RESEARCH ARTICLE OPEN ACCESS

Human Activity Recognition Using Prominent Art-Of-Techniques

Palak Rehan *, Kapil Sharma **
*(Department of Computer Science and Engineering, Guru Nanak Dev Engineering College, Ludhiana, India
** (Department of Computer Science and Engineering, Guru Nanak Dev Engineering College, Ludhiana, India

ABSTRACT
Human activity recognition (HAR) system has a remarkable contribution in developing smart home systems. Thus it is one of the prominent research areas for ubiquitous computing. HAR systems are based on three important aspects .Foremost aspect is to capture human behaviors through sensors. Second and most crucial aspect is extraction of relevant features from data captured via sensors. Last task is classification to recognize which human activity has taken place. Each aspect of HAR poses challenges which are yet to be solved. This paper enlightens state-of-art methods and research work carried in this field by carrying a literature survey. This paper describes each of prominent research models by highlighting their methodology, uniqueness and research gaps. Furthermore this review represents comparative study and result findings of discussed existing techniques.
***Keywords* -** Artificial Intelligence, Deep learning, Human Activity Recognition, Machine Learning, Neural Network

---

Date of Submission: 06-07-2020 Date of Acceptance: 21-07-2020

I. INTRODUCTION

Human activity recognition (HAR) is the process of automatic identification of human activity captured by the sensor. There is wide range of human activities from static postures (like standing or sitting) to steady transitions (like from sit to stand or stand to sit) and to repetitive activities (like running, jogging). Recognition of these human activities is very much vital to develop smart home devices or IOT devices which can predict events based on them. For instance, HAR can be applied on bank surveillance data to detect burglary and other crime events. . HAR is based on three general steps: capturing of human activities by sensors, preprocessing of sensor signals followed by extraction of relevant features and classification so as to identify which human activity has taken place.

Research has been carried out using various data capturing devices like video based sensors, wearable body sensors, wireless sensors and mobile phone sensors [1-16]. Out of them, wearable and mobile phone sensors are prominently used due to its less cost, easiness to use and mobility to carry anywhere. Wearable body sensors can be kept at different body locations (like ankle, wrist, arm, hip, thigh) chosen by researcher [1-3,9]. Trending research has been carried out using signals from smart phones [9, 11-15].

Generally, signals from accelerometer, gyroscope and magnetometer sensors have been processed by HAR researchers because these sensors are capable of capturing most relevant human information [4, 5]. Signal pre-processing is performed to reduce noise and spikes in signal. It is also desirable to filter out informative components from signals. For example, there are two components (gravity and motion component) in acceleration signals. Gravity component is responsible for distinguishing static postures and motion signals are useful for recognizing transitional and repetitive actions. After signal processing, feature extraction phase comes into picture. For motion signals, Statistical features (mean, power, standard deviation, entropy, correlation etc) are computed because of their high interclass variability [1,2,6,7,8]. Time domain features (like mean, acceleration, power, standard deviation etc) and frequency domain features (like location of peaks, DFT coefficients etc) are mostly extracted by Smartphone HAR researchers.[9, 12]

For classification which is the last stage of HAR, there are plenty of classifiers. Studies have revealed that choice of classifier is based on type of activity recognized, for instance decision trees classifiers are popularly used for classifying activity involving repetitive movements(like walking) [1,2,4,6,15] and type of sensors used, for instance , deep learning techniques are currently used to recognize recorded by Smartphone sensors[11,16]. There are three broad categories for HAR classification which is discussed in next section of this survey.

II. STATE-OF-THE-ART TECHNIQUES

HAR systems are broadly classified into three categories.

Category 1: HAR based on Classical Machine Learning methods

Category 2: HAR based on neural network methods.

Category 3: HAR based on Deep learning models.

There are numerous machine learning (ML) methods that have been used to design HAR system. Most commonly experimented techniques in this category are SVM, decision-trees, hidden Markov model and K-nearest neighbour. Studies shows that HAR based on neural network and deep learning models outperform than those based on classical ML methods. Approaches based on Artificial Neural network, Convolutional Neural network and recurrent neural networks falls under category 2. Category 3 is gaining lot of popularity in past few years due to its automatic feature extraction ability. Deep Belief network, Restricted Boltzmann machine and deep Boltzmann machine belongs to this category. Below, we describe survey of prominent proposed HRM system for each category.

In [11], author proposed a model based on category 3 approach.

