ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Machine Learning Model for Human Activity Recognition by Smartphone Dataset

Khushbu Dewangan [1], Abhishek Kumar Dewangan [2]

M.Tech. Student, Department Of Computer Science And Engineering, Shri Shankaracharya Group Of Institution, Junwani , Bhilai,Chhattisgarh, India [1]

Assistant Professor, Department Of Computer Science And Engineering, Shri Shankaracharya Group Of Institution , Junwani , Bhilai,Chhattisgarh, India [2]

ABSTRACT: Cell phones are complimenting more refined with each new model release. These days, cell phones ordinarily consolidate numerous differing and persuasive sensors for instance GPS, mouthpieces, light sensors, temperature sensors, attractive compasses, whirligigs, high-determination cameras and accelerometers. Human movement acknowledgment (HAR System) has enabled imaginative applications in disparate locales, for example, medicinal services, diversion and security. There are to approaches of HAR (i) Online Processing (ii) offline Processing. Our research emphasized on offline processing, in which Smartphone dataset has been used for Human activity prediction. In this paper we have proposed the use of PCA (Principle Component Analysis) algorithm for dimension reduction of input HAR dataset, further Cubic SVM (Support Vector Machine) has been as classifier for proposed machine learning model.

KEYWORDS: HAR;SVM;PCA

I. INTRODUCTION

A Human Activity Recognition (HAR) system can automatically recognize physical activities, which is a key research issue in mobile and ubiquitous computing. An HAR system performs tasks of recognizing different human daily activities from simple to complex. The sensors involved in an HAR system can be video sensors, inertia sensors, and environment sensors. The GPS receiver can also be used for activity recognition but is limited to outdoor environments.

Depend on the putting in of sensors, HAR systems can be separated into three categories: wearable devices based sensing systems, smartphone sensing systems, and smart living environments. Although wearable devices and smart living environments can deliver good activity detection results, smartphone based applications are an increasingly prominent solution as smartphones have become an indispensable part of our daily life. Especially with the rapid evolution of hardware, ever-increasing computing and networking capacity, and rich embedded sensors, smartphone based HAR systems can tell us different kinds of human activities in real time using machine learning techniques. In addition, using smartphones for human activity recognition has a wide range of applications including healthcare, daily fitness recording, anomalous situation alerting, personal biometric signature identification, and indoor localization and navigation. All this benefits from the fast development of mobile phone software and hardware.

In the Android environments, the most commonly used and installed sensors can be categorized as follows [Google, “Android API description”]:

- Motion sensors: the motion sensors are based on inertial force.
- Environmental sensors: these sensors measure environmental parameters, like temperature and pressure, using barometers or thermometers.
- Position sensors: these sensors include orientation sensors and magnetometers, measuring the physical position of the device.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

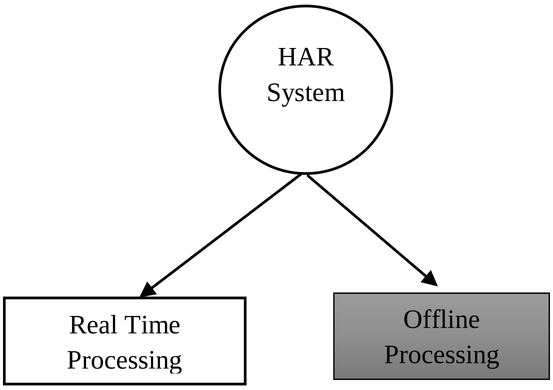

Fig. 1- HAR System : Functions as online(Real time) as well as offline.

Though there are previously several applications in the arcade as demonstrated in the introduction section, most of them do not fully utilize the smartphone embedded inertial sensors. The granularity of such applications is also not adequate. In some applications, only the events of walking and motionless are documented. Certain apps that only use GPS signals fail to function in indoor environments. The adequacy of utilizing machine learning systems on cell phone based sensor information is set apart to be researched, with the motivation behind perceiving human exercises. Distinctive space components and information handling systems should be contemplated.

