

A Novel Random Split Point Procedure Using Extremely Randomized (Extra) Trees Ensemble Method for Human Activity Recognition

B Padmaja^{1,*}, V V Rama Prasad² and K V N Sunitha³

¹Associate Professor of CSE, Institute of Aeronautical Engineering, Hyderabad, India, b.padmaja@gmail.com

²Professor of CSE, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, vvramaprasad@gmail.com

³Professor of CSE, BVRIT Hyderabad College of Engineering for Women, Hyderabad, India, kvnsunitha@gmail.com

Abstract

INTRODUCTION: Automatic detection and recognition of various human physical movements while performing daily life activities such as walking, jogging, running, sitting, standing etc. are usually considered as Activity Recognition (AR). AR is a prominent research area in many applications, such as elderly care, security and surveillance, smart homes, health and fitness. Extremely Randomized Trees Classifier (ET Classifier) is a type of ensemble learning technique used in Activity Recognition, which clusters several different decision trees into a forest from a single learning set and gives the classification result. But it suffers from high variance and over-fitting problem due to high inter-dependency among hyperparameters during model building.

OBJECTIVES: The primary objective of this paper is to propose a novel *Random Split Point* procedure for Extra tree classifier to make the existing approach more robust, less variance, less computational time in obtaining optimal split points and faster in model building. This approach generates K random split points from all the candidate features of the dataset and selects the best split point based on the maximum score obtained by information gain measure.

METHODS: In the proposed method to improve the randomization and accuracy of AR system, a novel random split-point procedure for ET classifier is proposed. This approach reduces the bias-variance problem induced due to the three hyperparameters such as K , n_{min} and M used in split-point procedure of existing ET classifier (K : number of randomly selected attributes at each node, n_{min} : minimum sample size for splitting a node, M : number of decision trees for ensemble). This approach generates K random split points from all the candidate features of the dataset and selects the best split point based on the maximum score obtained by information gain measure.

RESULTS: The proposed approach is experimented with two public AR datasets HAR and HAPT (UCI Machine Learning Repository) containing 6 and 12 activities respectively. In HAR dataset, smartphone sensed sensor signals of 3 static and 3 dynamic human daily activities are there, where as in HAPT dataset apart from these 6 daily activities, 6 postural transitions data is available. Experimental results and comparative analysis show that the proposed method outperforms over other existing techniques with an accuracy of 94.16% for HAR dataset and 92.63% for HAPT dataset. It also takes less computational time in finding optimal split-points and less model building time.

CONCLUSION: AR systems can be used as an intelligent system in healthcare to monitor the behaviour of healthy people by recognizing their daily activities. These systems also help in early detection of some chronic diseases and improve the quality of life. In this paper, an attempt is made to improve the accuracy of Activity Recognition over some existing methods.

Keywords: Activity Recognition, Extremely Randomized Tree Ensemble Method, Smartphone, Sensors, Random Split Point.

Received on 20 March 2020, accepted on 19 May 2020, published on 28 May 2020

Copyright © 2020 B Padmaja *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.28-5-2020.164824

*Corresponding author. Email: b.padmaja@gmail.com

1. Introduction

Human Activity Recognition primarily focuses on surveying the characteristics of physical and psychological human behaviour. Individual activities are identified through various built-in sensors in wearable devices and smart phones. An activity sensor is a device used to identify and trace body movements. Human activity recognition uses several sensors such as accelerometers, gyroscopes, and heart rate monitors while employing unprecedented machine learning procedures to translate low-level kind of sensor data and yield prolific contextual info in a real-life aspect. Individuals continually connect with their ease, little sized cell phones in their everyday exercises, which have prompted the ascent in the examination of fetching useful knowledge from information procured by ubiquitous sensors in smart phones [1, 2]. It has numerous applications and benefits in everyday life, such as life style improvement, health and fitness, smart homes, elderly care, etc. While HAR aims to incorporate motion and physical movement data, human behaviour analysis focuses on both physical movements and psychological states of the being.

