

# Human Activity Recognition using Smart Phone Sensors

Vinish Kumar, Anuj Sharma

**Abstract:** *In the new age technology, there exists many smart devices, which are using human activity data in reshaping the modern world dynamics of every aspect of our life be it health trackers, smartphones, intelligent systems. One futuristic concept is the connected devices that are way more efficient, adaptive, responsive and flexible to any conditions and reacts according to the data. For some connected devices to work more efficiently, human activity data is required. This data can be used to make devices smarter and using it can be useful in solving many problems of healthcare, efficient surveillance. Our work is an effort in efficient surveillance and using deep learning models, we detect the presence of human activities in different environments and use the data to analyze better to have efficient and effective surveillance. Many different models of deep learning model are used in our work from the likes of CNN (Convolutional Neural Networks) to LSTM (Long Short-Term Memory Networks). The data collected is from sensors' data which is present in the mobile and can make the predictions about various activities like sitting, walking, jumping and some other human activities. The prime focus here is to detect various canonical activities that are not given to the system.*

**Index Terms:** *Ubiquitous Sensing, Machine Learning, CNN, Human Activity Recognition, Tensor flow, Deep-Learning.*

## I. INTRODUCTION

Machine learning is the process of solving learning explicit real-life problems that cannot be solved by usual algorithmic approach. It basically involves the learning process of the system and then testing of the system for different sets of the problems. The problem ranges from identifying images as objects to detecting of the human activities. The learning process steps in with observations or real-life data experience in order to look for patterns in data and take effective decisions in the future perspective based on the example. The main objective is to allow systems to learn automatically without any human interference.

For this paper we have studied a collection of research papers which explains about various different methods for detecting human activities. The concepts of machine learning and deep learning are extensively used for human activity recognition. Activity recognition is the simple process by analyzing the people's behavior and their surroundings. For example: action recognition in the case of elderly people, infants or chronically ill patients. There are many research challenges associated with such cases. Therefore accurate, efficient and reliable activity detection is the need of the

society. Thus, it has the potential for improving the available activity recognition systems for such user groups. During the past decade, many experimental validations based on microelectronics and computer systems have been done to achieve the accurate action recognition. Uses of wearable sensors or mobile phone sensors have been extensively used. Their high computational power feature, compact size, and cost in control allow users to interact with the devices as part of their daily living. This became the evolution of *Ubiquitous Sensing* with main motive of extracting knowledge from the data gathered by various sensors.

Proposed work uses WISDM (Wireless Sensor Data Mining) data set which is CSV file and contains a statistical data and studies several examples and uses a sum total of six properties. All the properties are distributed into different types namely walking, jogging, upstairs, downstairs, sitting and standing [12].

## II. SOME MACHINE LEARNING METHODS

Machine learning algorithms divided into two ways supervised or unsupervised.

### A. Supervised machine learning algorithms

Supervised machine learning algorithms makes use from the knowledge of the past to make decisions on the new data using labelled examples to predict upcoming events. The learning algorithm forms a deduced function to make predictions on output values by proper examination of the dataset. After sufficient training, the model is able to provide targets for any new input.

### B. Unsupervised machine learning algorithms

Unsupervised machine learning algorithm uses neither labeling nor classification approach. It also studies about how the systems could determine a function to explain a hidden structure from unlabeled data. The model does not predict out the correct output, but decides the data and infer results from datasets so as to explain the hidden construction from unlabeled data.

### C. Semi-supervised machine learning algorithms

The semi-supervised machine learning algorithms use both concept of labeled and unlabelled data for training – typically a small amount of labeled data and a large amount of unlabeled data. Therefore they are sandwiched between supervised and unsupervised learning algorithms.

**Revised Manuscript Received on July 5, 2019.**

**Vinish Kumar**, Computer Science and Engineering, Institute of Technology, Roorkee, India, vinishk18@gmail.com

**Anuj Sharma**, Computer Science and Engineering Institute of Technology, Roorkee, India

In research it observed that the systems that implements semi supervised techniques are able to achieve improved learning accuracy. Usually, semi-supervised learning is selected when the labeled data requires skilled and related/relevant resources in order to learn from it. Otherwise, acquiring unlabelled data normally doesn't need additional resources.

#### **D. Reinforcement machine learning algorithms**

These algorithms exploit the learning technique that senses its environment by producing actions and determines errors or rewards. Trial, error search and delayed reward are the most important features of reinforcement learning. To maximize the performance of the system it determines the ideal behavior within a specific context. The reinforcement signal is nothing but simple reward feedback through which the system learns about which action is the best.