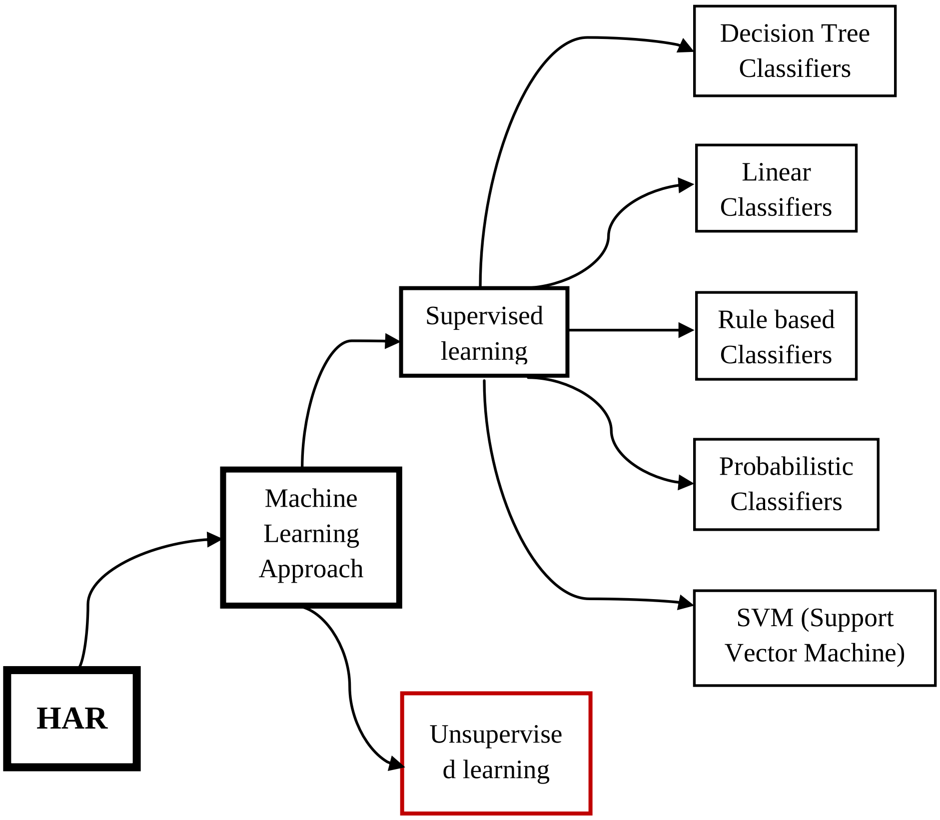

Fig.-2 Classification of HAR based on learning processes, supervised and unsupervised learning.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

II. LITERATURE_SURVEY

Subhas Chandra Mukhopadhyay [IEEE 2015] has reviewed the reported literature on wearable sensors and devices for monitoring human activities. The human activity monitoring is a vibrant area of research and a lot of commercial development are reported. It is expected that many more light-weight, high-performance wearable devices will be available for monitoring a wide range of activities. The challenges faced by the current design will also be addressed in future devices. The development of light-weight physiological sensors will lead to comfortable wearable devices to monitor different ranges of activities of inhabitants. Formal and Informal survey predicts an increase of interest and consequent usages of wearable devices in near future, the cost of the devices is also expected to fall resulting in of wide application in the society.

Oscar D. Lara and Miguel A. Labrador [IEEE 2013] surveys the state-of-the-art in human activity recognition based on wearable sensors. A two-level taxonomy is introduced that organizes HAR systems according to their response time and learning scheme. Twenty eight systems are qualitatively compared in regards to response time, learning approach, obtrusiveness, flexibility, recognition accuracy, and other important design issues. The fundamentals of feature extraction and machine learning are also included, as they are important components of every HAR system. Finally, various ideas are proposed for future research to extend this field to more realistic and pervasive scenarios.

Charissa Ann Ronaoet. al. [IEEE 2014] shown that a two-stage continuous HMM classifier is a feasible method towards universal activity recognition on smartphones. Continuous HMMs are able to specifically handle time series data such as accelerometer and gyroscope sensor values, and the two-stage architecture enabled us to use a significantly smaller number of features at the same time utilizing the most effective ones. We have also demonstrated that random forest variable importance measures, in combination with proper domain knowledge, is an effective approach in uncovering the most useful features from a large feature set.The proposed method consists of first-level CHMMs for coarse classification, which separates stationary and moving activities, and second-level CHMMs for fine classification, which classifies the data into their corresponding activity classes. Random Forests (RF) variable importance measures are exploited to determine the optimal feature subsets for both coarse and fine classification. Experiments show that with the use of a significantly reduced number of features, the proposed method shows competitive performance in comparison to other classification algorithms, achieving an over-all accuracy of 91.76%.