Motivation and Contribution Highlights

Human activity recognition technology exploits distinctive multi-modular information produced from different gadgets to distinguish human stance, physical activity status, and conduct activities. The interest in understanding human exercises has developed in the medicinal services area, particularly in elderly care support, restoration help, diabetes, and subjective issue. Strong proof shows that ordinary observing and acknowledgment of physical exercises can possibly help to oversee and diminish the danger of numerous sicknesses, for example, weight, cardiovascular and diabetes. Accelerometer and Gyroscope are the most broadly utilized cell phone sensors for human action perception. The contributions in this paper are as follows:-

- A Robust *Random Split Point* procedure is proposed here to reduce the variance effect on the ensemble model using the existing Extremely Randomized (Extra) trees classifier.
- Experimental analysis is carried out on two standard publicly available activity recognition datasets HAR and HAPT having six and twelve activities respectively.
- Later, comparative study is performed with some significant existing techniques based on various statistical performance measures obtained.

Organization of the Paper

Remaining paper is structured as: Section 2 provides an extensive summary of literature review in this domain. Proposed method along with their algorithmic steps is given in Section 3. Experimental analysis, results discussion and comparisons with other state of the art

existing approaches are presented in Section 4. Finally, Section 5 presents the conclusion of the paper.

2. Related Work

Several studies have been regulated over past years to make the collected sensor data more accurate and precise. While accelerometers and gyroscopes are the most frequently used sensors for tracing human activities, smart phones have also become a prominent choice to improve the accuracy of human activity recognition. *Ronao et al. (2017)* [3, 8] proposed a two-stage Continuous Hidden Markov Model (CHMM) to classify six physical activities of daily life. This approach reduced a lot of feature computation overhead, but it has high model building time compared to tree based classifier. Later, the authors experimented with a deep model to predict the accuracies of daily life activities. *Sadiq et al. (2015)* [4] used activity recognition mechanism for evaluating the performance of classifiers in disaster mitigation. *Chen et al. (2017)* [5] assessed the performance of modified Extreme Learning Machine (ELM) classifier for recognizing six human activities using smartphone sensor data. In the proposed approach, various positions of smartphones are considered during data collection and using weighted ELM the overall performance of the classifier was assessed. *Almaslukh et al. (2018)* [7] used a robust deep learning approach for recognizing human activities using smartphone sensor data. This approach has the advantage of automatic feature learning using a multi-layered approach with better accuracy results. *Cao et al. (2018)* [9] proposed a hierarchical group based context-aware classifier to improve the performance of activity recognition and to reduce the misclassification rate through context awareness instead of instead computation. But the proposed approach requires more computational resources and training time. *Kanjo et al. (2018)* [10] aimed using sensor data driven fusion approach to assess the impact of surrounding environment and physiological changes and emotion on activity recognition. *Jian et al. (2018)* [11] experimented on sequential human activity recognition using automatic feature learning mechanism of a hybrid deep framework. The authors used three deep models such as CNN, LSTM and ELM on opportunity dataset and achieved an accuracy of 91.8%. But this approach suffers from more model building time and computational resources. *Subasi et al. (2018)* [12] suggested Adaboost ensemble classifier to get higher accuracy of activity recognition. But this approach didn't use any feature extraction technique, hence suffers from biased results. *Brastein et al. (2017)* [13] aimed to identify activities from simple inertial sensors using decision tree ensemble algorithm XGBoost. This approach gives an accuracy of 94.6%, but suffers from an over-fitting problem. *Kwon et al. (2018)* [14] proposed an activity recognition system that collects data from the smart watch and used ANN classifier for activity recognition. But this approach suffers from an over-fitting