<table>
<tr><td>Paper</td><td>Author: Mohammed Mehedi Hassan et al[11]
Year of publication: 2018</td></tr>
<tr><td>Dataset</td><td>Public available dataset consisting of 10929 human events. Out of it, 7767 and 3162 events are used for training and testing phase respectively.</td></tr>
<tr><td>Activities</td><td>12 basic human activities like standing, sitting, lying down, stand to sit etc.</td></tr>
<tr><td>Input data detail</td><td>Sensor device : Smartphone
Sensors selected for capturing human events: triaxial accelerometer and gyroscopes.
Recording sampling frequency : 50Hz
Signals considered for processing : Triaxial angular velocity and linear acceleration signal sensed from smartphone gyroscope and accelerometer sensors respectively.</td></tr>
<tr><td>Signal preprocessing</td><td>Median and low-pass butter-worth filter with 20Hz cutoff frequency are applied to both signals to reduce noise. A low-pass Butter-worth filter is again applied to acceleration signal so as to filter out body acceleration and gravity information.</td></tr>
<tr><td>Features Extracted</td><td>Mean, median, correlation, signal magnitude area and various AR coefficients were extracted. Further processing of extracted signal is performed by kernel principal component analysis and linear discriminant analysis.</td></tr>
<tr><td>Recognition model</td><td>This phase consist of two part, pre-training and fine-tuning. Pre- training was based on Restricted Boltzmann Machine with 2 hidden layers, 100 neurons
In input layer, 60 in first hidden layer, 20 in second hidden layer and 12 in output layer.</td></tr>
<tr><td>Experiment results</td><td>System accuracy is 95.85% and mean recognition rate is 89.61%
<table>
<tr><th>Activity</th><th>Recognition rate</th><th>Mean</th></tr>
<tr><td>Standing</td><td>99.60%</td><td rowspan="12">89.61%</td></tr>
<tr><td>Sitting</td><td>95.97</td></tr>
<tr><td>Lying down</td><td>96.67</td></tr>
<tr><td>Walking</td><td>93.50</td></tr>
<tr><td>Walking-upstairs</td><td>97.12</td></tr>
<tr><td>Walking-downstairs</td><td>99.45</td></tr>
<tr><td>Stand-to-Sit</td><td>82.61</td></tr>
<tr><td>Sit-to-Stand</td><td>90.00</td></tr>
<tr><td>Sit-to-Lie</td><td>81.25</td></tr>
<tr><td>Lie-to-Sit</td><td>72.00</td></tr>
<tr><td>Stand-to-Lie</td><td>85.71</td></tr>
<tr><td>Lie-to-Stand</td><td>81.48</td></tr>
</table></td></tr>
</table>

Future Scope	This proposed system recognizes only basic activities which can be extended to real time environments involving complex activities.

In reference [12], author has experimented using category 2 model.

Paper	Author: Niloy Sikder et al[12] Year of publication: 2019
Dataset	UCI HAR dataset which is a standard dataset hosted by the University of California Irvine (UCI). Out of 10,299 samples, 7352 samples for training and 2,947 samples for testing are deployed.
Activities	6 human activities consisting of walking, walking upstairs, walking downstairs, sitting, standing and laying.
Input data detail	Sensor device : wrist-mounted Smartphone Sensors selected for capturing human events: accelerometer and gyroscopes. sampling frequency : 50Hz Signals considered for processing : Three axial angular velocity and three axial linear acceleration
Signal preprocessing	Median filter and third order low pass Butter-worth filter with 20hz cut off frequency are applied for noise reduction.
Features Extracted	Frequency features and Power features are extracted using well known Fast Fourier Transform(FFT) algorithm and Pwelch algorithm respectively
Recognition model	Finally, for classification, two layered Convolutional network is employed, one each for processing of frequency and power signals simultaneously. Convolution layers consist of 2*2 ,Relu Kernel. Two channel CNN are subjected to max pooling and at last the dense layer outputs of the channel are concatenated to acquire the final classification output.
Experiment results	System accuracy is 95.25% with precision of 95.32% and accuracy of 95.16
Future Scope	This model face some challenges in classification of sitting activities and 12% of them are wrongly classified as standing activities. Chosen dataset does not contain activities with postural transitions. So the above model can be explored to different datasets with wide range of motion. System performance can be improved by feeding new features to CNN channels and by increasing layers and nodes in it.