Table No.1: Literature Survey

S. No.	Author/Paper title/Year	Name of Algorithm/Tool/Method (discussed/Implemented)	Description	Accuracy
1.	BishoySefen, Sebastian Baumbach et. al. /Human Activity Recognition Using Sensor Data of Smartphones And Smartwatches/ ICAART 2016[1]	Naïve Bayes Classifier	In this paper, a platform to combine sensors of smartphones and smartwatches to classify various humanactivities was proposed. It recognizes activities in real-time Moreover, this approach is light-weight, computationally inexpensive, and able to run on handheld devices	87%

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

2.	DavideAnguita et. al. /Energy Efficient Smartphone-Based Activity Recognition using Fixed-Point Arithmetic/ JUCS 2013[5]	SVM,Fixed-Point Arithmetic	Propose a novel energy efficient approach for the recognition of human activities using smartphones as wearable sensing devices, targeting assisted living applications such as remote patient activity monitoring for the disabled and the elderly. The method exploits fixed-point arithmetic to propose a modified multiclass Support Vector Machine (SVM) learning algorithm	89%
3.	Jie Yin et. al./ Sensor-Based Abnormal Human-Activity Detection/ IEEE 2008[6]	Kernel nonlinear regression (KNLR)	Employs a one-class support vector machine (SVM) that is trained on commonly available normal activities, which filters out the activities that have a very high probability of being normal. We then derive abnormal activity models from a general normal model via a kernel nonlinear regression (KNLR)	84%
4.	Charissa Ann Ronao et. Al ./Human Activity Recognition Using Smartphone Sensors With Two-Stage Continuous Hidden Markov Models/ICNC 2014[7]	Two-stage Continuous hidden Markov model (CHMM)	Propose a two-stage continuous hidden Markov model (CHMM) approach for the task of activity recognition using accelerometer and gyroscope sensory data gathered from a smartphone	91.76%

III. PROBLEM IDENTIFICATION

A Human Activity Recognition (HAR) system can automatically recognize physical activities, which is a key research issue in mobile and ubiquitous computing. An HAR system performs tasks of recognizing different human daily activities from simple to complex. The sensors involved in an HAR system can be video sensors, inertia sensors, and environment sensors. After going through different literature we found that there are some bottleneck to develop an efficient HAR system:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

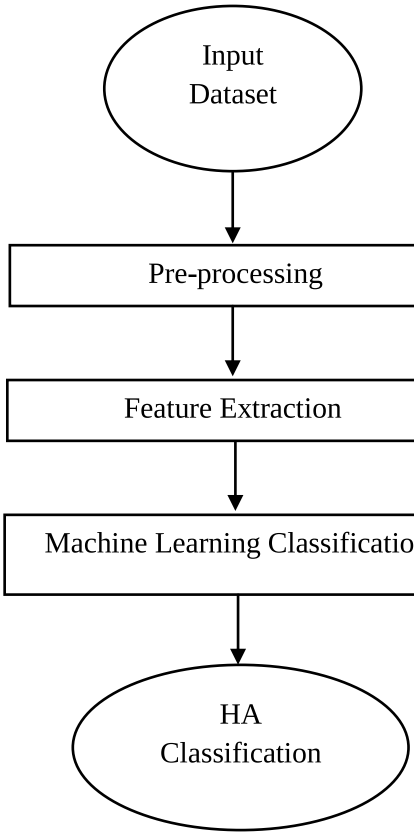

Fig.-3 Generalized Approach of HAR

This starts with homogenous input datasets and gives the best classification by applying different processing strategies sequentially. (Detailed explanation is given in the section proposed methodology)

1. To improve prediction accuracy we need to use efficient classifier.
2. Sensors involved in HAR must be wearable by person hence activity recognition is completely depend on how sensor wore i.e. moving of sensors also affect the prediction accuracy.
3. Online HAR system is not enough capable to recognize because of complex mathematical computation.
4. Historical data we improve the performance hence we need to store training data into device which will reduce the performance of HAR as well as device.
5. Dimension of input data is very high which decreases the performance of system.
6. More efficient feature extraction method can increase the recognition accuracy.
7. Accuracy decreases when test data is collected from same object.
8. The complexity of user activity can bring additional challenge to recognition model.

Henceforth we need an approach which will increase the accuracy of HAR and can improve the device performance, have proposed the HAR system based on machine learning by using multiple regression model we can decrease the mathematical complexity.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

IV. PROPOSED METHODOLOGY

To understand proposed system need to understand generalized approach of HAR (Human Activity Recognition) system, which involve different phases, figure 3 depicts the same.