problem and consecutive activities of a person can't be determined. *Jordao et al.* [15] focused on the process employed to generate data samples for activity recognition, because many traditional approaches are susceptible to bias leading to skewed results. The author observed that the accuracy of many datasets like MHealth is low due to imbalance in data and missing values. *Ku Nurhanim et al.* [2018] [16] used semi-non-overlapping mechanism and 10-fold cross validation for sample generation from smartphone sensor data for activity recognition. This approach used an ensemble method using classifiers such as Bagging, Adaboost, Rotation Forest, Ensemble nested dichotomy and Random Subspace and achieved an accuracy of 94.22%. *Cho and Yoon (2018)* [17] proposed 1D CNN model that employs divide and conquer based classifier learning couples with test data sharpening. The authors experimented with two standard datasets UCI HAR and Opportunity and obtained an accuracy of 91.62%. *Münzner et al. (2017)* [18] used CNN's on RBK and PAMAP2 datasets. *Nurhanim et al. (2017)* [19] study the performance of different classification kernels of the SVM for classifying various daily activities. Test subjects performed various physical

activities such as sitting, climbing stairs, and laying down which were tracked and measured using inertial sensor signals. The collected data was processed using signal processing methods and multiple features of time and frequency domain. *Luštrek et al. (2015)* [20] made use of smart phones to aid in better tracking of daily lifestyle activities of diabetes patients, which could be beneficial for physicians as well the patients themselves. *Ole M. B. et al. (2017)* [21] demonstrated that the best in class choice tree gathering calculation XGBoost gives an exactness of 94.6% approved on a free test set. *Kaur et al. (2016)* [22] applied human exercises forecast with the assistance of different AI models and information mining sets of tools. Cross-validation has been performed to check the consistency of the group model and precision of over 85% has been acquired. *Padmaja et al.* proposed a distributed and parallel decision forest approach for human activity recognition and also experimented on human stress behaviour using socio-mobile data [23, 24, 25]. Table 1 shows the state-of-the-art existing literature on activity recognition.

Table 1. Existing literature on activity recognition

Year & Reference	Classifier (s)	Type of Sensor	Type of Devices	Data set & Accuracy	Observation	Application (s)
2013 [1]	SVM (MC-HF-SVM)	Acc, Gyro	Smart phone	UCI HAR Dataset 89.3%	Low accuracy and high complexity in activity recognition	Assisted Healthcare
2015 [4]	KNN	Acc, Digi Compass	Smart phone	Collected Data 89%	Accuracy is low with insufficient number of classes and subjects (4 participants)	Crowd Disaster Mitigation
2016 [8]	Deep CNN	Acc, Gyro	Smart phone	UCI HAR Dataset, 94.79%	high computational resources	Surveillance-based security, context-aware computing, and ADL
2017 [3]	Continuous HMM	Acc, Gyro	Smart phone	UCI HAR Dataset 93.18%	More computational resources (offline Computation of data) (battery life)	Ambient assistive living, context-aware computing, surveillance-based security
2017 [5]	NN, RF, and SVM	Acc, Gyro	Smart phone	Collected Data, 94.95%	The dataset strength is relatively poor due to fewer subjects, environment controlled, and over sampled.	Daily life activity monitoring
2017 [6]	Online SVM	Acc, Gyro	Smart phone	Collected Data, 94.89%	Position-dependent HAR is not suitable for all subjects.	Healthcare Services.
2017 [13]	Tree Ensemble Algorithm XGBoost	Acc, Gyro	Smart phone	UCI HAR Dataset, 94.6%	Over-fitting	Health care, security monitoring
2018 [7]	CNN	Acc, Gyro	Smart phone	RealWorld HAR	Low accuracy and	Activities of daily

