International Journal of Big Data and Analytics in Healthcare
Volume 4 • Issue 1 • January-June 2019

Wearable Devices Data for Activity Prediction Using Machine Learning Algorithms

Lakshmi Prayaga, University of West Florida, Pensacola, USA
Krishna Devulapalli, Indian Institute of Chemical Technology, Secunderabad, India
Chandra Prayaga, University of West Florida, Pensacola, USA

ABSTRACT

Wearable devices are contributing heavily towards the proliferation of data and creating a rich minefield for data analytics. Recent trends in the design of wearable devices include several embedded sensors which also provide useful data for many applications. This research presents results obtained from studying human-activity related data, collected from wearable devices. The activities considered for this study were working at the computer, standing and walking, standing, walking, walking up and down the stairs, and talking while walking. The research entails the use of a portion of the data to train machine learning algorithms and build a model. The rest of the data is used as test data for predicting the activity of an individual. Details of data collection, processing, and presentation are also discussed. After studying the literature and the data sets, a Random Forest machine learning algorithm was determined to be best applicable algorithm for analyzing data from wearable devices. The software used in this research includes the R statistical package and the SensorLog app.

KEYWORDS

Activity Tracking, Data Analytics, Machine Learning, Random Forest Algorithm, Wearable Devices, Windowing And Smoothing

1. INTRODUCTION AND MOTIVATION

Wearable devices can generate multiple types of data such as heart rate, accelerometer, and gyroscope values, location, etc. This data is useful across multiple disciplines, including health care, cybersecurity, user interface design, personalizing social preferences, etc. Healthcare is one such institution that is leveraging this medium of data collection and performing analytics that is informative to healthcare providers, administrators, pharma companies and patients. Such analysis allows the audience in the domain of healthcare to maximize their returns either commercially or personally at an individual level.

Activity Recognition is an emerging field of research, born from the larger fields of ubiquitous computing, context-aware computing and pervasive computing (Pierluigi, Oriol & Petia, 2011; Sztyler & Stuckenschmidt, 2017). Recognizing everyday activities and its relation to overall wellness is generating a lot of interest in the research community of data scientists, pharmaceutical companies and healthcare professionals. Research also documents that monitoring physical activity in real life vs. a controlled environment provides a better context to evaluate patients and or other interested clients (Sztyler & Stuckenschmidt, 2016). The use of accelerometers and gyroscopes in wearable devices such as smartwatches and smartphones are now widely accepted for monitoring physical activity and

DOI: 10.4018/IJBDAH.2019010103

tailoring interventions as needed (del Rosario, Redmond & Lovell, 2015; Akker, Jones, Hermens & Hermie, 2014), without purchasing expensive wearable ambulatory monitors. Smart phones have also proven to be extremely useful to monitor the activity levels of construction workers, a context which provides a wealth of information for project management related to their work (Akhavian & Behzadan, 2016).

Fitness tracking devices are gaining in popularity and new devices are entering the market at regular intervals. Wearable devices include accelerometers, Gyroscope, barometers, and altimeters to provide high-quality data which is useful for tracking posture, activity, HR, sleep, etc. (Henriksen et al., 2018). A wearable device has the potential to be integrated as an intervention to increase physical activity embedding it as a change to lifestyle (Ridgers et al., 2018; Cadmus-Bertram, 2017; Maher, Ryan, Ambrosi & Edney, 2017) Many of these devices are being used for data collection and research on various aspects related to individual health including monitoring physical activity, sleep quality, heart rate, etc. and their impact on patient health.

The literature reviewed cited above demonstrates that recent studies have analyzed accelerometer data and have investigated the data for physical activity recognition. Nevertheless, few of them have undertaken the difficult task of performing experiments out-of-the-lab. The conditions to perform experiments out-of-the-lab create the need to build easy to use and easy to wear systems to free the testers from the expensive task of labeling the activities they perform. This study attempts to address this challenge and afford the ability to generate and analyze data outside the lab in an open and free environment using data recorded by the accelerometer on wearable devices or cell phones. Data generated in such a format can be used to train models using machine learning algorithms and use the models to test new data.

