Evaluating Machine Learning Techniques on Human Activity Recognition Using Accelerometer Data

Rimsha Khan[1], Muhammad Abbas[2], Rubia Anjum[3], Fatima Waheed[4]
Dept.of Computer and Software Engineering, College of E&ME,
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
rimsha.khan18@ce.ceme.edu.pk[1], m.abbas@ceme.nust.edu.pk[2],
rubia.anjum18@ce.ceme.edu.pk[3], fatima.waheed18@ce.ceme.edu.pk[4]

Sheeraz Ahmed
Dept.of Computer Science
Iqra National University
Peshawar, Pakistan
sheeraz.ahmad@inu.edu.pk

Fahad Bangash
Dept.of Electrical Engg,
Bahria University
Islamabad, Pakistan
engr.bangashh@gmail.com

Abstract— **Human activity recognition is gaining increasing importance because of its implication in remote monitoring application including security, health and fitness apps. This paper provides an analysis of different machine learning techniques for recognizing human activity. Firstly, all the recent work related to human activity recognition using accelerometer data is analyzed and presented in the paper. In this study the accelerometer used in smartphones as well as those embedded in wearable devices are compared and recognition methodologies applied on both the devices are presented. The dataset used in this project is a transformed version of "Activity Recognition using Cell Phone Accelerometers," by the Wireless Sensor Data Mining WSDM. Some important features were extracted from the data and based on it different models were assessed using Matlab Classification Learner App. Four distinct machine learning techniques were applied on the dataset, namely, linear regression, logistic regression, support vector machine and neural network. For the purposed of applying classifier Weka tool is used. The results of these algorithms are compared and presented in the form of tables and graphs and Bagged Tree is identified to be the best algorithm based on accuracy results.**

Keywords—.Human Activity Recognition (HAR), remote monitoring, accelerometer data, machine learning, weka.

I. INTRODUCTION

Human Activity Recognition is an engaging topic of research since quite some time but still has a lot of space for improvement. Various method presented by researchers through which human activity can be recognized are based on computer vision using RGB camera, sensors like Wifi sensors, magnetometers, gyroscope and accelerometer sensors. Recognizing activity through computer vision methods has limitations due to lightening, accurate foreground extraction technique and most importantly the person to be monitored needs to be in camera range of vision. Monitoring human activity of a more mobile person is far easier and pervasive through accelerometer data as it is embedded on a mobile device like smart phone or some other device and is easier to carry around. By using accelerometer data on a smart phone, a person's activity can be stored ubiquitously and sent to a server where it can be processed to recognize activity. Figure 1 shows an overview of the whole process.

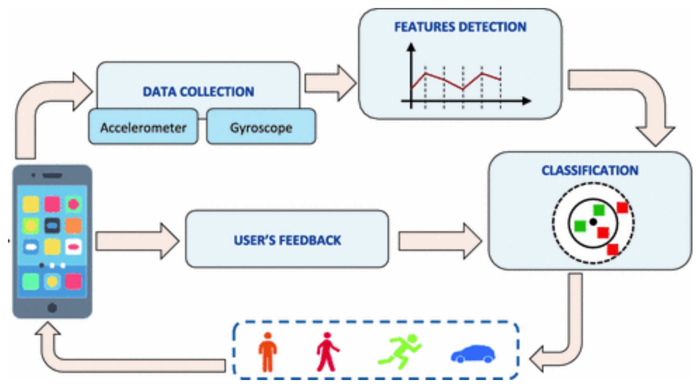


Figure 1 Overview of human activity recognition using data collected from smartphone

HAR is becoming more and more interesting in many application. Assistive technology like eldercare, storing calories lost based on exercise done in fitness app, and monitoring human gait in security application are few of amazing applications of HAR. Major challenge in HAR is that human body activities are very complex and its real-time recognition is very difficult. To overcome this challenge machine learning is applied on sensors data like accelerometer data. In this paper application of different methods of machine learning is analyzed and presented.

Different researchers have explored and presented different ways of recognizing and classifying human activity. The data is processed and classified into predefined learnt classes like walking, climbing, running etc. In any application of machine learning, feature selection is one of the most important step for creating a classifier. The features which are more informative, needs be selected before classification for generalized results. The features selected for classification are mean, standard deviation, peak and resultant values. These features help classify data from the tri-axial accelerometer, x, y, and z values into classes of walking, standing, running, sitting etc.