### **III. HYPOTHESIS AND PREDICTIONS**

The objective of this paper is to develop a model that is capable of recognizing daily activities of a human being under real-world conditions, using data collected by a single triaxial accelerometer built into an android based handheld device. A supervised classification technique was used to formulate the human activity recognition. During our experimental validation certain limitation are considered to achieve the said objective as: First, few activities like slow versus fast walking, aerobic dancing have not been studied. Second, minimal sensory input considered than existing research of the field, yet able to obtain a comparable accuracy. And, a robust classification methodology regardless of the classifier tools in use. The proposed model has been tested in an experiment with four users each performing the six physical activities: Walking (slow and fast patter), Running, Stairs-up and down, and Aerobic Dancing. Following sections introduce about the related work; data collection methodology and approach to detect and recognize activity from accelerometer data, and results of experimental validations.

Another approach toward sensing the human activity and recognition is based on RFID tag technique. Study of the field shows that there are three approaches to human-activity detection are being used: computer vision, sensory beacons, and the RFID. Computer vision involves many challenges like camera aperture, environmental factor, noise and the outfits of the human being. Active sensor beacon provides relatively accurate action identification, but it requires a consistent power source, making them impractical for long-term dense deployment. RFID tags have also proven the similar object-identification accuracy as active sensor beacons including advantage of being battery-free; however, it fails in detecting motion. Vision based human activity analysis has a great impact in its promising applications such as smart video surveillance, video event retrieval, and human-computer interaction (HCI). Vision based action recognition and localization is a daunting task due to the various constraints as research challenges like individual posture, gate, human motion or camera motion, view angle, clothing, illumination variations, occlusions (fully, partially, or self), and the

complex and cluttered background. The processing system consists of three different layers: body-part segmentation, pose and gesture layer, and the action and interaction layer. The first layer which is the lowest level deals with statistical information processing to calculate body-part extraction. The next layer deals with the temporal frame processing and using statistical tools to estimate the pose and the gesture of the segmented body parts. Gesture is the predicted by the continuous sub-sequence of the blob positions. The last layer deals with the detection and recognition of human activity. Human activities are identified by finding a temporal correlation that fulfils the necessary condition. Many spatial techniques are used for the segmentation. Bayesian networks and Hidden Markov models (HMMs) are used for posing and gesture layer respectively. At the highest layer, human actions and their interactions are represented by various models few of them have significant exposure as, semantic tool as a context-free grammar (CFG), the spatio-temporal long short-term memory (STLSTM) network which is an extension of the traditional LSTM-based learning to two concurrent domains (temporal and spatial domains) [10].

Human activity detection based on wearable sensor data such as acceleration signals, environment signals, location data, and physiological signals. There are two stages of human activity recognition i.e., training and testing of various actions. The training part requires a temporal dataset of measured attributes from multiple individuals performing each defined activity. The time series are divided into various uniform time quanta and then feature extraction is applied. These extracted features are then supplied to learning methods to portray an activity and recognize them. Testing also follows a similar process. The temporal features are calculated for the test data which is further evaluated with the pre - trained learning model, generating a predicted activity label [4].

Human activity detection in this paper uses a sensor data recorded through an android based handheld device. The device used in this research for detection is via mobile phone which has multiple sensors such as accelerometer, gyroscope and orientation sensors which tells movement and rotation around three axes (X, Y, Z). The machine learning algorithms used in this research are J48, SVM, Bayes theorem and Multilayer perceptron, LSTM and the proposed algorithm is implemented on an android platform [7].

The experiments were made using two datasets: observations through sensor device in hand position and observations through sensor device in pocket. Then random forest method is applied on both the observations that showed the highest accuracy for both the cases. Combination of the three classifiers, presuming the best, using the mean probabilistic method turned out to be the best classifier for activity recognition, outperforming all individual classifiers accurate up to 91.15% [5].

In times of omnipresent computing and artificial intelligence, the sensor-based human activity detection is successful in determining strange human activities. Two techniques were used one being the usage of wearable sensors and other is the propagation of the model using kernel linear regression (KLNLR). Standard methods are inefficient as these methods get a lot of high positive rate as the methods are partial towards usual activities. The approach used here compensates among strange detection rate and false alarm using unsupervised approach [10].

One of the approaches uses context-free-grammar based policy in order to constitute in order to detect and depict various human activities like approach, depart, point, handshakes, hug, punches, kicks and push. Broadly the activities can be categorized into three types namely atomic action, composite action and interaction. The model used is able to depict as well as detect the various actions and interactions with high precision by using a series of images frames that detects movement and motions [11].