				Data set, 88%	high computation overhead	living (ADL)
2018 [9]	Hierarchical Group-based Context-aware Classifier (GCHAR)	Acc, Gyro	Smart phone	UCI HAR Dataset, 94.16%	Training time is high, more computational resources	Health monitoring, fitness tracking
2018 [10]	SVM, RF KNN	HR, body temp., EEG, Blood volume pulse, noise, air pressure, UV etc.	Smart phone	Collected Data	More training time and high cost	Effect of stressors (pollution, noise, crowded area) on human health
2018 [11]	CNN, LSTM, ELM classifier	Acc, Gyro, Mag	Smart phone	OPPORTUNITY dataset, 91.8%	More run time and high Computational resources	Athletic competition, medical care, smart home, elderly care
2018 [12]	Adaboost ensemble classifier	Acc, Gyro, Mag	Smart phone	UCI REALDISP dataset, 94.98%	No feature extraction techniques used, biased results.	ADL, healthcare, AAL, home monitoring, personal fitness assistants
2018 [16]	Bagging, Adaboost, Rotation forest, Ensemble nested dichotomies and Random subspace	Acc, Gyro	Smart phone	UCI HAR dataset, 94.22%	Over fitting	Rehabilitation, computer games, animation
2018 [14]	ANN	Acc, Location	Smart watch	Collected Data, 95%	Consecutive activities of a person are difficult to determine, over fitting problem.	Healthcare, fitness, and abnormal behavior detection
2018 [17]	1D CNN	Acc, Gyro	Smart phone	UCI HAR and OPPORTUNITY dataset, 91.62%	More model building time	Activity Monitoring, Fitness Tracking
2019 [15]	CNN	Acc, Gyro, Mag	Smart phone	MHEALTH (89.02%), PAMAP2 and WISDM, 94.64%	Low accuracy	Healthcare, smart environments
2019 [2]	XGBoost, RF, Extra Trees and Softmax Regression	Acc, Gyro	Smart phone	UCI HAR Dataset 94.88%	In Cascaded Ensemble Learning model (Softmax Regression + ExtraTrees) was the best combination, in achieving high performance.	Healthy lifestyle maintenance and patient rehabilitation management.
This work	Modified Extra Tree classifier	Acc, Gyro	Smart phone	UCI HAR and HAPT datasets	Less variance, less model building time, more accuracy	Healthcare, Fitness Tracking, Elderly care

* Acc : Acclerometer, Gyro : Gyroscope, Mag : Magnetometer, Digi : Digital, HR : Heart Rate

3. Proposed Method for Activity Recognition

3.1 HAR and HAPT Datasets:

“Activity Recognition using Smartphone” dataset was built and made universally accessible by Davide Anguita

et al. [1]. The smart phone sensor data was gathered from the experiments carried out on a group of 30 volunteers whose age was between 19 and 48 years. The set of physical activities taken by the authors includes walking, sitting, standing, laying, walking upstairs and walking downstairs. The authors attached a Samsung Galaxy SII smart phone to each subject to capture sensor data. Signals sent by the accelerometer and gyroscope, embedded within the Samsung Galaxy S II, and are captured through a smart phone app. An accelerometer is

used to determine the acceleration of the device. Values along the X, Y and Z axis are used to identify motions such as swinging, tilting, vibration, etc. A gyroscope, on the other hand, utilizes the angular velocity to calculate the rotation or twist in a smart phone device. While an accelerometer detects directional movement, a gyroscope detects the lateral orientation of the device. They captured the sensor signals at a constant rate of 50Hz, which were subsequently preprocessed to reduce noise. The signals were preprocessed for noise reduction with a median filter and a 3rd order low-pass Butterworth filter with a 20Hz cutoff frequency. The Butterworth filter was employed to distinguish the acceleration signal into body acceleration and gravitational acceleration. The processed signals were sampled into a fixed window of length 2.56 seconds with a 50% overlap. Each window had 128 data points for every original features recorded, which are body acceleration, body gyroscope and gravity acceleration over X, Y and Z axis. The inertial signals were feature engineered and several, time and frequency, features were extracted from each window. Feature engineering yielded a feature vector of 561 attributes. The authors randomly split the dataset into 70:30 ratios which formed a distribution of 21 subjects for training and 9 subjects for testing. In HAPT dataset, the process of data collection is as same as HAR, but it contains 12 human activities, three static, three dynamic and six transitions between activities such as sit_to_stand and stand_to_sit etc.

3.2 Proposed Work:

This section presents the proposed framework representation along with the algorithmic steps. The Extra tree classifier builds multiple trees by making bootstrap – False, which means it samples without replacement. Then the classifier chooses the optimal split-point for each one of the K randomly chosen features at every node, which means this algorithm selects a split-point randomly. The existing Extra tree classifier, chooses a subset ‘S’ and selects an attribute ‘a’ for random split. The value of ‘a’ is chosen randomly after finding the maximum and minimum value in S [a_{max}, a_{min}]. The procedure followed in existing Extra tree classifier is given in Table 2. This algorithm draws samples without replacement and the chances of getting the same split-point is high as it selects a split-point randomly from its maximum and minimum range of values.

