Activity Recognition using Empatica E4 Wristworn Device

Akhil Menon
Computer Information and Science Department
Temple University
Tug81428@temple.edu

Abstract – This paper is the initial phase of a larger project framework which aims to combat smoking behavior of users with the help of IoT based technologies. The purpose of this paper is to document the process of human activity recognition (HAR) using a wristworn device (Empatica E4) and to test the efficiency of different machine learning algorithms with the training data set. The HAR system includes collection of raw data, labeling the data with corresponding activities, extracting & selecting features and finally running the data through machine learning algorithms that classifies the activities. The algorithms evaluated in this paper are Support Vector Machine, Logistic Regression, K-NN, Random Forest and Naïve Bayes. The results proved that SVM gives the maximum accuracy of 89%, with a precision of 90%, followed by Random Forest with 88% accuracy and precision of 87%. This paper contributes the following:- 1. Features used in the process of activity recognition 2. Parameters used in the algorithms 3. Output 4. Performance metrics

I. INTRODUCTION

Recent developments in the field of IoT based technologies that employ the use of sensor such as proximity, location, accelerometer, gyroscope, IMU sensors have enabled and encouraged researchers to analyze as well as contribute to various fields such as Healthcare, Security, Psychology/social behavior monitoring, human computer interaction and many other fields. Human Activity Recognition systems using wearable devices with embedded sensors help analyse and recognise different activities and gestures. Automatic activity recognition is a data driven process that uses a machine learning techniques to process sensor data.

Smoking is one of the primary causes of deaths in the world, however, it is one such cause that can be prevented. According to Centers for Disease Control and Prevention, tobacco smoking is responsible for over 480,000 deaths per year in the United States alone and nearly 6 million, worldwide. In addition, illnesses related to and caused by smoking in United States costs $170 billion in direct medical care and $156 billion in lost productivity in 2017 [1]. Individuals who smoke regularly mentioned that they initiated this habit due to this belief that smoking reduces stress, anxiety and gives instant relief. Many have also claimed to have started smoking due to social stigma, which includes perception of seeming fashionable, peer pressure and social acceptance. Research shows that 68% adults attempt to quit, however, the addiction to nicotine makes it challenging and the withdrawal symptoms tends to pull the individual back on the cycle of smoking. The approach to combat the addiction in turn, the smoking behavior, in this project is three folded :-

1. Recognize the individuals daily pattern of smoking
2. Replace the smoking activity with a activity that is considered positive to the user such as meditation and/or jogging.
3. Manage withdrawal symptom and emotions

In this paper, a wearable wristband device is used to collect data and with the help of a self learning program, recognize common activities of individuals in daily life. Thereafter, the model can be used specifically to recognize the daily smoking patterns of users, non-intrusively. HAR model is generally a pipeline, that consists of several tasks. Initially, embedded sensors collect raw data from users that is cleaned and preprocessed. Cleaning data refers to converting raw data to a specific and common format, removing/filling empty values, followed by discretizing and getting a uniform data that is called training data. This training data is passed through a machine learning module. Machine learning program basically refers to a self learning program that derives features from classified data, recognises patterns and thereafter, predicts the output for future/test data.

II. MATERIAL AND METHOD

A. Data collection

A wristworn device called Emaptica E4 was used to collect data. The device encompassed Bluetooth, Flash Memory, PPG (Photoplethysmography sensors) from which Heart Rate is derived, EDA (Electrodermal Activity) sensor, Temperature sensor and Tri-axial accelerometer sensor. The device dimensions are 44mm * 40mm * 16mm and weighed 25g. EDA, Heart-rate and Accelerometer sensor data were sampled at 4Hz, 1Hz and 32Hz respectively. The device consisted of Lithium battery, 3.7V, 260mAh with battery life of nearly 36hours in record mode and 20 hours in streaming mode. [2] The device was attached to the author on 29th and 30th of April. A list of activities were conducted, each activity was repeated for 15 times and data was retrieved/downloaded from the Empatica E4 web portal which is called Empatica Manager. The data is organised in a csv folders with 1 sensor data per folder. Each data folder consisted of UNIX-formatted timestamp on the 1st cell and the sampling rate mentioned on the 2nd cell of the first column. Serial number was added starting from 0 corresponding to first timestamp value and UNIX time stamp was converted to US Eastern Daylight Time with the formula {(Timestamp/86400)+DATE(1970,1,1)+(-5/24)} [3]. Hence, the timestamp for each individual data value was obtained in the format HH:MM:SS.000.
Activities that were conducted were :- Combing hair, Drinking water, Eating with a spoon, Itching/touching on the cheek, Answering the phone and applying chapstick on the lips. These activities were repeated 15 times with a 5 sec pause between each repetition and was recorded on