#### IV. MEHTODS AND EXPERIMENTAL DESIGN

The methods used to carry out the process of data collection as well as the technology used is explained in the following part step by step where each step is explained comprehensively about the what role it played to carry out the working of the proposed algorithm to meet the objective.

##### A. Proposed method

Figure 1 is the process flow chart of the proposed method that is used in the experiment with their adequate explanations.

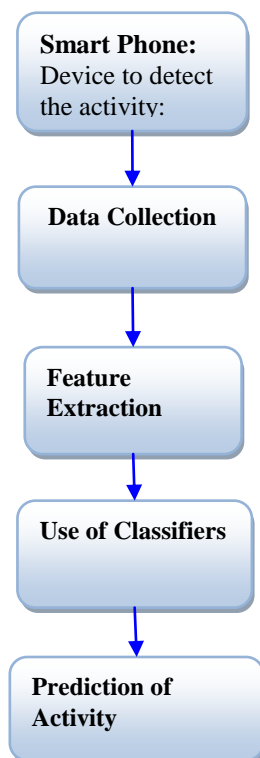


Fig. 1: Flow chart of the methods

##### a. Smart phone based approach

The smart phone was used to detect the activities of human

like walking, running and sitting. Android based smart handheld device application utilizes its embedded sensors by collecting sensor readings and predicting the context.

##### b. Data collection

It is done when a person put the mobile in his/her pocket and performs some activity without caring about the orientation of the mobile phone kept in pocket. Sensor data is recorded in a CSV files. A preprocessing step is applied by trimming the beginning and end of each activity from the data files. It is applied to achieve the increased classification accuracy. After every 20 value sensor data is collected again.

##### c. Feature Extraction

X, Y, Z axes are regarded as three individual features for every individual sensor. Monitoring of four sensors is done along with five-time domains features for each axis, 60 features are extracted for each instance in total.

##### d. Classifiers

Few supervised learning algorithms are used:

- *J48*: The resulting outcome of J48 algorithm is nothing but a decision tree, here entropy is calculated and finally information gain is calculated.
- *SVM*: Since the models contains more than one class the SVM is used to create a hyperplane which separates two classed but to separates multi class it uses strategies like, one-against-one and one-against-all.
- *Bayes Theorem*: It is used to calculate the probability of the features that will be extracted for input and it is used to because it consumes lesser time to train and classify.
- *Multilayer Perceptron*: It is a feed forward network which uses three of more layers and it is trained through back propagation where weights are updated according to error between observed output and the expected output.
- *LSTM*: Used to store long term memory for storing. LSTM has additional memory units for storing the sequence of data. LSTM are a specialized version of RNN which is able to handle long term dependencies. These networks form a chain like structure and the basic idea behind LSTM is being a horizontal line running through the whole network. A convolution neural net layer based gate controls the addition or removal of information to a specific cell state.

##### e. Context Free Grammar

Context free grammar has been used in order to detect and depict various human activities to classify into 3 types.

The position of the device whole time was free, able to tilt lean forward, backward, diagonally with accordance with the body of the subject.



### B. Human detection and activity classification based on micro-Doppler signatures using DCNN

As a supervised learning algorithm, DCNN attempts to learn by mapping between the input data and its associated level provided by human annotators. The DCNN has three main components. The convolution filter, exploits the sliding window protocol, is the very first component that works on small local receptive fields of input data. Each filter can be considered as a specific feature detector. The convolution filter output is fed to a nonlinear activation function in the second phase that enables nonlinear transformation to discriminate among classes easily. The RELU function is used as an activation function. The third component is the pooling that reduces the data size. The final prediction can become robust to the translation of input data through pooling. When DCNN is used as a classifier all input nodes of the layer are connected to all output nodes. The DCNN parameters are trained by back propagation with SGD, to prevent overfitting dropout is widely used a regularization scheme, it randomly omits hidden nodes with a predefined probability that is independent of each training sample.

### C. Human Activity Recognition using two-Stream RNN/CNN in 3D Videos

In this method the Human action recognition is implemented using RNN and CNN structures and by merging the features into a linear support vector machine, which predicts the actions. The implementation combines the RNN, CNN, and Fusion in to one. To compute the decision fusion, first we extract the SoftMax output, and search for the fusion parameters, trust weights  $w_c$  and  $w_r$  for RNN and CNN through validation split. For feature fusion, RNN features are extracted (from the fully-connected layer, and CNN features from the fully connected layer. Now these two features are concatenated into one feature array to get the fusion and apply L2 normalization on to it.

