# Recognition of Human Activities Using Rain Forest Algorithm with Smartphone Sensors

**Kruti Nile[1],Vaishali Salve[2]**
[1] VJTI College
[2]Sinhgad College Of Science

***Abstract-*** *Improvement in everyday life, may it be lifestyle improvement or health care, requires recognition of activities. There have been a number of methods for measuring physical activities, such as self-reporting, attaching wearable sensors, etc. Since a smartphone has become widespread rapidly, physical activity can be easily measured by accelerometers in the smartphone. There are six actions that are selected for recognition: walking, standing, sitting, lying down, walking upstairs, walking downstairs. Although there is a lot of study done successfully for activity recognition exploiting smartphone's accelerometer data, this paper aims to improve the efficiency of recognition of activities in terms of time and accuracy by using much better and efficient machine learning algorithm, The Random Forest Algorithm*

***Keywords***- Random Forest Algorithm(RF), Support Vector Machine(SVM), Accelerometer, Human Activity Recognition(HAR).

## I. INTRODUCTION

There is a close relation between daily physical activities and health status of a person. World Health Organization (WHO) identified that physical inactivity has been considered as the fourth leading risk factor for the deaths in the world. Physical inactivity may cause a lot of severe chronic diseases, such as coronary heart diseases(CHD), blood liquid profile, hypertension, ischemic stroke, type 2 diabetes, colon cancers, glucose tolerance and insulin sensitivity, anxiety and depression, fall injuries, etc. Therefore, in wellbeing, healthcare, and sports, it is often important to capture the activities being performed by a subject. Efforts are being widely made to encourage and improve the quantity as well as the quality of physical activities and effectively keep track of these activities for a good and healthy lifestyle. For instance, the intensity of physical activity is worth measuring because it significantly affects the health of an individual. One way of measuring physical activity is by self-reporting. However, it has some drawbacks, such as, it may fail to include every activity pattern, also may fail to measure an accurate amount of physical activity. Another way is to use wearable sensors by attaching on a human body. Although wearable sensors can retrieve records of physical activity, the subject may feel uncomfortable wearing the devices. Smartphones are being widely used by almost everyone not just for calling, as name suggests they are getting more and more smarter. Thus, the idea of utilization of these sensors to make a smartphone recognize human activities becomes more realistic. This technology has the potential to promote the development of assistive technologies, for example helping the elderly people to live a better life. The human activities recognition, uses sensors to recognize human actions, to produce more simple and handy system with high precision.

## II. WORKING

Human activity recognition has been considered as one of the challenging topics in the information technologies. To classify one's activity accurately, it has been widely studied with various sensors, such as accelerometers, GPS, compasses, microphones, temperature sensors, etc. In the last decades, there were several machine learning methods that were be used for classifier and recognition of human physical activities including Naive Bayes, Support Vector Machines (SVMs), Threshold based and Markov chain. Although there has not been any study being done that can find out the best method for human physical activities classification, but SVMs have been successfully widely used in many research related to handwriting recognition and speech recognition. But SVM's although is able to faithfully classify human activities is a slow algorithm in terms of real time classification. Also SVM's acts very good for bi-class classification problems. HAR is a multi-class classification problem. Therefore, this paper makes an effort to design an efficient system for HAR and trying to improve the efficiency of system by using a better algorithm than SVM, The Random Forest Algorithm.

## III. METHODOLOGY

**There are two main steps involved in designing a model:**

1.) **Training**
2.) **Testing**

Human Activities Recognition System are formed from various functional blocks. Each block performs a different task for training process and identifying actions. The human activities recognition system consists of four main functional modules:

- Data acquisition and data processing module
- Feature extraction module
- Training module
- Human activities recognition module.