Random Forest machine learning algorithm was used in this study. A review of recent literature suggests that when the Random Forest algorithm's performance was compared to other techniques such as support vector machine, C4.5 and *k*-nearest neighbor methods, Random forest was the most accurate and suitable for the analysis of data from wearable devices (Balli, Sağbaş & Peker, 2019; Henriksen et al., 2018; Zhang, Stogin & Alshurafa, 2018). Random Forest was also documented to be specifically reliable to predict the gait of a subject. (Ahamed et al., 2018) which is applicable to the current research. Random forest is one of the most popular machine learning algorithms. Machine learning algorithms are successful because they provide in general a good predictive performance, low overfitting, and easy interpretability. This interpretability is given by the fact that it is straightforward to derive the importance of each variable on the tree decision. In other words, it is easy to compute how much each variable is contributing to the decision. Random Forest algorithm has also yielded high accuracy in classification problems due to the identification of important features (Natarajan, Kumar & Selvaraj, 2018).

Data preparation and preprocessing using Random Forest involves feature selection, a process which we have used in our case and described in section 4 below. Feature selection algorithms fall into three categories: filters, wrappers, and embedded techniques. The Random forest algorithm fits into the embedded techniques category. Embedded methods combine the qualities of filter and wrapper methods. They are implemented by algorithms that have their own built-in feature selection methods. Some of the benefits of embedded methods are:

- They have a high accuracy
- They generalize better
- They are interpretable

R Language is an open source popular language for statistical analysis and for data science applications. R Language provides thousands of packages covering various applications. In the present study, the Random Forest package of R, which implements Breiman's random forest algorithm, was utilized for performing the calculations.

2. EXAMPLE USE CASES OF THIS STUDY

Studies show that nearly 40 million workers in Europe are diagnosed with Musculoskeletal disorders (MSD) attributed to their work due to repetitive movements and improper postures), and close to 40% of chronic conditions for patients over the age of 16 years was also attributed to MSD (Cammarota, 2003). Several studies have documented the efficacy and correlation of good posture in maintaining good health. McGinnis et al. (2017) cited other researchers documenting the correlation of maintaining good posture and reduction of stress levels, reduction of depression and other chronic conditions. These researchers have found that maintaining proper posture is a key element to ensure the overall quality of health of the human body. Additionally, when conditions such as fractures, muscle tears, and other bone and muscle related incidents happen, maintaining good posture is essential for speedy recovery.

When patients visit their physicians or health care providers with fractures and/or other related injuries, x-rays are used to determine the extent of the injury and appropriate treatment is provided to the patient. Healthcare providers are trained to determine the correct posture that the patient must strive to maintain to recover from the injury.

The problem, in this case, is that the patient does not know how to check if he/she is maintaining the right posture and if necessary, rectify the posture, until the next scheduled visit to the facility of the provider, who hosts expensive infrastructure and possesses the domain knowledge required to evaluate the patient. The physician then takes another set of x-rays, studies them and observes the patient's posture to determine the extent of improvement. If the posture is not conducive to a speedy recovery, the physician might again send the patient to a physiotherapist who can assist the patient to improve the posture. This cycle repeats over and over again until satisfactory results are obtained.

Ideally, it is desirable to propose a system of intervention that patients can use, in between visits to the healthcare provider, which could assist the patient in determining, and, if necessary, correcting, their posture based on some quantifiable values. This solution is two-fold in that we use machine learning algorithms for training and testing activity related data and determine the accuracy of activity prediction. In the present study, we apply the Random Forest machine learning technique to the acceleration data collected from eight volunteers' smartphones for training and testing the model. The volunteers used the SensorLog app which affords the ability to export the accelerometer data to a CSV file. Volunteers performed each of the six activities described below for five to ten minutes and uploaded the corresponding files for each activity to the researchers.

3. METHODOLOGY

The statistical data analysis package R and the SensorLog mobile app were used to analyze the data for this study. R is a popular statistical package used to apply machine learning algorithms and perform data analytics. The data for this research project is collected from an app (SensorLog) installed on the smartphone. Most raw data collected by the sensors from the wearable devices must be interpreted and transformed into a format that can be understood by the naïve user. Mobile apps are generally used for data transformation. In our study, the SensorLog app available both on the App Store and Google play store was used to collect the raw data from the wearable devices and transmit it as a CSV file via the Internet. Figures 1 to 4 provided below show the interface provided by the SensorLog app.

User Interface of the SensorLog App

The SensorLog app has a user interface that can be configured to collect specific data and also options to save the data to a particular file format. In this research the researchers collected the data and saved it as a csv file which can be downloaded or shared. The four figures below describe the user interface of the SensorLog.

Figure 1 displays the accelerometer data collected by the sensorLog app.

Figure 2 displays the screen to specify and configure the file type, recording rate, device ID etc.

Figure 3 displays the screen for the Logged Files that are available

Figure 4 displays the available options to share and or download the logged file

A description of the data collection, processing and analyzing are presented below.