### D. Significance of work

Human activity recognition is a field of study concerned with identifying the movement of person based on the sensor data. Movements are often typical activities performed indoors such walking, running, sitting, lying, climbing stairs etc. smart phones and other personal tracking devices used for fitness and health monitoring are cheap and ubiquitous. As such, sensor data from these devices is cheaper to collect, more common, and therefore is a more commonly studied version of the general activity recognition problem. Recognition methods can be applied to detect a terrorist activity at a certain place by providing the individuals with a sensor device and by detecting any suspicious activity based on the sensor movements. Also, the health related problems such as heart attack can be detected by measuring physiological sensors.

Deep neural network models have started delivering on

their promises of feature learning and are achieving state-of-the-art results for human activity recognition. They are capable of performing automatic feature learning from the raw sensor data and out-perform models fit on hand-crafted domain-specific features.

Our work aims to recognize human activities that can further help in surveillance systems and monitoring strange activities for patients that need constant watch.

## V. STRUCTURE OF HUMAN ACTIVITY RECOGNITION

### A. Two layered MEMM model

The below model depicts how different layer works or operates and its effects on the system.

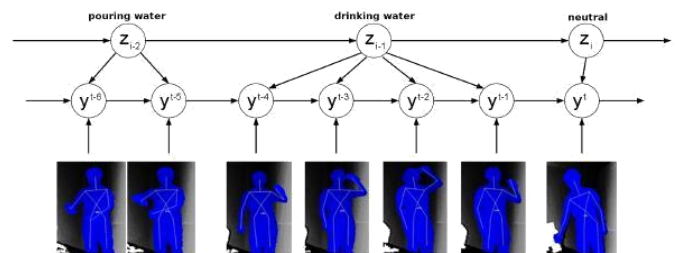


Fig 2: Two layered MEMM model

### B. Basic working of Hough Voting for Activity detection and recognition

The below model depicts the working of training and testing data, how the layer takes input from the training data, then train the data based on the features and then test it using a sample from the testing data.

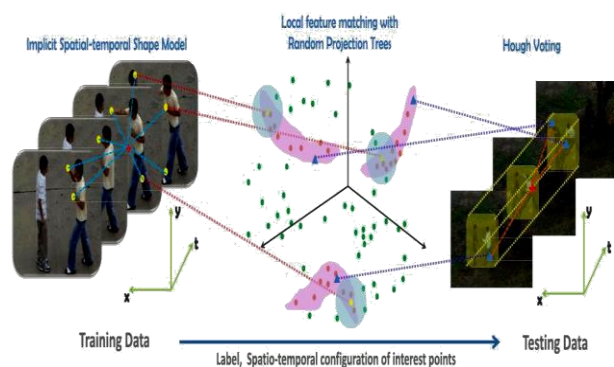


Fig 3: Basic working of Hough Voting for Activity detection and recognition

## VI. EXPERIMENTAL VALIDATION AND RESULT

### A. Algorithms used in experiment.

There are number of algorithms used in the experiment to find out which algorithm suits best to give highest accuracy to detect human activities below is the result and accuracy of the algorithms used.

## B. Testing of the implementation module

These are the different results after the training of the HAR dataset on the different parameters like different learning rate of the model using convolution neural networks. Varying training epochs and learning rates are used for achieving the better accuracy.

**Table 1:** Overview of the algorithms used

Algorithms	Accuracy Observed		Activities detected
CNN,RNN[6]	Cross subject	Cross view	Activities recognized from optical data
	83.74 %	93.65 %	
Naïve-Bayes Classifier [2]	84.7%		Tests on seen and unseen human activity
Combined-SVM Multilayer-perceptron Logit Boost [5]	94.15%		Supervised learning for classification of activities and accelerometer data
Decision Tree C4.5[4]	92.6%		Human activity recognition using android platform
KNN [3]	70%		Used to recognize activities like shake, hug, kick, point, punch, push
Hough Forest [3]	77.7%		Used to recognize actions like shake, hug, kick, point, punch, push

**Table 2:** Confusion matrix of the predicted labels

<div> <div>Predicted Label</div> <div>True Label</div> </div>	Walking	Walking Upstairs	Walking Downstairs	Sitting	Standing	Laying
Walking	446	2	26	0	2	0
Walking Upstairs	5	441	25	0	0	0
Walking Downstairs	1	0	419	0	0	0
Sitting	1	1	0	39 6	87	6
Standing	2	1	0	87	442	0
Laying	0	0	0	0	0	537