As from figure 3 pre-processing is the primary phase of HAR system. Pre-processing involves cleaning operation: in this phase consists removing unimportant or disturbing elements for the next phases of analysis and in the normalization of some missing data. Remove redundant data which will decrease the performance of classification. Next phase is feature extraction, the raw sensor data contain fixed-width sliding windows of 2.56sec (128 readings/window). In classification phase we need to choose best classifier, supervised learning classifier, because the input dataset is homogenous dataset.

A. Proposed Algorithm :

Step-1: Load Training data and test data.
//Remove redundant data
Step-2: Create Table variable.
Step-3: Apply pre-processing through PCA (Principle Component analysis).
Step-4: Extract feature from pruned data.
Step-5: Apply Cubic SVM as classifier

B. PCA Algorithm :

Step-1. Select a normalized direction in m-dimensional space along which the variance in X is maximized. Save this vector as p1.
Step-2. Find another direction along which variance is maximized, however, because of the orthonormality condition, restrict the search to all directions orthogonal to all previous selected directions. Save this vector as pi.
Step-3. Repeat this procedure until m vectors are selected.

Principal Component Analysis (PCA): In this technique, variables are transformed into a new set of variables, which are linear combination of original variables. These new set of variables are known as principle components. They are obtained in such a way that first principle component accounts for most of the possible variation of original data after which each succeeding component has the highest possible variance.

C. Cubic SVM (Support Vector Machine) Classification:

Step-1. Extract predictors and response.
Step-2. Processes the data into the right shape for training.
Step-3. Create the result with predict function.
Step-4. Set up holdout validation.
Step-5. Specify all the classifier options and trains the classifier.
Step-6. Compute validation accuracy.

V. RESULT AND DISCUSSION

For implementation of our proposed algorithm we have opted Matlab 2016 further in this section, we will evaluate the effectiveness of our proposed algorithm. Here we have presented an experimental valuation using data. As a source dataset for experimental evaluation we have used HAR dataset. Furthermore we have shown accuracy of proposed algorithm with classificationlearner.

Dataset URL: https://archive.ics.uci.edu/ml/machine-learning- databases/00240/UCI%20HAR%20Dataset.zip

The features selected for this database come from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ. These time domain signals (prefix 't' to denote time) were captured at a constant rate of 50 Hz. Then they

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

were filtered using a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise. Similarly, the acceleration signal was then separated into body and gravity acceleration signals (tBodyAcc-XYZ and tGravityAcc-XYZ) using another low pass Butterworth filter with a corner frequency of 0.3 Hz.
Subsequently, the body linear acceleration and angular velocity were derived in time to obtain Jerk signals (tBodyAccJerk-XYZ and tBodyGyroJerk-XYZ). Also the magnitude of these three-dimensional signals were calculated using the Euclidean norm (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag).