video for precise and easy manner of labeling.

B. Preprocessing

In this paper, accelerometer heart-rate and EDA sensor data was used for HAR. Data was labelled by referencing the video. Thereafter, the data was discretised into window of 2s, no overlap. Magnitude of the acceleration was calculated in order to derive a set of feature characteristics from the window size.

$$\overline{A} = \sqrt{A_x^2 + A_y^2 + A_z^2}$$

This represents acceleration signals in the x,y,z axes respectively.

Features :

Features extracted include : Absolute signal area, Spectral Entropy, Minimum value, max value, Mean, Standard deviation, Dominant frequency, Kurtosis and Signal. These features are used in the classification algorithms mebtioned in the further sections.

C. Classification Algorithms

In this paper, several algorithms are tested and implemented. Table 1 shows the list of algorithms evaluated along with the parameters used. While there are several effective machine learning tools to test the data and create a model, Orange software was used in this paper. Orange describes each algorithms distinctively and has pre-set parameter values for each algorithm that are equivalent to the standard parameter values mentioned in the Scikit-learn library [4]. However, in this paper, different combination of the parameters values were tested, according the training set data, to acquire efficient results.

Table 1

Algorithms	**Parameters used**
Random Forest	Number of trees : 10, Depth of tree : 3, Limit splitting of subset :3
KNN	Number of neighbors : 3, Metric : Euclidean,
Logistic Regression	Regularization type : Lasso(L1), C = 1
SVM	C=1, Regression loss epsilon = 0.10, Tolerance = 0.001,Kernel = RBF *(exp(-g|x-y|^2))*, where g = 1

D. Output

Figure 1, 2 and 3 shows Heart-rate, EDA and magnitude of acceleration against time respectively. The following colors on the graph represents the corresponding activites.

Red = Combing
Magenta = Drinking water
Cyan = Itching/touching cheeks
Green = Eating
Blue = Brushing
Black = Answering the phone
Yellow = Applying chaptsick

FIGURE 1: HEART-RATE VS TIME

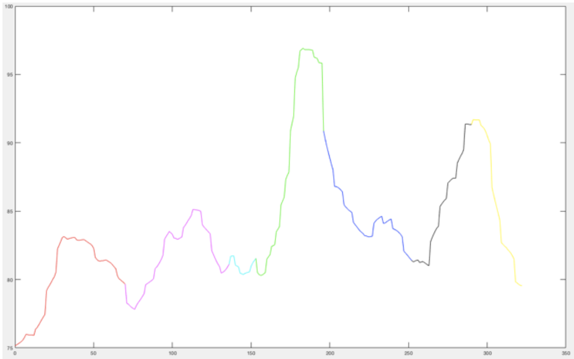


FIGURE 2 : EDA VS TIME

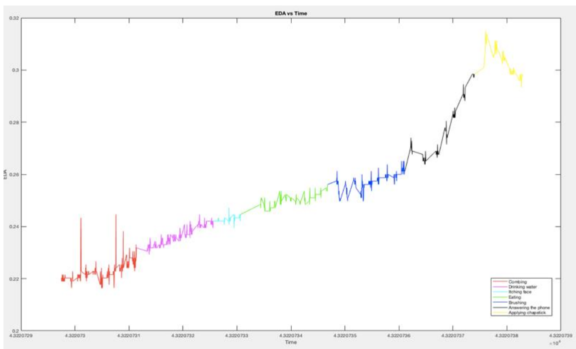
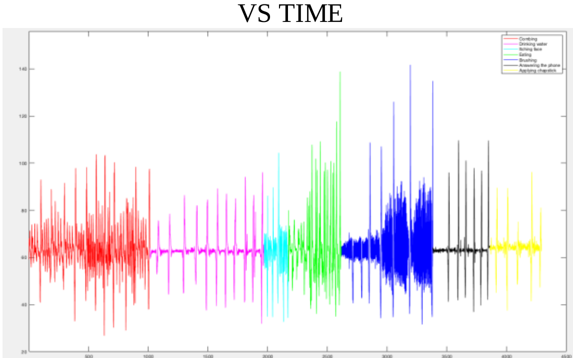