Table 2. Existing Procedure for selecting a random split point in ET classifier

<p>Split_a_node (S) <i>Input:</i> Sample subset S consisting of set of features <i>Output:</i> A split point</p>

<ul style="list-style-type: none"> - If Stop_split(S) is TRUE then return nothing. - Otherwise select K features {a₁, a₂, ..., a_k} from all candidate features; - Draw K splits {s₁, s₂, ..., s_k}, where s_i = Pick_a_Random_Split(S, a_i), where i = 1, 2, ..., K; - Return a split s* based on maximum score where Score(s*, S) = max_{i=1, 2, ..., K} Score(s_i, S).
<p>Pick_a_Random_Split (S, a) <i>Input:</i> Sample subset S and a feature / attribute a <i>Output:</i> A split location</p> <ul style="list-style-type: none"> - Let a_{max} and a_{min} be the maximum and minimum value of the attribute a in S; - Draw a random split-point a_c uniformly from [a_{min}, a_{max}]; - Return the split [a < a_c].
<p>Stop_split (S) <i>Input:</i> Sample subset S <i>Output:</i> A Boolean value TRUE or FALSE</p> <ul style="list-style-type: none"> - If S < n_{min}, then return TRUE; - If all attributes are constant in S, then return TRUE; - If the output is constant in S, then return TRUE; - Otherwise, return FALSE.

In this work, a novel procedure for random-split-point is proposed. The overall block diagram of this work is shown in figure 1.

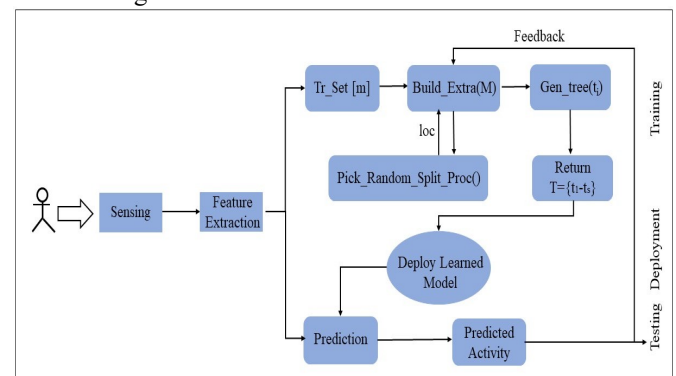


Figure 1. Overall Block Diagram of Proposed Work

The raw sensor data is first pre-processed using various noise filters and then 561 handcrafted features are generated. Then the training and testing set is prepared by taking 70% and 30% of the dataset. The training set data (M) goes as an input to *Build_Extra* trees procedure, which exploits a sub-procedure called *Random_Split_Procedure* and outputs location indexing. Further it generates sub model (t_i); $i \rightarrow [1, S]$ and for each sub-model t_i , it computes local misclassification rate. In ensemble procedure, *Statistical Mode* [set of values] is performed. Finally the procedure returns $T = \{t_1 \dots t_s\}$. Learned model is deployed to predict the labels for test data. The performance of the framework is judged by measures such as precision, recall, F1-Score and accuracy etc. Table 3 contains the procedure for building tree ensemble method using ET classifier through *Random_Split_Procedure*. In the proposed approach, for each sub model, accuracy and misclassification rate is computed and initial weights are updated.

$$\text{Misclassification Rate} = \sum_{i=1}^n w_i * \text{error}_i / \sum_{i=1}^n w$$

For each misclassification rate, stage is calculated and weights are updated for every misclassified instance.

$$\text{Stage} = \ln((1 - \text{misclassification rate}) / (\text{misclassification rate}))$$

Then the new weights are updated using the formula:

$$\text{new_weight} = \text{weight} * e^{\text{stage} * \text{error}}$$

Finally, M sub models are generated and then it will predict the output of each model using the formula:

$$\text{Prediction} = \text{Stage} * (\text{IF}(\text{Output} == 1) \text{ THEN } -1 \text{ ELSE } +1)$$