3.1 Description of Data

The dataset consisting of uncalibrated accelerometer data with a sampling frequency of 30 Hz, is collected from 8 participants using their wearable devices (mobile phones) performing six activities. These activities are referred to as labels and are codified as follows

1: Working at a computer
2: Standing and walking
3: Standing
4: Walking
5: Walking up and down stairs
6: Talking while standing

For each participant, the corresponding csv file which can be downloaded contains the following information: sequential number, x acceleration, y acceleration, z acceleration and activity label. The activity label is codified as numbers 1 to 6 where each activity has a corresponding number associated with it. Working at the computer has a code of 1, Standing has a code of 3, etc.

Note: In the data tables and plots in this paper, the classes are labelled according to the above list of activities

Data Preparation and Preprocessing

Table 1 presented below provides a list of variables for which raw data were collected from the wearable device corresponding to the activity - sitting at a computer.

From this list, a set of values for the following fields were used for the prediction of the activity. These include

```
LoggingTime,
LoggingSample,
AccelerometerAccelerationX(G),
AccelerometerAccelerationY(G),
AccelerometerAccelerationZ(G).
```

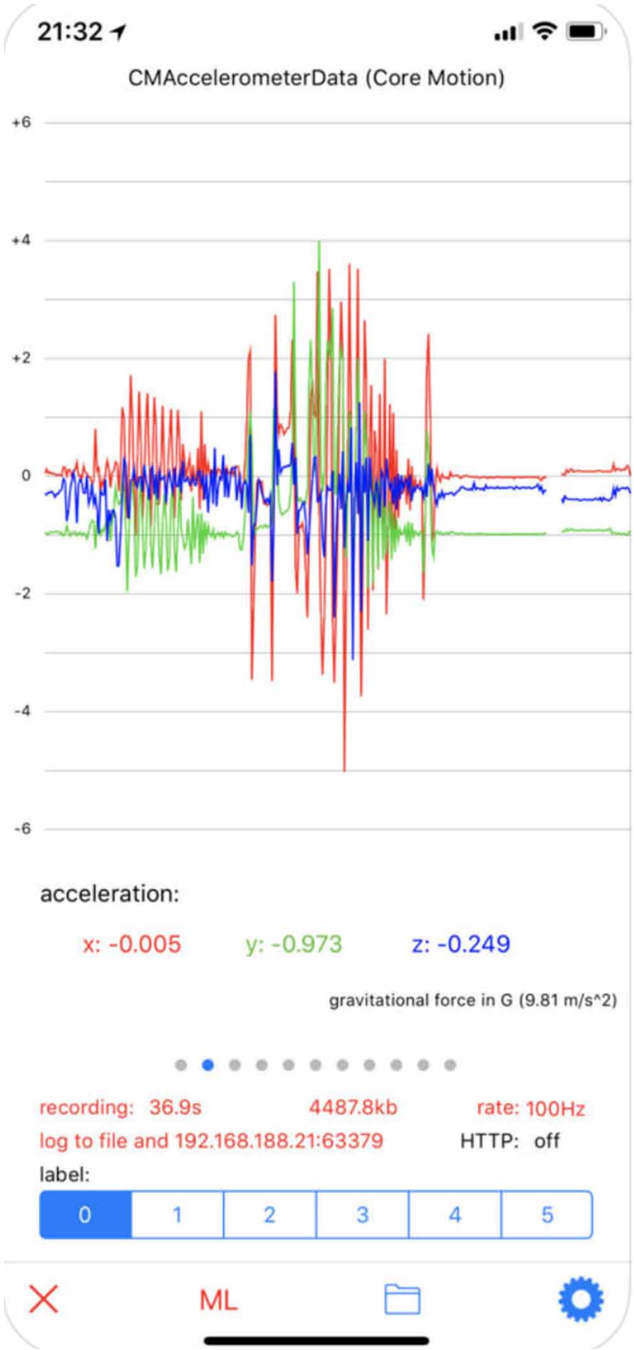
Acceleration data recorded in the dataset are coded according to the following mapping: [0; +30] = [-1.5g; +1.5g]. Ehatisham-ul-Haq, Azam, Naeem, Rehman and Khalid (2017) observed that the time series generated by smartphones generally contains noise generated by the participants and by the smartphones. So, the coded data is smoothed by Holt Winter exponential smoothing model. It is to be noted that data obtained under carefully controlled conditions may contain much less noise and yield much better accuracy (Tillis 2016).