FIGURE 3 : MAGNITUTDE OF ACCELERATION VS TIME



E. Performance of classification

Following metrics were used to evaluate the classification algorithms

$$\text{Classification accuracy} = \frac{TP+TN}{TP+TN+FP+FN'}$$

$$\text{Precision} = \frac{TP}{TP+FP'}$$

$$\text{Recall} = \frac{TP}{TP+FN'}$$

$$\text{F-Score(F1)} = 2* \frac{Precision \times Recall}{Precision + Recall}$$

<u>Output table</u> :

Method	AUC	CA	F1	Precision	Recall
kNN	0.863	0.834	0.827	0.834	0.834
SVM	0.922	0.892	0.887	0.901	0.892
Random Forest	0.946	0.881	0.876	0.877	0.881
Naive Bayes	0.943	0.767	0.791	0.842	0.767
Logistic Regression	0.899	0.815	0.777	0.770	0.815

Figure 4 is the comparison of output from different machine learning models

Accuracy score of SVM is maximum = 89.2% , with precision = 90.1%
Followed by Random Forest = 88.1% with precision = 87.7%

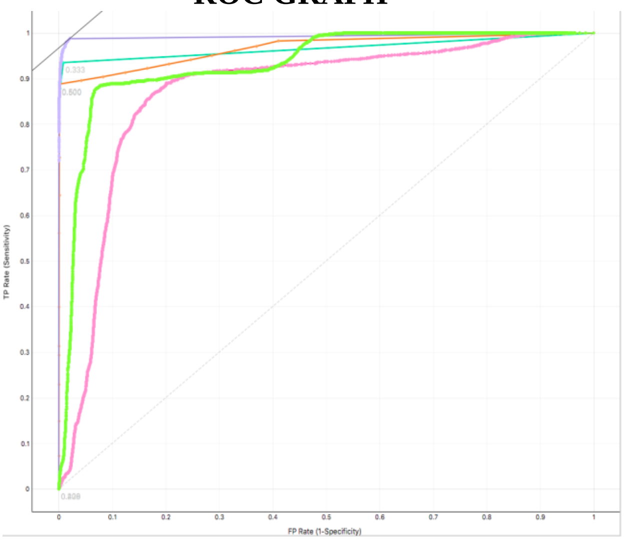
Confusion Matrix

	A	B	C	D	E	F	G	H
A	72	0	1	0	0	0	0	8
B	0	63	0	0	0	0	0	7
C	1	0	102	0	0	0	0	7
D	0	0	0	136	1	0	0	23
E	0	0	0	9	96	0	0	25
F	0	0	1	0	0	12	0	57
G	0	2	0	0	0	0	20	10
H	7	33	32	42	4	0	15	19 17

Confusion Matrix Classes
A = Combing
B = Drinking water
C = Itching/Touching the cheek

D = Eating
E = Brushing
F = Answering the phone
G = Applying chapstick on the lips
H = None

ROC GRAPH



TP RATE VS FP RATE

III. RELATED WORK

Wearable senors : There are several existing systems that focus on hand activity recognition using wearable devices and have a primary goal of recognising the activity smoking. These include StopWatch [5], SmokeBeat [6] and RisQ [7]. RisQ used 9 axis IMU and it gives an accuracy of 95% which is highest amongst the 3 systems. SmokeBeat and StopWatch uses accelerometer as well as gyroscope sensor data from users to provide low cost systems. The accuracy for SmokeBeat and StopWatch are 86% and 80% respectively. All 3 systems require the users, in the test trials, to respond to a notification on the mobile app. In RisQ, users had to manually notify the app of the activity being conducted and SmokeBeat app sent a notification whenever it detected users' activity of smoking, to which the user had to either confirm or deny. The accuracy of the system, this approach, is unpredictable because it solely depends on the users' honesty.