The rest of the paper is organized as follows. Section 2 presents a literature review of the recent work done on human activity recognition through accelerometer data. First different techniques applied on smartphone accelerometer data are discussed and then the work done recognizing activity from wearable accelerometer data is discussed. Section 3 describes the data set used in experimentation in

Authorized licensed use limited to: University of Prince Edward Island. Downloaded on November 10,2020 at 16:20:24 UTC from IEEE Xplore. Restrictions apply.

detail. Section 4 presents classification done on Matlab Classification Learning App, its analysis and results. The complete methodology of developing classifier, graphical analysis and results is presented in the paper. Section 5 elaborated the methodology of evaluation carried out in the project. Each of the algorithm and its results are discussed in detail. Section 6 presents the comparison of results and the best algorithm is identified based on that comparison in Section 7 which also provides the conclusion derived from this work.

II. LITERATURE REVIEW

Smartphone sensors like accelerometer and gyroscope have paved the way for more and more sensor based applications like Human Activity recognition for the welfare of people. Simple daily life activities are recognized and based on it more complex activities can be identified. In the paper, a complex activity like Muslim prayer is recognized by firstly collecting data, preprocessing the data collected, extracting important features, classifying into classes and finally aggregation [1]. In another paper machine learning technique is introduced which uses ensemble manifold rank preserving (EMRP) algorithm to predict human activity on two datasets, namely, the SCUT NAA dataset and the NMHA [3].

Convolution Neural Networks technique is used for classification and the datasets used were Actitracker dataset along with Skoda and Opportunity datasets [4]. Automatic extraction of discriminating features is described in the paper [5].The data taken is recorded from mobile phone accelerometer which is placed in person's pocket and hand. In the paper [6], different classifier were analyzed and five of the best, high performance classifiers were selected. The results of the proposed model were up to 91.15% accurate. An Algorithm called LHUC (Learning Hidden Unit Contributions) which is based Neural Networks is applied in Smartphone sensor.

Other than mobile phone accelerometers, accelerometers are also sometimes embedded in other wearable devices with the benefit that it is not as sensitive to damage as smart phones and can be used by children and elderly people care. In a paper machine learning technique called the random forest is applied on data recorded from accelerometer sensors embedded in a wearable device [7]. The results of the experimentation were up to 94% accurate. In the paper, it is described how features selection mechanism is applied on statistically derived features [8]. Human activity is recognize using a wearable device which is described in the paper.

Feature selection is a critical part of machine learning methodology. In the paper those features are extracted and presented that helps classify physical activities better [9]. 94% of accuracy is obtained using the Random Forest algorithm. Numerous activities are recognized using a wearable accelerometer and classified into classes of activities and sub-activities successfully. Using K-nearest neighbor classification six activities were correctly classified into the class of the type of activity being performed. In another [10] work deep learning technique is used to for activity recognition. Firstly useful features are extracted automatically and computational cost is decreased. A wearable device is worn on arm which has a tri-axial accelerometer embedded in it which recognized the movement of the arm.

Different techniques of human activity recognition are studied and analyzed in various different works. In a work human activity recognition techniques are discussed which are used in both wearable as well as smartphones [11]. Different techniques for activity recognition, including temporal pattern mining and ANFIS are proposed. The techniques discussed in the paper are evaluate for both wearable and smartphone accelerometer data and gave accurate results.

III. DATA SET

The dataset used in this project is "Activity Recognition using Cell Phone Accelerometers," released by released by the Wireless Sensor Data Mining (WISDM) Lab [12]. Transformed data that is experimented with follows Attribute-Relation File Format. For transformation 10 seconds of accelerometer data samples were taken and transformed into a single tuple of 46 values, each having a unique identifier represented by UNIQUE_ID. The features generated were mostly statistical measures. The features are average, peak, absolute and standard deviation and resultant which is the average of square root of sum of x, y, z values. Finally class is the activity that a user is performing during the example.

IV. CLASSIFICATION ON MATLAB CLASSIFICATION LEARNING APP

A. Overview

In this paper a method for automatic detection of human activity, given a mobile phone accelerometer data, is presented. Target is to develop a classifier that gives best results using Matlab platform and Matlab Classification Learner App. The strategy is trained and tested on the data set, Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. The data set consists of gyroscope and accelerometer data recorded for activities like, walking, climbing stairs, siting, standing and laying.

