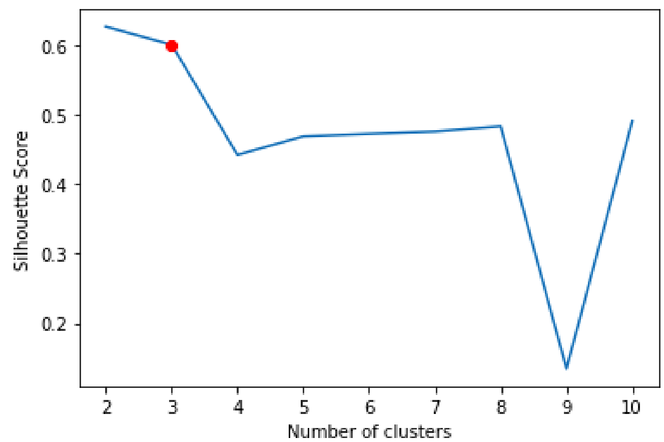

Fig. 9. Silhouette Scores

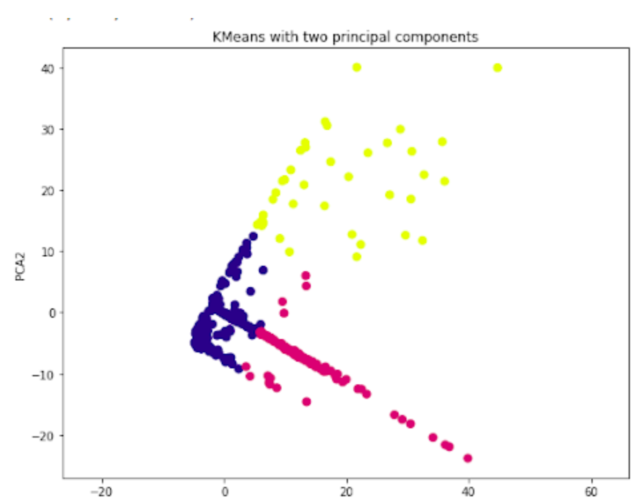

Fig. 10. Output for three clusters

H. Multiclass Classification using Multilayer Perceptron

Multi class Classification of user activity was performed based on the the data collected from the mentioned sensors. The results can be seen as follows:

Wacc, Acc, Gyro, Loc:

Activity	Prediction Score
Walking	0.9606
Lying Down	0.9403
Sitting	0.8573
Standing	0.9330
Sleeping	0.9403

Loc, Audio:

Activity	Prediction Score
Walking	0.9359
Lying Down	0.9606
Sitting	0.8864
Standing	0.9112
Sleeping	0.9519

I. Binary Classification Using ANN

Binary Classification of user activity of 'walking' was performed based on the the data collected from the mentioned sensors. The results can be seen as follows: (Fig 11 Fig 12)

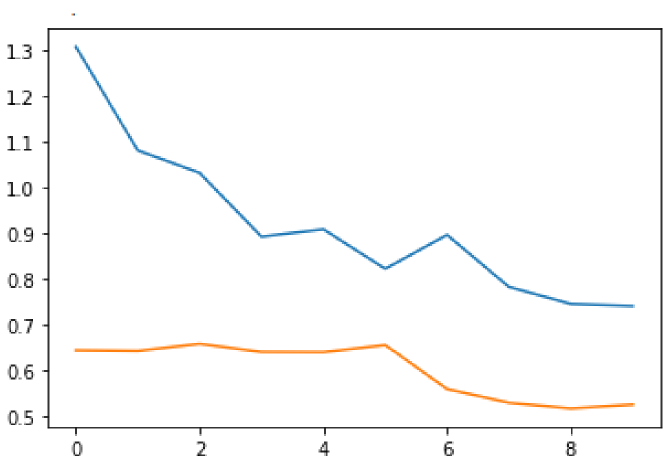

Fig. 11. Loss Graph - Prediction Accuracy: 0.9330

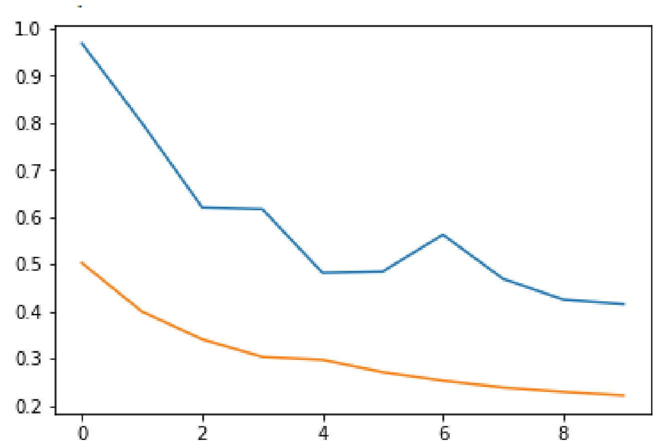

Fig. 12. Loss Graph - Prediction Accuracy: 0.9304

VI. Conclusion and Future Work

The extrasensory dataset provides enough data for us to classify amongst the various activities that a human being is doing. However, to recognize a human being individually, the problem lies in the fact that we do not have annotated data that can be used to train models to uniquely detect a human being. Hence, a major step needs to be taken in the direction where we can publicly can make annotated data available for human behaviour detection and unique identification.

Creating new strategies to manage the unbalanced multi-label data and efficient selection of sensor to decrease computation, as well as incorporating additional equivalent sensors to increase the detection rate even more, are interesting future topics for research.

Acknowledgment

Behavioural Context recognition is one of the fundamental task under activity recognition with broad applications in various fields. We would like to thank Dr. Suchetana Chakraborty for continuous guidance and providing us the opportunity to explore this area.

References

[1] I. K. Ihianle, A. O. Nwajana, S. H. Ebenuwa, R. I. Otuka, K. Owa and M. O. Orisatoki, "A Deep Learning Approach for Human Activities Recognition From Multimodal Sensing Devices," in IEEE Access, vol. 8, pp. 179028-179038, 2020, doi: 10.1109/ACCESS.2020.3027979.

[2] P. Wang, Y. Zhang and W. Jiang, "Application of K-Nearest Neighbor (KNN) Algorithm for Human Action Recognition," 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2021, pp. 492-496, doi: 10.1109/IMCEC51613.2021.9482165.

[3] Vaizman, Y., Weibel, N., Lanckriet, G. (2018). Context recognition in-the-wild: Unified model for multi-modal sensors and multi-label classification. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1(4), 1-22.

[4] A. Saeed, T. Ozcelebi, S. Trajanovski and J. J. Lukkien, "End-to-End Multi-Modal Behavioral Context Recognition in a Real-Life Setting," 2019 22th International Conference on Information Fusion (FUSION), 2019, pp. 1-8.

[5] Zheng, Y. (2015). Human activity recognition based on the hierarchical feature selection and classification framework. Journal of Electrical and Computer Engineering, 2015.

[6] L. Cheng, Y. Guan, Kecheng Zhu and Yiyang Li, "Recognition of human activities using machine learning methods with wearable sensors," 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), 2017, pp. 1-7, doi: 10.1109/CCWC.2017.7868369.

[7] Schober, Patrick MD, PhD, MMedStat*; Vetter, Thomas R. MD, MPH† Logistic Regression in Medical Research, Anesthesia Analgesia: February 2021 - Volume 132 - Issue 2 - p 365-366 doi: 10.1213/ANE.0000000000005247