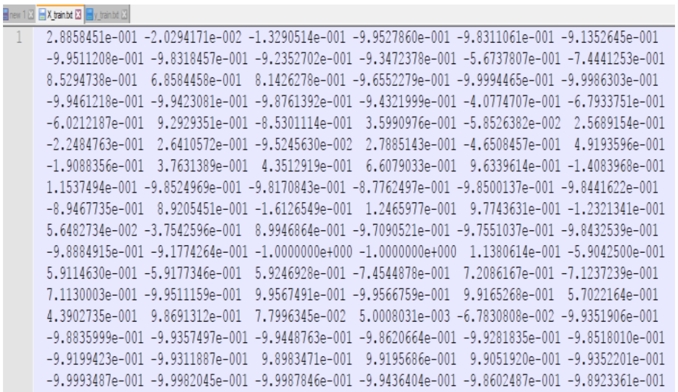
new 1 X_train.txt y_train.txt
1 2.8858451e-001 -2.0294171e-002 -1.3290514e-001 -9.9527860e-001 -9.8311061e-001 -9.1352645e-001
-9.9511208e-001 -9.8318457e-001 -9.2352702e-001 -9.3472378e-001 -5.6737807e-001 -7.4441253e-001
8.5294738e-001 6.8584458e-001 8.1426278e-001 -9.6552279e-001 -9.9994465e-001 -9.9986303e-001
-9.9461218e-001 -9.9423081e-001 -9.8761392e-001 -9.4321999e-001 -4.0774707e-001 -6.7933751e-001
-6.0212187e-001 9.2929351e-001 -8.5301114e-001 3.5990976e-001 -5.8526382e-002 2.5689154e-001
-2.2484763e-001 2.6410572e-001 -9.5245630e-002 2.7885143e-001 -4.6508457e-001 4.9193596e-001
-1.9088356e-001 3.7631389e-001 4.3512919e-001 6.6079033e-001 9.6339614e-001 -1.4083968e-001
1.1537494e-001 -9.8524969e-001 -9.8170843e-001 -8.7762497e-001 -9.8500137e-001 -9.8441622e-001
-8.9467735e-001 8.9205451e-001 -1.6126549e-001 1.2465977e-001 9.7743631e-001 -1.2321341e-001
5.6482734e-002 -3.7542596e-001 8.9946864e-001 -9.7090521e-001 -9.7551037e-001 -9.8432539e-001
-9.8884915e-001 -9.1774264e-001 -1.0000000e+000 -1.0000000e+000 1.1380614e-001 -5.9042500e-001
5.9114630e-001 -5.9177346e-001 5.9246928e-001 -7.4544878e-001 7.2086167e-001 -7.1237239e-001
7.1130003e-001 -9.9511159e-001 9.9567491e-001 -9.9566759e-001 9.9165268e-001 5.7022164e-001
4.3902735e-001 9.8691312e-001 7.7996345e-002 5.0008031e-003 -6.7830808e-002 -9.9351906e-001
-9.8835999e-001 -9.9357497e-001 -9.9448763e-001 -9.8620664e-001 -9.9281835e-001 -9.8518010e-001
-9.9199423e-001 -9.9311887e-001 9.8983471e-001 9.9195686e-001 9.9051920e-001 -9.9352201e-001
-9.9993487e-001 -9.9982045e-001 -9.9987846e-001 -9.9436404e-001 -9.8602487e-001 -9.8923361e-001

Fig. 4 Training Dataset

Figure 4 shows the snippet of training dataset. Dataset description given as above.

Further in dataset finally a Fast Fourier Transform (FFT) was applied to some of these signals producing fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag. (Note the 'f' to indicate frequency domain signals). These signals were used to estimate variables of the feature vector for each pattern: '-XYZ' is used to denote 3-axial signals in the X, Y and Z directions.