B. Pre-Processing of Training Data

First the raw sensor train and test data were loaded from the data set folder. The data in training .txt file is converted into tables and saved as .mat files for later access. The raw sensor training data is analyzed by plotting the accelerometer x y and z axis on three different graphs. In pre-processing of data mean is taken for every 128 points as there were 128 readings per window in the raw training data.

C. Additional Feature Extraction

For better learning some more features were extracted from the training data including, mean, standard deviation and

Authorized licensed use limited to: University of Prince Edward Island. Downloaded on November 10,2020 at 16:20:24 UTC from IEEE Xplore. Restrictions apply.

principal component analysis. These features were used for learning

D. Using Classification Learner for Training a Model and Assessing its Performance

Matlab Classification Learner Application is used for automated training of different models which could be used to classify the test data. The training data is used to train different models including Support Vector Machines (SVM) and k Nearest Neighbor KNN models. To find the best classification model the results from all the training models were compared. Figure 2 shows Scatter Plot for Ensemble Model and figure 3 shows the results of three different models on the dataset. It is concluded that the bagged trees ensemble model give best results i.e. 96.2%. The model is exported to the Matlab workspace for future access and for activity prediction of test data.

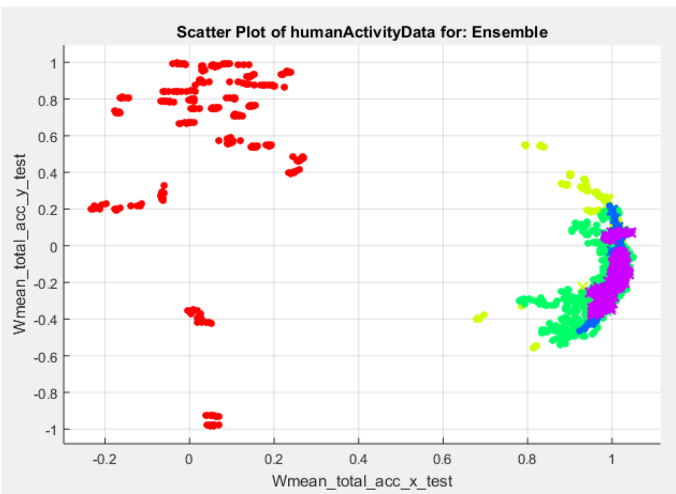


Figure 2 Scatter Plot of the data for Ensemble Model

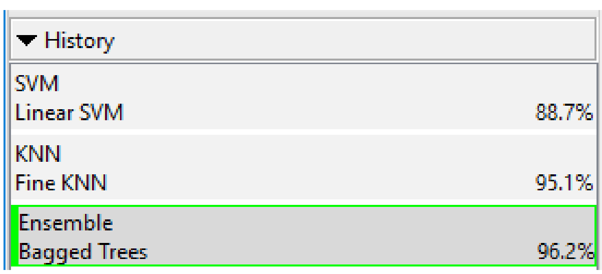


Figure3 Results of Models

E. Activity Prediction of Sensor Test Data

After training the raw sensor test data, mat files were loaded and its features were extracted. The extracted features were again mean, standard deviation and principal component analysis. The classifier data exported previously is used to predict activity of the raw sensor test data. To visualize results the accelerometer data is plotted on a continuous graph and the prediction results were displayed on Matlab interface.

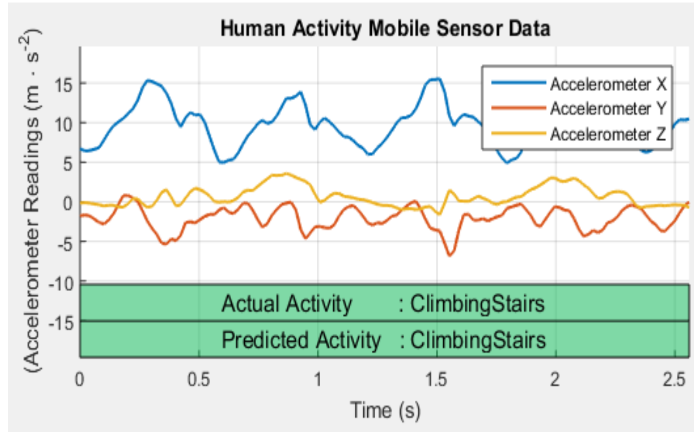
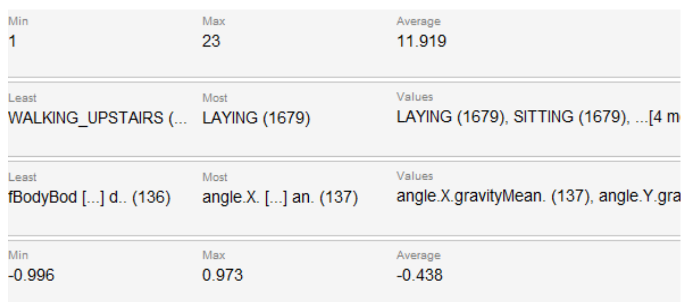