As suggested by Pierluigi et al. (2011), features have been extracted by windowing of 75 samples, corresponding to 2.5 seconds of accelerometer data, with 50% of overlapping between windows. From each window, fifteen features have been extracted corresponding to means, standard deviations, minimum, maximum and median values for the three axes x, y, and z. As stated earlier, R language software was utilized for generating these new set of features.

3.2 Classification and Prediction of Activities

In order to classify and predict the individual's activity based on the newly derived features, the Random Forest machine learning model was utilized. For fitting Random Forest, software programs were developed using the R language.

Figure 1. SensorLog accelerometer data



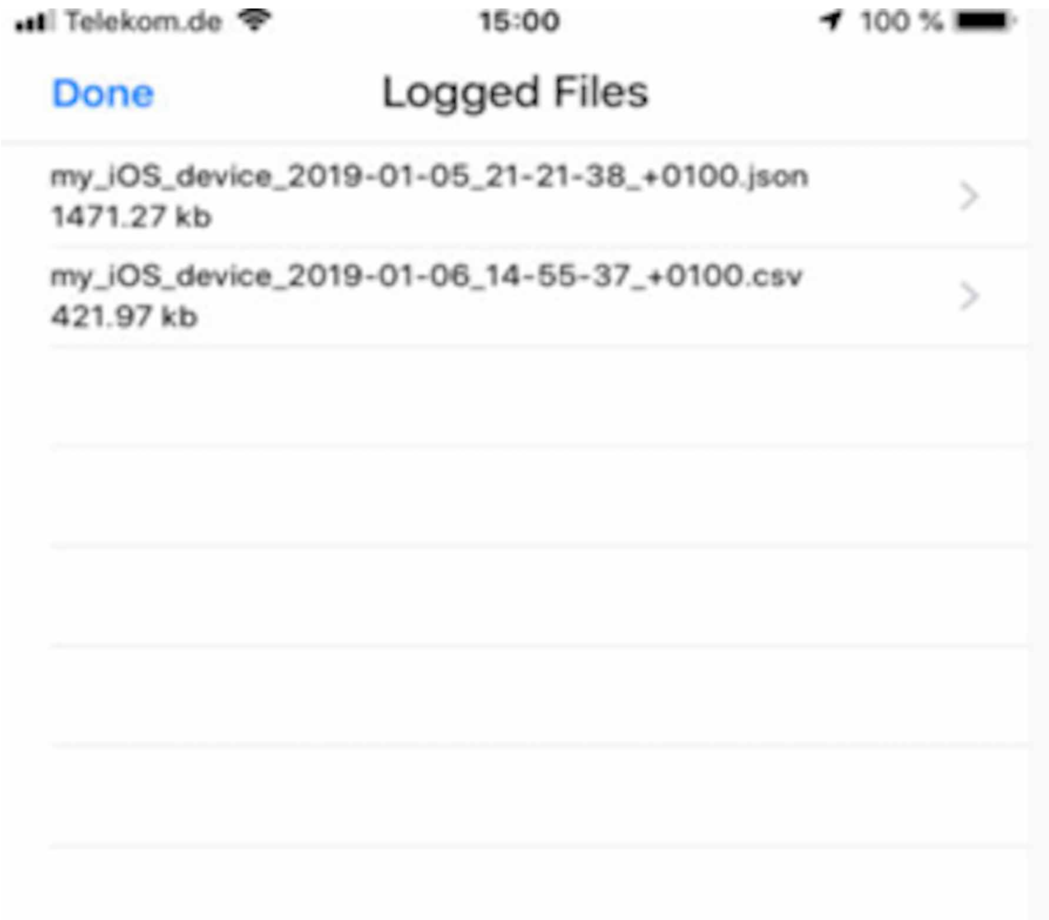
The newly derived dataset consisting of 15 features is divided into two datasets viz., training dataset and test dataset. The training dataset is obtained by taking 80% of the randomly selected observations and the remaining 20% of the dataset is utilized as the test dataset. Random Forest model is fitted with the training dataset and tested with the test dataset. With the Random Forest model, the important variables were also identified.

Figure 2. SensorLog configuration options

4. RESULTS AND DISCUSSION

Results from the Random Forest Model are presented in Tables 2, 3 and 4. Table 2 includes the resulting OOB error and confusion matrix obtained from the Training dataset. The Confusion Matrix of the results displayed in Table 2 demonstrates the accuracy of the model. In the Confusion Matrix, the diagonal elements show the number of observations which are correctly classified for each activity, non-diagonal elements are the number of observations, which are not correctly classified. The overall accuracy of the Random Forest Model is given by Out of the Bag (OOB) estimate of error rate, which

Figure 3. Files generated by SensorLog app



is 11.67%. This clearly demonstrates that the Random Forest Model has accurately classified all the activities with 88.33% accuracy in the training dataset.