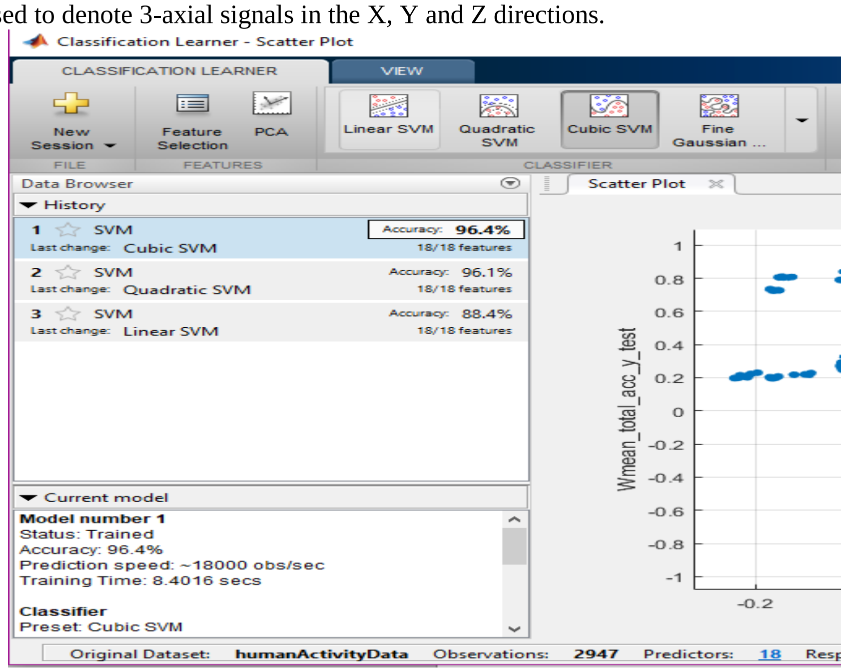

Fig. 5 Classification output of dataset

Table No.2: Accuracy comparison

S. No.	Approach	Accuracy %
1.	Earlier	91
2.	Proposed	96.1

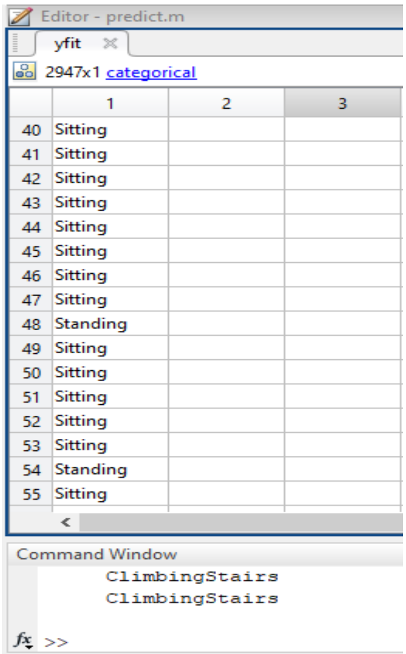

Fig. 6 Output as Predicted Value of HAR: Shows the most appropriate results and store these all data.

VI. CONCLUSION

Our research towards HAR, concludes that by applying better pre-processing algorithm over input dataset, accuracy of HAR system can be improved, from result section we can conclude that accuracy increased.

In future we can work upon time complexity of proposed algorithm, we can also apply our proposed algorithm to different application like personality detection etc..

VII. ACKNOWLWDGEMENT

I want to thank my guide prof. Abhishek Kumar Dewangan and the management of Shri Shankaracharya Group of Institute and Technical Campus, Junwani, Bhilai, India; for their inspiring encouragement and support towards the completion of this result.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

REFERENCES

[1] BishoySefen, Sebastian Baumbach Human Activity Recognition Using Sensor Data of Smartphones and Smartwatches/vol.2: ICAART,488-493, 2016.
[2] ZhinoYousefi/ Human Activity Recognition Using Time Series Classification/ TSpace 2015.
[3] Muhammad Shoaib A Survey of Online Activity Recognition Using Mobile Phones/ Sensors 2015, 15(1),2059-2085.
[4] Oscar D. Lara and Miguel A. Labrador / A Survey on Human Activity Recognition using Wearable Sensors/ IEEE 2013,15(3),1192-1209.
[5] DavideAnguita Energy Efficient Smartphone-Based Activity Recognition using Fixed-Point Arithmetic/ JUCS 2013,19(9),1295-1314.
[6] Jie Yin Sensor-Based Abnormal Human-Activity Detection/ IEEE 2008,20(8),1082-1090.
[7] Charissa Ann Ronaoet. Human Activity Recognition Using Smartphone Sensors With Two-Stage Continuous Hidden Markov Models/ICNC 201410th International conference on natural computation,pp-681-686.
[8] A. J. Perez, M. A. Labrador, and S. J. Barbeau, “G-sense: A scalable architecture for global sensing and monitoring,” IEEE Network, vol. 24, no. 4, pp. 57–64, 2010.
[9] Y. Jia, “Diatetic and exercise therapy against diabetes mellitus,” in Second International Conference on Intelligent Networks and Intelligent Systems, pp. 693–696, 2009.
[10] J. Yin, Q. Yang, and J. Pan, “Sensor-based abnormal human-activity detection,” IEEE Trans. Knowl. Data Eng., vol. 20, no. 8, pp. 1082– 1090, 2008.
[11] O. X. Schlmilch, B. Witzschel, M. Cantor, E. Kahl, R. Mehmke, and C. Runge, “Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring,” Computers in Human Behavior, vol. 15, no. 5, pp. 571–583, 1999.
[12] F. Foerster, M. Smeja, and J. Fahrenberg, “Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring,” Computers in Human Behavior, vol. 15, no. 5, pp. 571–583, 1999.
[13] E. Kim, S. Helal, and D. Cook, “Human activity recognition and pattern discovery,” IEEE Pervasive Computing, vol. 9, no. 1, pp. 48–53, 2010.

BIOGRAPHY

Khushbu Dewangan is a M.Tech. Student [Computer Science And Engineering] whose work is on machine learning model for human activity recognition by smartphone dataset. The principal impact of his work has been on machine learning model.

Abhishek kumar dewangan is currently a Assistant Professor in the department of Computer Science And Engineering at the university At C.S.V.T.U. the state university of (C.G.) India where he also holds adjunct professorship in the department of Computer Science And Engineering . He teaches courses in machine learning model.