Activity	**Accuracy (%)**
WLk	97.38
WUp	94.90
WDn	95.48
Stn	96.24
Lay	99.81
Total	**95.25**
Precision(%)	95.32
Recall(%)	95.16
F1-score(%)	95.24

In reference [13], author has proposed an innovative HAR system by combining category 2 and category 3 techniques.

Paper	Author: Ran Zhu et al[13] Year of publication: 2019
Dataset	Researcher has created dataset by capturing activities of 100 participants ageing between 12 to 51 years.

<table>
<tr><td>Activities</td><td>Seven human activities consisting of Going Upstairs (GU), Going Downstairs (GD), Standing (SD), Running (RU), Walking (WK), Bicycling (BY) and Swinging (SW)</td></tr>
<tr><td>Input data detail</td><td>Sensor device : Smartphone
Sensors selected for capturing human events: accelerometer and gyroscopes and magnetometer
sampling frequency : 50Hz
Signals considered for processing: nine axis motion signals of selected sensors.</td></tr>
<tr><td>Signal preprocessing</td><td>Raw signals are directly used except simple processing. Simple processing involves sliding window segmentation with 50 seconds of step size to each sensor data. This resulted into 235977 samples each of size 200(data of four seconds)*3(for each of three sensor)*3(for each x, y, z axis).</td></tr>
<tr><td>Methodology</td><td>Two CNN models namely CNN-7(seven class network) and CNN-2(Two class network) are used for training and weighted voting of these two CNN models is considered in testing phase to recognize a human activity.
Proposed CNN architecture consists of five kinds of layers
• Input layer of input sample size
• 2 Convolutional layers with 64 and 32 kernels respectively for feature extraction from input data.
• Max-pooling layer for robustness and size reduction of extracted features.
• Fully connected layer
• Output layer of Softmax function</td></tr>
<tr><td>Experiment results</td><td>96.11 % accuracy. This model has efficiently identified two confusing activities like going upstairs and walking.
<table>
<tr><th></th><th>GU</th><th>DU</th><th>SD</th><th>RU</th><th>WK</th><th>BY</th><th>SW</th></tr>
<tr><td>Test1</td><td>99.28</td><td>95.61</td><td>98.20</td><td>99.08</td><td>92.94</td><td>96.38</td><td>100</td></tr>
<tr><td>Test2</td><td>82.89</td><td>99.43</td><td>99.77</td><td>100</td><td>90.24</td><td>99.29</td><td>100</td></tr>
<tr><td>Test3</td><td>96.42</td><td>98.56</td><td>97.24</td><td>100</td><td>97.58</td><td>95.12</td><td>95.43</td></tr>
<tr><td>Test4</td><td>86.34</td><td>98.96</td><td>99.93</td><td>100</td><td>85.23</td><td>99.86</td><td>99.67</td></tr>
<tr><td>Test5</td><td>95.74</td><td>98.31</td><td>100</td><td>100</td><td>96.27</td><td>99.88</td><td>100</td></tr>
<tr><td>Test6</td><td>80.18</td><td>98.99</td><td>100</td><td>100</td><td>85.42</td><td>99.40</td><td>99.69</td></tr>
<tr><td>Test7</td><td>93.05</td><td>96.59</td><td>96.88</td><td>100</td><td>92.13</td><td>93.01</td><td>99.36</td></tr>
<tr><td>Test8</td><td>92.06</td><td>98.13</td><td>99.59</td><td>99.71</td><td>94.11</td><td>99.13</td><td>99.68</td></tr>
<tr><td>Test9</td><td>90.29</td><td>98.32</td><td>97.76</td><td>99.49</td><td>88.99</td><td>98.29</td><td>99.71</td></tr>
<tr><td>Test10</td><td>93.24</td><td>97.01</td><td>97.11</td><td>99.41</td><td>90.98</td><td>90.54</td><td>99.68</td></tr>
</table></td></tr>
<tr><td>Future Scope</td><td>This work can be extended to wide dataset and activities captured by different placements of Smartphone.</td></tr>
</table>

In reference [14], author has proposed a novel HAR system based on shallow Recurrent Neural Network (RNN) combined with Long short term memory (LSTN) deep learning algorithm. Thus, researcher has developed a hybrid approach combining category 2 and category 3 techniques.