The labels of the six classes listed in the top row and left column in the confusion matrix are as follows:

1: Working at Computer
2: Standing and Walking
3: Standing
4: Walking
5: Walking Up and Down Stairs
6: Talking while Standing

Prediction for Test Dataset

The results of the test dataset for the Random Forest Model are given in Table 3a, Table 3b and Table 3c. Table 3a displays the confusion matrix, Table 3b. displays the overall statistics and Table 3c. displays the statistics by class obtained from the test dataset. From the overall statistics, we observe that the overall accuracy of the model for the Test Dataset is 85.94% accurate.

Figure 4. SensorLog file export options

The labels of the six classes listed in the top row and left column in the confusion matrix are as follows:

1: Working at Computer
2: Standing and Walking
3: Standing
4: Walking
5: Walking Up and Down Stairs

Table 1. List of variables

loggingTime(txt)	motionUserAccelerationY(G)
loggingSample(N)	motionUserAccelerationZ(G)
locationTimestamp_since1970(s)	motionAttitudeReferenceFrame(txt)
locationLatitude(WGS84)	motionQuaternionX(R)
locationLongitude(WGS84)	motionQuaternionY(R)
locationAltitude(m)	motionQuaternionZ(R)
locationSpeed(m/s)	motionQuaternionW(R)
locationCourse(å¡)	motionGravityX(G)
locationVerticalAccuracy(m)	motionGravityY(G)
locationHorizontalAccuracy(m)	motionGravityZ(G
locationFloor(Z)	activityTimestamp_sinceReboot(s)
accelerometerTimestamp_sinceReboot(s)	activity(txt)
accelerometerAccelerationX(G)	activityActivityConfidence(Z)
accelerometerAccelerationY(G)	activityActivityStartDate(txt)
accelerometerAccelerationZ(G)	pedometerStartDate(txt)
motionTimestamp_sinceReboot(s)	pedometerNumberofSteps(N)
motionYaw(rad)	pedometerAverageActivePace(s/m)
motionRoll(rad)	pedometerCurrentPace(s/m)
motionPitch(rad)	pedometerCurrentCadence(steps/s)
motionRotationRateX(rad/s)	pedometerDistance(m)
motionRotationRateY(rad/s)	pedometerFloorAscended(N)
motionRotationRateZ(rad/s)	pedometerFloorDescended(N)
motionUserAccelerationX(G)	pedometerEndDate(txt)

Table 2. Random forest model results for the training dataset

	1	2	3	4	5	6
1	454	23	7	9	1	7
2	24	628	10	7	21	21
3	8	29	433	1	4	11
4	3	8	0	758	27	37
5	1	15	1	28	418	19
6	4	23	6	30	23	398

Call: randomForest(formula = activity ~ ., data = train, ntree = 500, mtry = 5)
Type of random forest: classification
Number of trees: 500
No. of variables tried at each split: 5
OOB estimate of error rate: 11.67%

Table 3a. Predictions for test dataset – confusion matrix

	1	2	3	4	5	6
1	90	4	3	0	0	5
2	13	162	4	4	8	6
3	2	4	124	0	0	2
4	1	1	0	188	9	6
5	3	2	2	8	97	9
6	3	3	2	10	9	91

Table 3b. Predictions for test dataset – overall statistics

Accuracy	0.8594
95% CI	(0.8346, 0.8818)
No Information Rate	0.24
P-Value [Acc > NIR]	< 2.2e-16
Kappa	0.8291
Mcnemar’s Test P-Value	NA

Table 3c. Predictions for test dataset – statistics by class

	Class1	Class2	Class3	Class4	Class5	Class6
Sensitivity	0.8036	0.9205	0.9185	0.8952	0.7886	0.7647
Specificity	0.9843	0.9499	0.9892	0.9744	0.9681	0.9643
Pos Pred Value	0.8824	0.8223	0.9394	0.9171	0.8017	0.7712
Neg Pred Value	0.9715	0.9794	0.9852	0.9672	0.9655	0.9630
Prevalence	0.1280	0.2011	0.1543	0.2400	0.1406	0.1360
Detection Rate	0.1029	0.1851	0.1417	0.2149	0.1109	0.1040
Detection Prevalence	0.1166	0.2251	0.1509	0.2343	0.1383	0.1349
Balanced Accuracy	0.8939	0.9352	0.9539	0.9348	0.8784	0.8645

6: Talking while Standing

The labels of the six classes are as follows:

1: Working at Computer
2: Standing and Walking
3: Standing
4: Walking
5: Walking Up and Down Stairs
6: Talking while Standing

The RandomForest package in R reports the results for the training dataset by giving the OOB estimate of error rate, and the results for the Test dataset by providing the overall statistics, as reported above.