Table 3. Procedure for building tree ensemble method using ET classifier using *Random_Split_Procedure*

<p>Build_ET_Ensemble (ES)</p> <p><i>Input:</i> Tri-axial accelerometer and gyroscope training dataset <i>TS</i>.</p> <p><i>Output:</i> ET ensemble $TE = \{t_1, t_2, \dots, t_M\}$.</p> <ul style="list-style-type: none"> - for $i = 1$ to M <ul style="list-style-type: none"> - Create a tree $t_i = \text{Build_ET}(TS)$. - Return TE.
<p>Build_ET (TS)</p> <p><i>Input:</i> Tri-axial accelerometer and gyroscope training</p>

dataset *TS*.

Output: A tree t .

- Return a leaf node by averaging the frequency of each class in *TS*
- if $|TS| < n_{min}$
- Otherwise
 1. Select randomly K features $\{f_1, f_2, \dots, f_k\}$, by making bootstrap = **False** from all candidate features in *TS*;
 2. Generate K split-points $\{s_1, s_2, \dots, s_k\}$ by using the *Random_Split_Procedure* $s_i = (TS, f_i)$ where $i = 1$ to K ;
 3. Select the best split point s^* by computing the maximum_score $(s^*, TS) = \max_{i=1, \dots, k}$ for a sample *TS* and a split s , using the formula $\text{Score}(s^*, TS) = 2 * I_c(TS) / H_S(TS) + H_C(TS)$ where $I_c(TS)$ is the information gain, $H_S(TS)$ is the split entropy and $H_C(TS)$ is the log entropy in *TS*;
 4. Split the dataset *TS* into two subsets S_{left} and S_{right} based on the test score s^* ;
 5. Build t_{left} and t_{right} by calling **Build_ET** recursively from the two subsets S_{left} and S_{right} ;
 6. Create a new tree using the split node s^* , attach left and right sub-trees t_{left} and t_{right} to this node and return the tree t .

Random_Split_Procedure (F, V_F, δ, β, N)

Input: A training dataset *TS* with N as the total number

of samples, F as the feature set, V_F as each feature vector in set $F = \{f_1, \dots, f_{N-1}\}$,

Output: A split node index.

- For a numerical feature f , use the Congruential equation

$$R_{j+1} = (\delta * R_j + \beta) \bmod N \quad \text{where } 0 < j < N$$

- Here, R_0 is a positive integer, called the initial value ($0 < R_0 < N$), δ is an integer, also known as multiplier, β is also another positive integer and N is the sample size.
- Choose the value of δ , β , N with the following rules -
 1. N and β must be co-prime with each other.
 2. $(\delta - 1)$ must be fully divisible with all prime factors of N .
- Return split node index

4. Experimental Results

The experimental setup consists system environment (Hardware and Software specifications) as – OS: Ubuntu 16.04 LTS, 64 bit, Python 3.7 version is used in implementation. The performance measures considered for activity recognition are – precision, recall, F1-Score and accuracy. The performance of AR system is maximized using recall which is the ability of a model to find all the relevant cases within the dataset. Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives. While recall factor shows the ability for finding out all the relevant instances in the dataset, precision judges the total proportion of data points which the model infers was relevant in actual were relevant. Figure 2 and 3 shows the confusion matrix for the proposed system for activity recognition using HAR and HAPT datasets.