Figure 4 Human Activity Prediction on Test Data

F. General Statistics

The dataset is analyzed through different aspects. The findings of analysis are given. The minimum, maximum and average values for each attribute were found. Minimum reading is that of walking upstairs and least reading were that of laying.



Min 1	Max 23	Average 11.919
Least WALKING_UPSTAIRS (...	Most LAYING (1679)	Values LAYING (1679), SITTING (1679), ...[4 m
Least fBodyBod [...] d.. (136)	Most angle.X. [...] an. (137)	Values angle.X.gravityMean. (137), angle.Y.gra
Min -0.996	Max 0.973	Average -0.438

Figure 5 General statistics for subject, activity, attribute, value respectively (top to bottom)

G. Correlation

The degree of association between each attribute is found by finding the correlation. The mathematical derivation of correlation is computed through summation and product with results between negative 1 and positive 1 which shows negative and positive association respectively. The correlation matrix is presented in figure 6.

Attribut...	id	activity	attribute	value
id	1	-0.018	-0.000	0.012
activity	-0.018	1	-0.000	0.425
attribute	-0.000	-0.000	1	-0.444
value	0.012	0.425	-0.444	1

Figure 6 Correlation Matrix

Authorized licensed use limited to: University of Prince Edward Island. Downloaded on November 10,2020 at 16:20:24 UTC from IEEE Xplore. Restrictions apply.

H. Decision Tree

The decision tree shows that how much possibility of outcome of certain event. By attribute value different events outcome are categorize and it is computed by using model in which role is set on labels which are assigned to different activities. The decision tree is shown in figure 7.

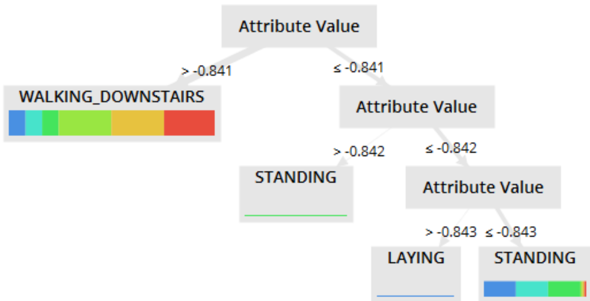


Figure 7 Decision Tree

I. Results

In this paper, a Human Activity Recognition classifier is presented the purpose of this classifier is to identify different activities of human that are walking, standing and laying. After testing classifier on different models it is shown that Bagged Trees gives the best results. Two data sets are created named as "Training Data" and "Testing Data" on the basis of the data obtained from the data from gyroscope and accelerometer in smartphones. For the training 70% data is selected and for testing 30 % data. The results of different models are shown in table 1.

Table 1. Accuracy Results

Models	***Accuracy Percentage***
SVM	88.7%
KNN	95.1%
Bagged Tree	96.2%

V. Classification on Weka

A. Overview

In the paper the different algorithms are applied on the data and evaluated with the goal of finding the best algorithm based on accuracy results. The platform used for applying classifiers on the data is The Waikato Environment for Knowledge Analysis (WEKA) [12]. Weka is chosen for analysis as it is the best tool for research in machine learning. It comprises of functions of a number of machine learning algorithms which can easily be implemented on an Attribute-Relation File Format (.arrf) dataset using graphical user interface. It allows evaluation and comparison of different algorithms on same dataset. The data can be cross validated or a random split can be applied over it to divide the data into training and testing data.

B. Pre-Processing of Training Data

Firstly the raw sensor dataset is loaded into Weka workspace. The dataset is in standard Weka Attribute-Relation File Format (arrf). Weka allows users to pre-process data from its pre-processing tab. There are a variety of filters on attributes which can be applied on raw data. For the purpose of gaining more accurate results filters such as normalize and nominal to binary were applied on the data

C. Classification Algorithms

Four different machine learning models are applied in the data set in order to evaluate their results and identify the best model for the particular dataset. The results for each of the model is discussed in detail.

a) Linear Regression

Linear classification is a machine learning model that linearly separates data into predefined classes.