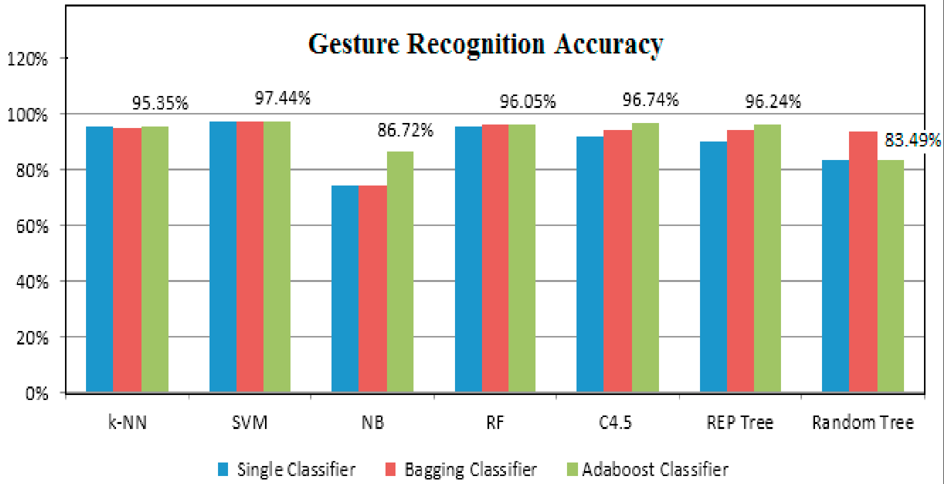
Paper	Author: Preeti Aggarwal et al[14] Year of publication: 2019
Dataset	Standard dataset(WISDM) consisted of 6 activities performed by 29 participants

<table>
<tr><td>Activities</td><td>Six activities which include walking, jogging, upstairs, downstairs, sitting and standing.</td></tr>
<tr><td>Input data detail</td><td>Sensor device: Android Smartphone and later tested on Raspberry Pi.
Location of device: Front leg pocket of participants
Sensors selected for capturing human events: accelerometer
sampling frequency : 20Hz
Signals considered for processing: Tri-axial accelerometer signals.</td></tr>
<tr><td>Data preprocessing</td><td>Accelerometer reading is partitioned into time step of 100, fixed window size of 80, batch size of 64 and 75 number of epochs.</td></tr>
<tr><td>Architecture</td><td>RNN-LSTN model developed by author make use of LSTN memory cells by replacing RNN nodes in traditional RNN model. It consists of two hidden layers containing 30 neurons each.</td></tr>
<tr><td>Classification</td><td>• 70% of data was used for training. Softmax classifier with bias weight initialization of 1.0 is employed.

• For learning, Adam optimizer is used with learning rate of 0.0025 and loss rate of 0.0015.</td></tr>
<tr><td>Experiment results</td><td>This model showed 95.78% of overall accuracy.
<table>
<tr><th></th><th>Accuracy(%)</th><th>Precision(%)</th><th>Recall(%)</th><th>F1-Score(%)</th></tr>
<tr><td>Lightweight RNN-LSTM Model</td><td>95.78</td><td>95.81</td><td>95.78</td><td>95.73</td></tr>
</table>
<table>
<tr><td>Activities</td><td>Accuracy Rate</td></tr>
<tr><td>Downstairs</td><td>93%</td></tr>
<tr><td>Jogging</td><td>99%</td></tr>
<tr><td>Sitting</td><td>97%</td></tr>
<tr><td>Standing</td><td>95%</td></tr>
<tr><td>Walking</td><td>99%</td></tr>
<tr><td>Upstairs</td><td>81%</td></tr>
</table></td></tr>
<tr><td>Future Scope</td><td>Upstairs activity achieved only 81% of accuracy rate and most of these activities are wrongly identified as walking activities. In future, this rate can be increased. Dynamic windowing architecture can be used in place of fixed windowing scheme. Instead of single tri-axial accelerometer, multiple sensors can be applied.</td></tr>
</table>

In reference [15], author has implemented HAR system grounded on machine learning algorithms. Thus, this research falls under category1.

Paper	Author: Abdulhamit Subasia et al [15] Year of publication: 2019
Dataset	Standard UCI dataset comprising of 6 human activities of 30 volunteers with age between 19-48 years.
Activities	Six basic activities namely standing, sitting, lying, walking, ascending stairs and descending stairs.