4.2 Identifying Important Variables

The important variables identified by the Random Forest Model are given in Table 4. The MeanDecreaseGini values give the relative importance of each individual variable, which is provided under the column "Overall". From this table, it is observed that the variables ymax, ymean, xmax, which have higher MeanDecreaseGini values are the important features followed by other features. The overall OOB error for different activities as functions of the number of trees in the Random Forest model is given in Figure 5.

Explanation of Figure 5: The OOB (out of the Bag) error estimate is similar to that obtained by N-fold cross-validation (Hastie, Tibshirani & Friedman 2016). It steadily decreases as the number of trees are added in the model and reaches lowest values around 500 trees. The labels of the six classes are as follows:

1: Working at Computer
2: Standing and Walking
3: Standing
4: Walking
5: Walking Up and Down Stairs
6: Talking while Standing

Table 4 below lists the hierarchy of importance of the variables

Figure 5. Overall OOB error for different activities

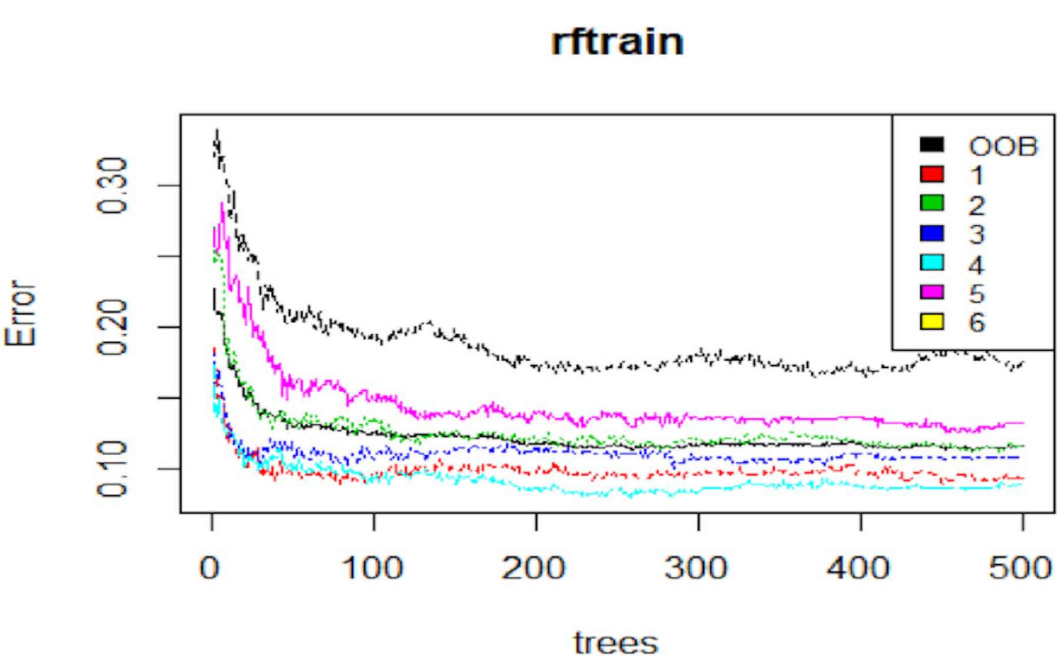


Table 4. Overall variables importance

xmean	187.2409
ymean	252.3255
zmean	134.8651
xsd	200.4376
ysd	135.1334
zsd	168.6573
xmin	188.7961
ymin	219.7125
zmin	139.0769
xmax	234.8544
ymax	298.4171
zmax	181.0266
xmedian	191.5964
ymedian	211.1288
zmedian	137.1785

Figure 6 also visually displays the same hierarchy of importance of variables. The RandomForest package automatically identifies the important variables/features based on MeandecreaseGini values. Figure 6 is obtained by considering the Mean Decrease in Node impurity.