IV. FUTURE WORK

Empatica E4 consists accelerometer, EDA and heart-rate sensors which gives fairly good accuracy scores. In the future, HAR testing with Emaptica Embrace device (consisting of accelerometer, EDA and gyroscope) is recommended and the results from E4 and Embrace should be compared. This is because the related work illustrates accurate results with the gyroscope data, therefore the comparison will depict the importance of Heartrate sensors vs Gyroscope sensor. In order to accurately detect and classify smoking activity, heart-rate sensor data must be considered along with the other data since the change in heart rate is directly related to smoking activity [8]. Secondly, in terms of recognizing the activity of smoking, a precise start and end of activity must be recognised so as to accurately acquire the duration of the smoking activity. In the related articles such as SmokeBeat, the author used a mobile app to record the beginning and end of the smoking activity. This can be developed and modified to have a more nonintrusive recognition. Additionally, another area of focus could be a respiratory sensor that would help in the classification of the activity of smoking from other hand related activity.

V. CONCLUSION

This paper illustrates the use of accelerometer, heart-rate and EDA sensor data to recognise activities. It also depicts that SVM is the most suitable algorithm to use for activity recognition for the respective device and type of data. An accuracy of 89% and precision of 90% was achieved. Different activites/classes were identified efficiently. The paper concludes by stating that activity recognition was successfully achieved and user trials can follow this process to collect data from user environment. In the future, appropriate regulations and approvals for user testing must be obtained and the device should be tested for the actual activity of smoking. Since the heartrate is correlated to smoking, Empatica E4 would prove efficient and accurate in detecting the activity.

VI. REFERENCES

1) "Smoking & Tobacco Use," *Centers for Disease Control and Prevention*, 20-Feb-2018. [Online]. Available: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/fast_facts/index.htm. [Accessed: 07-May-2018].

2) "E4 Get Started Guide - Start acquiring physiological signals". [Online]. Available: https://www.empatica.com/get-started-e4. [Accessed: 07- May-2018].

3) "Session start-time format and synchronization". [Online]. Available : https://support.empatica.com/hc/en-us/articles/202800715-Session-start-time-format-and-synchronization. [Accessed : 07-May-2018]

4) "Scikit-Learn", Internet : http://scikit-learn.org/stable/ [Accessed 07- May-2018]

5) Andrew Skinner, Christopher J Stone, Hazel Doughty, Marcus Munafo. "*StopWatch: The Preliminary Evaluation of a Smartwatch-Based System for Passive Detection of Cigarette Smoking*", Oxford University Press, 24 January 2018. [Online]. Available : https://doi.org/10.1093/ntr/nty008. [Accessed : 07-May-2018]

6) Reuven Dar, "*Effect of Real-Time Monitoring and Notification of Smoking Episodes on Smoking Reduction: A Pilot Study of a Novel Smoking Cessation App*", 06-November-2017, [Online]. Available : https://doi.org/10.1093/ntr/ntx223. [Accessed : 07-May-2018]

7) Abhinav Parate, Meng-Chieh Chiu, Chaniel Chadowitz, Deepak Ganesan, Evangelos Kalogerakis. "*RisQ: Recognizing Smoking Gestures with Inertial Sensors on a Wristband*" in Mobisys, 16-June-2014, [Online]. Available : https://people.cs.umass.edu/~dganesan/papers/MobiSys14-RisQ.pdf. [Accessed : 07-May-2018].

8) Junichi Minami, Toshihiko Ishimitsu, Hiroaki Matsuoka. Effects of Smoking Cessation on Blood Pressure and Heart Rate in Habitual Smokers. 09-Novemnber-1998. [Online]. Available : http://hyper.ahajournals.org/content/33/1/586