In order to apply linear regression on the data the pre-processed data, Weka library function is applied. The data set is split into 70% training and 30% for testing. It took 0.2 seconds for the tool to build the model. The root mean square error is 0.115 which is equivalent to 11.5% error or 89.5% accuracy. In a total number of 5418 instances the correlation coefficient is 0.8335.

b) Logistc Regression

In Logistic regression the possibility of an output variable being in one of the two realization is calculated by a linear combination of input variables which is transformed by a logistic function, sigmoid. In this work the data is fed into logistic function and the results of 10 folds cross validation are 4606 correctly classified instances out of a total of 5418 which gives an accuracy of 85.01%.

c) Support Vector Machine Classification

Support Vector machine is a type of supervised learning model in which a non-probabilistic binary classification is done through dividing the data into categories through a clear gap which is as wide as possible unlike the linear and logistic in which there is a line to separate classes. In this work Self Organizing maps (SOM) algorithm is applied on the data set and gives an accuracy of 83.69%

d) Neural Network Classification

Neural network is a layer classifier in which each layer takes some input values along with weights, applies some function and feeds the output to the next layer as input. In this work neural network is applied by applying the neural multilayer perceptron function in Weka tool library. The results of the classifier are shown in figure 11. The accuracy of the model applied on the data set id 87.8% as it correctly identifies 571 instances out of the total 650 instances that were analyzed.

VI. Comparison

Based on the results of the four algorithms, linear regression, logistic regression, support vector machine and neural networks, discussed in detail in the previous section, a comparison is shown in this section. For each of the algorithms accuracy values are calculated form mean absolute error and correctly classified instances. A parametric

Authorized licensed use limited to: University of Prince Edward Island. Downloaded on November 10,2020 at 16:20:24 UTC from IEEE Xplore. Restrictions apply.

study of the above mentioned algorithms is presented in Table 2.

Table 2. Comparative Analysis of Results from WEKA

Comparison Features	Linear Regression	Logistic Regression	Support Vector Machine	Neural Multi-layer Perceptron
Mean Absolute error	0.0798	0.633	0.2269	0.2269
Root mean squared error	0.115	0.1888	0.3176	0.3176
Relative absolute error (%)	92.02	27.10	92.31	92.31
Root relative squared error (%)	55.24	53.99	90.35	90.35
Time to build model (seconds)	2.3	0.36	12.3	11.1

Accuracy results for Linear Regression is 89%, Logistic Regression is 85%, Support Vector Machine is 83% and that of Neural Network is 87%. These results are presented in the form of a graph in figure 12. Based on the accuracy results linear regression shows the highest rate of correctly identified instances closely followed by neural, logistic and support vector. For this dataset human activity can be analyzed better using linear regression but the rest of the three algorithms are also good enough because there results are also good and close to linear in the range of 80%-90%. If compare the time taken to build the model then logistic regression is faster is it took 0.36 seconds to complete while support vector is the slowest consuming 12.3 seconds, hence not as good for real time applications.

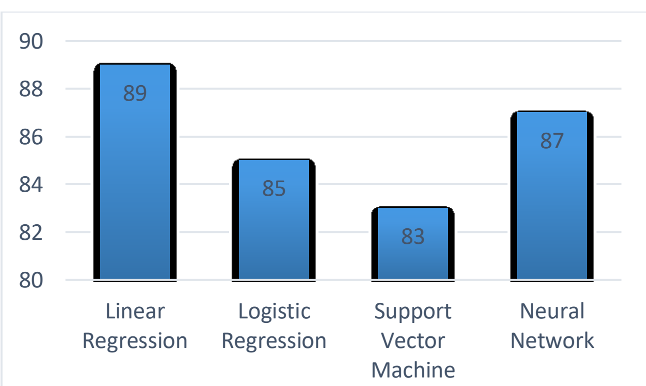


Figure 8 Graphical representation of accuracy of algorithms.