**Table 3:** Testing of the implementation module

Dropout Probability	Learning Rate	Training Epoch	Training accuracy	Testing Accuracy
0.7	0	10	0.617	0.577
0.7	0.1	100	0.691	0.182
0.8	0	10	0.682	0.628
0.8	0.1	100	0.791	0.182
0.9	0	10	0.767	0.711
0.9	0.1	100	0.791	0.182

## VII. CONCLUSION

The concepts of machine learning, and deep learning are extensively used for human activity recognition. Activity recognition is the simple process by analyzing the people's behavior and their surroundings. For example: action recognition in the case of elderly people, infants or chronically ill patients. There are many research challenges associated with such cases. Therefore accurate, efficient and reliable activity detection is the need of the society. During our experimental validation of the human activity recognition using various existing machine learning algorithms and our proposed method it is observed that (with certain limitations like the sensors displacement and its positioning, environmental conditions, etc) CNN, RNN with cross attains the highest accuracy above 93% while decision tree gives around 93% , and our proposed algorithms attains the 94.15% accuracy.

## VIII. FUTURE WORK

In future, we plan to implement motion based human activity recognition using a visual or image data with the help of open-cv and the required modules. Local spatio-temporal features can be used for representing and recognizing motion patterns such as human actions. The other physiological signals such as GPS can also be used to track the human activity of the user based on the location of the user, for example if he is in a restaurant, he must be eating or drinking, so actions can be detected combined with some other factors can be implemented.

## ACKNOWLEDGEMENT

We pay our deep sense of gratitude to Mr. Santosh Kumar Verma, Assistant Professor - CSE, IET, JK Lakshmiipat University, Jaipur, Rajasthan, India for his generous guidance, help and useful suggestions. He was extremely supportive and helped us, his insightful comments and constructive suggestions helped us improving the quality of this work.

## REFERENCES

1. Wilde, Adriana G. "An overview of human activity detection technologies for pervasive systems." Department of Informatics University of Fribourg, Switzerland 212 (2010): 72-112.
2. Sung, Jaeyong, et al. "Unstructured human activity detection from rgbd images." Robotics and Automation (ICRA), 2012 International Conference on. IEEE, 2012.
3. Yu, Gang, Junsong Yuan, and Zicheng Liu. "Propagative hough voting for human activity detection and recognition." IEEE Transactions on Circuits and Systems for Video Technology 25.1 (2015): 87-98.
4. Lara, Oscar D., and Miguel A. Labrador. "A survey on human activity recognition using wearable sensors." IEEE Communications Surveys and Tutorials 15.3 (2013): 1192-1209.
5. Bayat, Akram, Marc Pomplun, and Duc A. Tran. "A study on human activity recognition using accelerometer data from smartphones." Procedia Computer Science 34 (2014): 450-457.

6. Zhao, Rui, Haider Ali, and Patrick van der Smagt. "Two-stream RNN/CNN for action recognition in 3D videos." Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on. IEEE, 2017.
7. Yin, Xizhe, et al. "Human activity detection based on multiple smart phone sensors and machine learning algorithms." Computer Supported Cooperative Work in Design (CSCWD), 2015 IEEE 19th International Conference on. IEEE, 2015.
8. Kim, Youngwook, and Taesup Moon. "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks." IEEE geoscience and remote sensing letters 13.1 (2016): 8-12.
9. Liu, Jun, et al. "Spatio-temporal lstm with trust gates for 3d human action recognition." European Conference on Computer Vision. Springer, Cham, 2016.
10. Yin, Jie, Qiang Yang, and Jeffrey Junfeng Pan. "Sensor-based abnormal human-activity detection." IEEE Transactions on Knowledge & Data Engineering 8 (2007): 1082-1090.
11. Ryoo, Michael S., and Jake K. Aggarwal. "Recognition of composite human activities through context-free grammar based representation." CVPR, IEEE computer society conference on. Vol. 2. IEEE, 2006.
12. Web-link of the Data-Set used in this paper  
"http://www.cis.fordham.edu/wisdm/dataset.php"

### AUTHORS PROFILE



**Vinish Kumar** is working as an Associate Professor in Sanskar College of Engineering and Technology, Ghaziabad, India. He is pursuing his Masters degree from ITR Roorkee, India. Personal profile which contains their education details, He has several national/international research articles in his credit.



**Anuj Sharma** Dr. Sharma is Director and Professor in Institute of Technology Roorkee, Roorkee, India. He has several national/international research articles in his credit.