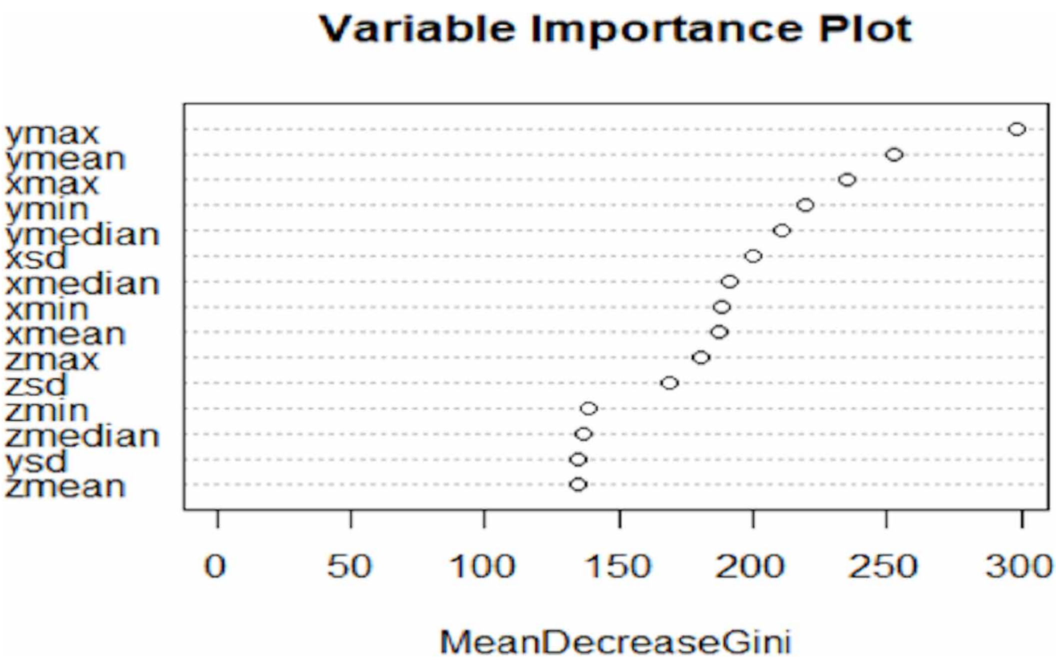
5. CONCLUSION AND FUTURE PROSPECTS

An attempt is made in this study to predict the activity of the individual by utilizing the accelerometer data obtained from smartphones. As the data is available with a frequency of 30 Hz, a window of 75 observations is taken, corresponding to two and a half seconds, with 50% overlapping. From the rolling windows, fifteen features have been extracted for the three x, y, and z accelerometer data. The Random Forest Machine Learning algorithm was utilized to predict the six activities utilizing the data. 80% of the randomly selected data was utilized as a training set and the remaining 20% of the data was utilized for testing the model. Random Forest model has identified the activities with 88.33% accuracy in the training dataset and 85.94% accuracy in the test dataset.

6. FUTURE WORK

Some of the limitations in this study are that a. practical usability of the study and b. extensibility. To ensure practical usability of this study the researchers plan to design user interfaces (UI) to allow users to monitor and test their own data. Examples include using a UI such as a shiny app that allows a user to upload data acquired from a wearable device, and obtain a decision on whether their activity and or posture was accurately being monitored. Data collected from such applications can be used by medical practitioners to determine the impact on the patient's health. Additionally, the researchers plan to expand this study to include a much larger population in different domains of knowledge.

Figure 6. Variables importance plot



REFERENCES

Ahamed, N. U., Kobsar, D., Benson, L., Clermont, C., Kohrs, R., Osis, S. T., & Ferber, R. (2018). Using wearable sensors to classify subject-specific running biomechanical gait patterns based on changes in environmental weather conditions. *PLoS One*, *13*(9), e0203839. doi:10.1371/journal.pone.0203839 PMID:30226903

Akhavian, R., & Behzadan, A.H. (2016). Smartphone-based construction workers' activity recognition and classification. *Automation in Construction*, 71, 198-209.

op den Akker, H., Jones, V. M., & Hermens, H. J. (2014). Tailoring real-time physical activity coaching systems: a literature survey and model. User modeling and user-adapted interaction, 24(5), 351-392.