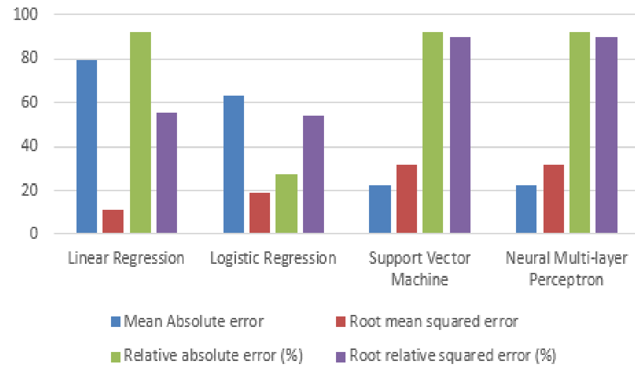


Figure 9 Graphical representation of RME, RSE, RAE, RSE

VII. Conclusion

Increasing amount of work is being done to correctly identify Human Activity as it has many important applications. Using accelerometer data to recognize human activity is very convenient compared to classic ways recognition because it is in-expensive and mobile. Accelerometers are installed in all modern smartphones and is available everywhere. An accelerometer can also be embedded in a device other than smartphones. Recognizing activity in real time is great challenge which can be achieved efficiently only through using machine learning techniques. In this paper some important machine learning techniques have been analyzed and their results compared using Matlab Classification Learner and Weka tool. In Matlab Classification Learner the best model is Bagged Tree with 96.2% accuracy results is identified and it also gives good results in testing phase and on linear regression gives most accurate results compared to logistic, support vector and neural networks for this data set but with slight difference. Logistic regression id better because it is faster hence, should be used where real time recognition is required.

Refferences

[1] K. Eskaf, W. M. Aly and A. Aly, "Aggregated Activity Recognition Using Smart Devices," *3rd International Conference on Soft Computing & Machine Intelligence (ISCMI)*, Dubai, 2016, pp. 214-218.

[2] Y. Tian and W. Chen, "MEMS-based human activity recognition using smartphone," *35th Chinese Control Conference (CCC)*, Chengdu, 2016, pp. 3984-3989.

[3] D. Tao, L. Jin, Y. Yuan and Y. Xue, "Ensemble Manifold Rank Preserving for Acceleration-Based Human Activity Recognition," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1392-1404, June 2016.

[4] M. Zeng *et al*., "Convolutional Neural Networks for human activity recognition using mobile sensors," *6th International Conference on Mobile Computing, Applications and Services*, Austin, TX, 2014, pp. 197-205.

[5] P.Casale, O.Pujol, and P.Radeva, "Human Activity Recognition from Accelerometer Data Using a Wearable Device," *5th Iberian Conference, IbPRIA, Las Palmas de Gran Canaria,* Spain, 2011, pp. 289-296.

[6] S. Matsui, N. Inoue, Y. Akagi, G. Nagino and K. Shinoda, "User adaptation of convolutional neural network for human

Authorized licensed use limited to: University of Prince Edward Island. Downloaded on November 10,2020 at 16:20:24 UTC from IEEE Xplore. Restrictions apply.

activity recognition," *25th European Signal Processing Conference (EUSIPCO)*, Kos, Greece, 2017, pp. 753-757.

[7] M. Zubair, K. Song and C. Yoon, "Human activity recognition using wearable accelerometer sensors," *2016 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, Seoul, 2016, pp. 1-5.

[8] P.Casale, sO.Pujol, and P.Radeva, "Human Activity Recognition from Accelerometer Data Using a Wearable Device," *5th Iberian Conference, IbPRIA, Las Palmas de Gran Canaria,* Spain, 2011, pp. 289-296.

[9] E. Fullerton, B. Heller and M. Munoz-Organero, "Recognizing Human Activity in Free-Living Using Multiple Body-Worn Accelerometers," in *IEEE Sensors Journal*, vol. 17, no. 16, pp. 5290-5297, Aug.15, 2017.

[10] M. Panwar *et al*., "CNN based approach for activity recognition using a wrist-worn accelerometer," *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Seogwipo, 2017, pp. 2438-2441.

[11] M. Poorani, V. Vaidehi and P. Varalakshmi, "Performance analysis of triaxial accelerometer for activity recognition," *Eighth International Conference on Advanced Computing (ICoAC)*, Chennai, 2017, pp. 170-175.

[12] S. Kalpana. "Classification of IRIS Dataset using Weka." *International Journal of Computer Applications & Information Technology*, 2020, *12*(1), pp.287-29

Authorized licensed use limited to: University of Prince Edward Island. Downloaded on November 10,2020 at 16:20:24 UTC from IEEE Xplore. Restrictions apply.