Input data detail	Sensor device: Samsung Galaxy S II Smartphone Location of device: mounted on waist of participants Data used: acceleration data sampling frequency : 50Hz
Data preprocessing	Noise-filters are applied and then sampled in fixed width sliding windows of 2.56 sec and 50% overlap.
Feature Extraction	561 features extracted from each window by calculating time and frequency domain.
Classification	Single, bagging and adaptive boosting ensemble classifier (Adaboost) classifiers are combined separately with each of K-Nearest Neighbour (K-NN), Support Vector Machine (SVM), Naive Bayesian Networks (NB), RF, C4.5, REP Tree and Random Tree algorithms.
Experiment results	There is 97.44% accuracy of Smartphone-based HAR system when Adaboost ensemble classifier is merged with SVM. 
Future Scope	Dynamic windowing architecture can be used in place of fixed windowing scheme. This study can be extended to healthcare monitoring and smart city systems with better prediction rates.

III. CONCLUSION

This paper has surveyed existing HAR Systems and explained technical aspects of various phases involved in these systems. It also contains comprehensive study of prominent research papers with proposed HAR systems using various techniques like machine learning, neural networks and deep learning. This study briefly reviewed all necessary information like dataset, input detail, methodology, results and future scope related to each explored paper. From mere results (accuracy factor) of proposed systems, we could not conclude that one HAR research work is superior to other. The reason is that these research results are based on various factors, for instance HAR systems working on dataset with only postural activities may achieve higher accuracy than those including transitional activities. Therefore, during this research survey, we found that a HAR system outcome depends on many factors like activity chosen, location of sensor, features extracted, number of samples selected for training and testing, classifier used, classification method used etc. So we deduce that no single method can be considered best for recognition of any activity.

REFERENCES

[1]. Bao L, "Activity recognition from user annotated acceleration data", *Pervasive computing*(2004)
[2]. Ravi N, Dandekar N, "Activity recognition from accelerometer data", Proceedings of the 17th Conference on *Innovative Applications of Artificial Intelligence (IAAI'05), vol. 3*, 2005.

[3]. Karantonis DM, Narayanan MR, “Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring”, *IEEE Trans Inf Technology Biomed* 2006;10(1):156-67.

[4]. Martin H, Bernardos AM, “Activity logging using lightweight classification techniques in mobile devices”, *Pervasive Ubiquitous Computing 2013*; 17(4):675-95.

[5]. Berchtold M, Braunschweig T, “Activity Recognition Service for mobile phones”, *International Symposium on Wearable Computers (ISWC)* 2010.

[6]. Kwapisz JR, Weiss GM, “Activity recognition using cell phone accelerometers”, *ACM SIGKDD Explorer Newsletter* 2011; (December):74-82.

[7]. Thiemjarus HAS, ”Accurate activity recognition using a mobile phone regardless of device orientation and location”, *International Conference on Body Sensor Networks (BSN);* 2011.

[8]. Anguita D, Ghio A, ”Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine”, *IWAAL’12 Proceedings of the 4th International Conference on Ambient Assisted Living and Home Care*, 2012.

[9]. Morales, David, ”Physical activity recognition by smartphones”, *Biocybernetics and Biomedical Engineering*. 37. 10.1016/j.bbe.2017.04.004.

[10]. Jobanputra, Charmi & Bavishi, ”Human Activity Recognition”, *Procedia Computer Science. 155. 698-703*. 10.1016/j.procs.2019.08.100.

[11]. Hassan, Mohammad & Uddin, ”A robust human activity recognition system using Smartphone sensors and deep learning”, *Future Generation Computer Systems.* 81. 10.1016/j.future.2017.11.029.

[12]. Sikder, Niloy & Chowdhury, ”Human Activity Recognition Using Multichannel Convolutional Neural Network”

[13]. R. Zhu et al, "Efficient Human Activity Recognition Solving the Confusing Activities Via Deep Ensemble Learning," *in IEEE Access, vol. 7, pp. 75490-75499,* 2019.

[14]. Agarwal, Preeti & Alam, ”A Lightweight Deep Learning Model for Human Activity Recognition on Edge Devices”

[15]. Abdulhamit Subasi, Asalah Fllatah, ”Smartphone-Based Human Activity Recognition Using Bagging and Boosting,*Procedia Computer Science,Volume 163,Pages 54-61*,ISSN 1877-0509”

[16]. Nweke, Henry & Wah, ”Deep Learning Algorithms for Human Activity Recognition using Mobile and Wearable Sensor Networks”, State of the Art and Research Challenges. *Expert Systems with Applications.* 105. 10.1016/j.eswa.2018.03.056.