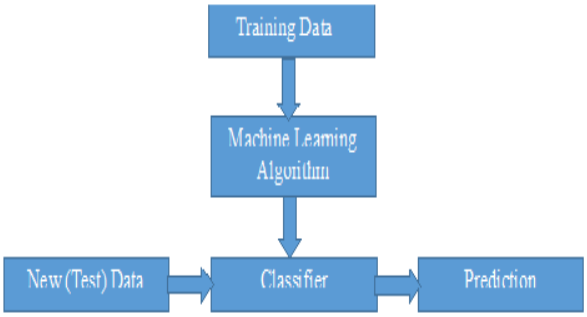



Fig.1.1 Training Process

**1.) Training Environment:**

**A. Data acquisition & data processing:**

In this module the data is collected. The data is processed like filtering, cleaning the data etc. This data is collected from the user's smartphone sensor modules i.e. accelerometer and gyroscope performing the human activities. In order to execute the data acquisition process, this module is implemented in two steps. Step 1: establishing a connection between the controller and the data recording device. Step 2: Turn the control signal to start the gathering process.

B. **Feature Extraction:**

The raw data collected from the sensor cannot be used directly for machine learning algorithm. Thus, a module, which conducts feature extraction and converts this raw data into training samples, is implemented. This data needs to be brought into proper form. Data is collected from 2 types of sensor: gyro sensor and tri-axial accelerometer sensor. Each sensor returns three values corresponding to three dimensional x, y, z. The sensor runs at 50Hz (collecting 50 values per second) to store raw data in units of sample. Each sample includes 128 values corresponding to each sample time record of 2.56 seconds.

Table 1.1

| | |
|---|---|
| Mean | Normal statistical mean |
| Standard Deviation | Normal statistical Standard deviation |
| Mad | Average of standard deviation |
| Max | Maximum |
| Min | Minimum |
| Energy | Energy of signal |
| Entropy | Entropy of signal |

These features in table 1.1 calculated in the time domain. There are two versions of this system include MATLAB version for testing in PC and Android version for implementing in a smartphone. In MATLAB, total 561 features are calculated. On other side, Android version has 248 features. The raw data is collected from 10 volunteers aged from 19 to 48. All of them have normal health status. The phone placed in their pants pocket.

**C. Training:**

Training converts extracted features to the recognition model, which is used as a template for activity recognition. Results of recognition are analyzed and compared to find out the fault location as well as features are calculated from those faults in order to adjust the calculation to achieve better recognition model.

**D. Recognition:**

To design the system for real time operation, the recognition module is processed in a short time of 3-5 seconds. In particular, this period of time includes the time of writing data and the time of recognition for activity. Writing data module and processing data from sensor module records data for 2-3 seconds (128 values from each sensor). The amount of data is processed through the features extraction module. These features are calculated in time domain over time. Totally, thus 248 features are calculated and six activities namely sitting, standing, laying, walking, walking upstairs and walking downstairs are considered for recognition. Then, these features are recognized with the training model, which is obtained from training module. Recognition results can be displayed to the user on a PC which can be further sent for application like to a healthcare or elderly people doctor where it's not feasible for a doctor to always be around a patient to keep a track of all physical activities or in sports for stamina analysis or fitness analysis etc. of a sports person.

## IV. METHODS

### SUPPORT VECTOR MACHINE:

Support Vector Machine(SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. This SVM is known to work really well with clear margin of separation and high dimensional spaces. Also works very well where number of dimension is greater than the number of samples. This algorithm uses a subset of training points in the decision function (called support vectors), so it is also memory efficient. In order to find out the best hyperplane for data classification, SVMs search the hyperplane which has the largest margin.

### RANDOM FOREST:

Random forest algorithm(RF) is a supervised classification algorithm. As the name suggest, this algorithm creates the forest with a number of trees. In general, more the trees in the forest more robust the forest looks like. In the same way in the random forest classifier, higher the number of trees in the forest higher the accuracy results. Random forest classifier is known to handle the missing values. Even though the number of trees are more in the forest, this doesn't over fit the model. Random Forest simply takes the test features and use the rules of each randomly created decision tree to predict the outcome and stores the predicted outcome (target) then calculates the votes for each predicted target. The high voted predicted target is considered as the final prediction.