Balli, S., Sağbaş, E. A., & Peker, M. (2019). Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm. *Measurement and Control*, *52*(1-2), 37–45. doi:10.1177/0020294018813692

Cadmus-Bertram, L. (2017). Using Fitness Trackers in Clinical Research: What Nurse Practitioners Need to Know. *The Journal for Nurse Practitioners*, *13*(1), 34–40. doi:10.1016/j.nurpra.2016.10.012 PMID:28603469

Cammarota, A. (2003), "The commission's initiative on MSDS: Recent developments in social partner consultation at the European level. *Presented at the Conference on MSDs — A Challenge for the Telecommunications Industry*, Lisbon, Portugal, October 20–21 (pp. 20-21).

del Rosario, M. B., Redmond, S. J., & Lovell, N. H. (2015). Tracking the Evolution of Smartphone Sensing for Monitoring Human Movement. *Sensors*, *15*(8), 18901–18933. doi:10.3390/s150818901 PMID:26263998

Hastie, T., Tibshirani, R., & Friedman, J. (2016). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

Henriksen, A., Haugen Mikalsen, M., Woldaregay, A. Z., Muzny, M., Hartvigsen, G., Hopstock, L. A., & Grimsgaard, S. (2018). Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables. *Journal of Medical Internet Research*, *20*(3), e110. doi:10.2196/jmir.9157 PMID:29567635

Henriksen, A., Haugen Mikalsen, M., Woldaregay, A. Z., Muzny, M., Hartvigsen, G., Hopstock, L. A., & Grimsgaard, S. (2018). Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables. *Journal of Medical Internet Research*, *20*(3), e110. doi:10.2196/jmir.9157 PMID:29567635

Ehatisham-ul-Haq, M., Azam, M. A., Naeem, U., ur Rèhman, S., & Khalid, A. (2017). Identifying Smartphone Users based on their Activity Patterns via Mobile Sensing. *Procedia Computer Science*, *113*, 202–209.

Maher, C., Ryan, J., Ambrosi, C., & Edney, S. (2017). Users' experiences of wearable activity trackers: A cross-sectional study. *BMC Public Health*, *17*(1), 880. doi:10.1186/s12889-017-4888-1 PMID:29141607

Natarajan, B., Sundhara Kumar, K. B., & Selvaraj, C. (2017). Empirical study of feature selection methods over classification algorithms. *International Journal of Intelligent Systems Technologies and Applications*.

Pierluigi, C., Oriol, P., & Petia, R. (2011), Human Activity Recognition from Accelerometer Data Using a Wearable Device. In *Pattern Recognition and Image Analysis: 5th Iberian Conference, IbPRIA 2011*, Las Palmas de Gran Canaria, Spain, June 8-10 (pp. 289-296).

Ridgers, N. D., Timperio, A., Brown, H., Ball, K., Macfarlane, S., Lai, S. K., & Salmon, J. et al. (2018). Wearable activity tracker use among Australian adolescents: Usability and acceptability study. *JMIR mHealth and uHealth*, *6*(4), e86. doi:10.2196/mhealth.9199 PMID:29643054

McGinnis, R. S., DiCristofaro, S., Mahadevan, N., Sen-Gupta, E., Silva, I., Jortberg, E., ... & Patel, S. (2017, August). Longitudinal Data from Wearable Sensor System Suggests Movement Improves Standing Posture. In *Proceedings of the 41st Annual Meeting of the American Society of Biomechanics*, Boulder, CO (pp. 8-11).

Tillis, R. (2016). Machine Learning Project – random forest – Sensor Data. *RPubs.*

Zhang, S., Stogin, W., & Alshurafa, N. (2018). I sense overeating: Motif-based machine learning framework to detect overeating using wrist-worn sensing. *Information Fusion*, *41*, 37–47.

Sztyler, T., & Stuckenschmidt, H. (2016). On-body localization of wearable devices: An investigation of position-aware activity recognition. In *Proceedings of the 2016 IEEE International Conference on Pervasive Computing and Communications (PerCom).* IEEE. doi:10.1109/PERCOM.2016.7456521

Sztyler, T., & Stuckenschmidt, H. (2017). Online personalization of cross-subjects based activity recognition models on wearable devices. In *Proceedings of the 2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)* (pp. 180-189). IEEE. doi:10.1109/PERCOM.2017.7917864

Lakshmi Prayaga is an Associate Professor in the department of Information Technology at the University of West Florida. Her research interests include data analytics, data visualizations, educational robotics, games and simulations. She teaches several of the courses related to these topics. She also received several grants for research projects on these topics.

Krishna Devulapalli is a retired Chief Scientist from the IT Department of Indian Institute of Chemical Technology, Hyderabad, India. Currently, he is a freelance Data Scientist, TapChief Expert in Data Science and Guest Faculty in BioStatistics, NIPER, Guwahati, India. His areas of interest include Machine Learning, Data Visualization, Statistical Modeling, Pattern Recognition, etc. He has published numerous research papers in various journals in the areas of Chemometrics, BioStatistics, HealthCare Analytics, etc. He has delivered lectures in all these areas for a number of students from various Universities and currently providing online Capstone Project work mentoring in the data science domain, career counseling, online mock interviews, etc.