### 2.) Testing Environment:

The experiment was carried out on 10 volunteers aged 19-48, having normal health. Each volunteer performing six basic activities. The process of data collection controlled by a computer program, and data named separately to manage, store. Data accompanied by information on the action name, the name of volunteers and time data logging. Then data to be randomized into two data sets: 70% for the training, 30% for the inspection process. The training will be done by machine learning algorithm. Samsung S2 used for this test (any smartphone can be used though). Smartphones contain two essential sensor including accelerometer sensor and gyroscope. The sensor data recorded at 50Hz, suitable for recording data on human activities. For the process of identification, an application on smartphones running Android OS to be developed. The process begins by identifying the collection of raw data from the sensor. This data to be divided into small data samples, each sample as a sequence of 128 values corresponding to the time of collection of 2.56 seconds. Then this sampled data sent for characteristic calculation. Finally, these characteristics to be compared with the set of trained characteristics to identify the action and classify accordingly.

## V. RESULTS

The summary of prediction results of the classification is calculated below for both the algorithms. The number of correct and incorrect predictions are summarized with count values and broken down by each class. Table2.1 gives matrix for SVM and table2.2 for Random Forest. Both the algorithms are able to classify very efficiently with misclassification of very few number of time.

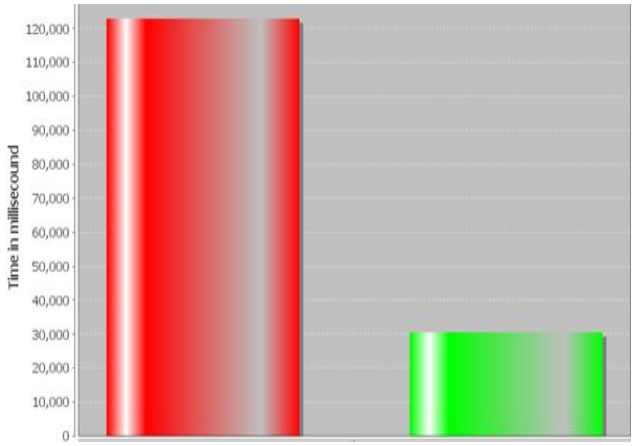
Table 2.1

| RANDOM FOREST | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Standing | Sitting | Laying | Walking | Down stairs | Upstairs | Percentage |
| Standing | 1234 | 0 | 0 | 0 | 0 | 0 | 100 |
| Sitting | 11 | 1275 | 0 | 0 | 0 | 0 | 99.144 |
| Laying | 1 | 0 | 1406 | 0 | 0 | 0 | 99.128 |
| Walking | 0 | 0 | 0 | 1225 | 0 | 1 | 99.918 |
| Downstairs | 0 | 0 | 0 | 5 | 981 | 0 | 99.492 |
| Upstairs | 0 | 0 | 0 | 2 | 4 | 1067 | 99.44 |
| Total | | | | | | | 99.667 |

Table 2.2

| SVM | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Standing | Sitting | Laying | Walking | Down stairs | Upstairs | Percentage |
| Standing | 1360 | 14 | 0 | 0 | 0 | 0 | 98.981 |
| Sitting | 9 | 1277 | 0 | 0 | 0 | 0 | 99.301 |
| Laying | 0 | 0 | 1407 | 0 | 0 | 0 | 100 |
| Walking | 0 | 0 | 0 | 1226 | 0 | 0 | 100 |
| Downstairs | 0 | 0 | 0 | 0 | 986 | 0 | 100 |
| Upstairs | 0 | 0 | 0 | 0 | 0 | 1073 | 100 |
| Total | | | | | | | 99.708 |

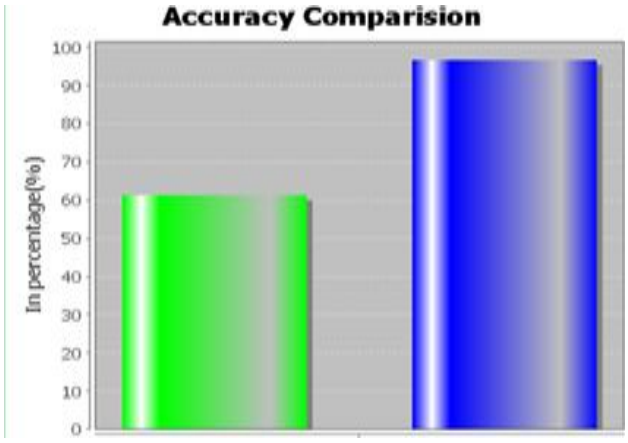
### 1) Time:




Graph1.1

This graph1.1 is a time graph comparison of two algorithms SVM(bar1) and Random Forest(bar2). Therefore, its clearly being able to prove that Random Forest is more time efficient algorithm. SVM takes approximately 120seconds to classify whereas Random Forest is able to do the same task with rather more accuracy in 30seconds. Therefore, proving Random Forest algorithm 1.5times faster than SVM. SVM doesn't perform well, when the dataset is large because the required training time is higher. This comparison may not make a significant difference when the data is limited but is a very important factor when dataset is huge i.e. when real time data is considered which is the case most of the times in real world.

**1) Accuracy:**




Grap.1.2

This graph1.2 shows the comparison of accuracy. Random Forest(bar2) proves to be more accurate than SVM(bar1). SVM showed an accuracy of 61.3% whereas Random Forest showed an accuracy of 96.7%. SVM doesn't perform very well, when the data set has more noise i.e. target classes are overlapping. SVM doesn't directly provide probability estimates, these are calculated using an expensive five-fold cross-validation. The overfitting problem never comes when we use the random forest algorithm therefore making it more accurate.

## VI. CONCLUSION AND FUTURE WORK

In this paper, a system has been developed to recognize the human activities which are necessary in life for betterment of lifestyle, healthcare and fitness. Inbuilt sensors in smartphones were used. The experiments showed Random Forest to be a more efficient multi classification algorithm than SVM in terms of time and accuracy. This system can be applied in many fields of practice like health care, sports, lifestyle. However, the system still has certain restrictions. The recognition is limited to certain action. In the future, efforts can be made to include more and more activities for recognition.

## REFERENCES

[1] Tran, Duc Ngoc, and Duy Dinh Phan. "Human Activities Recognition in Android Smartphone Using Support Vector Machine." *Intelligent Systems, Modelling and Simulation (ISMS), 2016 7th International Conference on*. IEEE, 2016.

[2] Cheng, Long, et al. "Recognition of human activities using machine learning methods with wearable sensors." *Computing and Communication Workshop and Conference (CCWC), 2017 IEEE 7th Annual*. IEEE, 2017.

[3] Sztyler Timo. "Towards real world activity recognition from wearable devices." *Pervasive Computing and Communications Workshops (Per Com Workshops), 2017 IEEE International Conference on*. IEEE, 2017.

[4] Yazdansepas, Delaram, et al. "A multi-featured approach for wearable sensor-based human activity recognition." *Healthcare Informatics (ICHI), 2016 IEEE International Conference on*. IEEE, 2016.

[5] Irene, S., et al. "Development of ZigBee Tri axial Accelerometer Based Human Activity Recognition System." *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on*. IEEE, 2015.

[6] McIlwraith, Douglas, Julien Pansiot, and Guang-Zhong Yang. "Wearable and ambient sensor fusion for the characterisation of human motion." *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010.

[7] Mojidra, Hemali S., and Viral H. Borisagar. "A literature survey on human activity recognition via hidden markov model." *IJCA Proceedings on International Conference on Recent Trends in Information Technology and Computer Science*. 2012.

[8] Pansiot, Julien, et al. "Ambient and wearable sensor fusion for activity recognition in healthcare monitoring systems." *4th international workshop on wearable and implantable body sensor networks (BSN 2007)*. Springer Berlin Heidelberg, 2007.

[9] Ravi, Daniele, et al. "Deep learning for human activity recognition: A resource efficient implementation on low-power devices." *Wearable and Implantable Body Sensor Networks (BSN), 2016 IEEE 13th International Conference on*. IEEE, 2016.

[10] Hsieh, Shang-Lin, et al. "A wrist-worn fall detection system using accelerometers and gyroscopes." *Networking, Sensing and Control (ICNSC), 2014 IEEE 11th International Conference on*. IEEE, 2014.