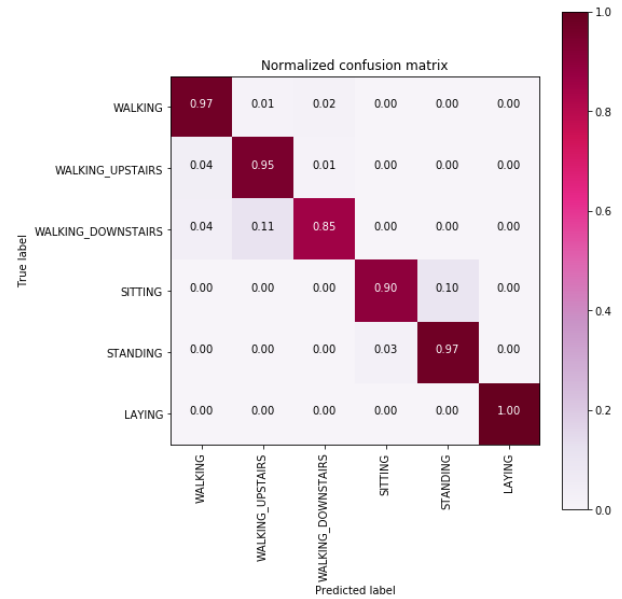


Figure 2. Confusion Matrix of the proposed work using HAR dataset

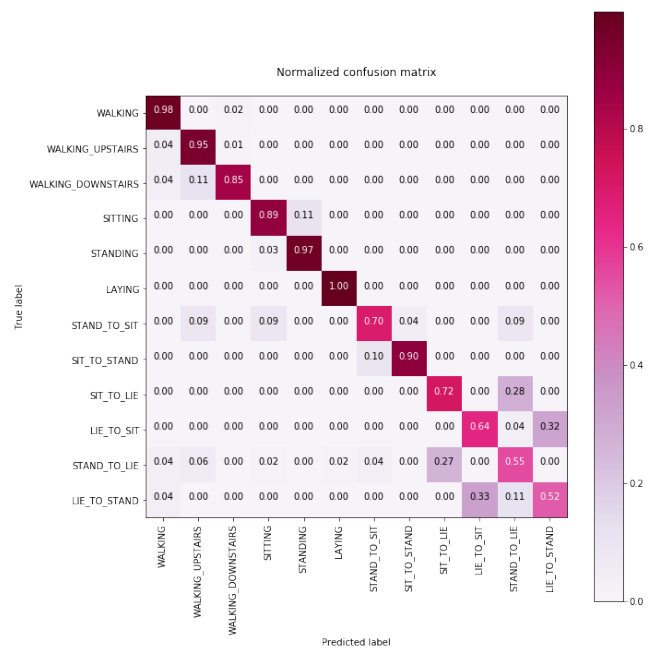


Figure 3. Confusion Matrix of the proposed work using HAPT dataset

Table 4. shows the comparative statistical performance measures of proposed approach using HAR and HAPT datasets.

Parameters	Precision	Recall	F1-score	Overall
------------	-----------	--------	----------	---------

	(in %)	(in %)	(in %)	accuracy (in %)
HAR Dataset	0.95	0.94	0.94	94.16
HAPT Dataset	0.93	0.93	0.92	92.63

Where precision = $TP / (TP+FP)$, Recall = $TP / (TP+FN)$, F1-score = $2 * ((Precision * Recall) / (Precision + Recall))$ and Accuracy = $((TP + TN) / (TP + FP + TN + FN))$

The proposed approach is compared with state-of-the-art classical machine learning algorithms used popularly for activity recognition. In CART algorithm, trees are grown from the learning sample and pruned by estimating the errors using 10-fold cross validation approach. In KNN algorithm, the value of $K = 7$ has given the best accuracy for the AR datasets. Bayes algorithm is proven to be not suitable for these datasets because of poor performance obtained. RF algorithm outperforms significantly with good accuracy. Each time it builds a tree by using the bootstrap copy of the learning sample. At each test node, the optimal split is obtained by searching a random subset of K candidate attributes. This algorithm performs well in terms of degree of randomization, if K is small compared to number of attributes n , otherwise RF algorithm suffers from over-fitting problem. The existing ET algorithm performs well in terms of randomization compared to RF, but prone to bias-variance problem.

Table 5 and 6 show the comparative analysis of recognition accuracies and Cohen's Kappa score of proposed approach with existing classifiers.

Table 5. Comparative analysis of recognition accuracies of proposed approach with existing classifiers

RECOGNITION ACCURACY (in %)						
Classifier	KNN	CART	BAYES	RF	ET	Proposed Method
HAR Dataset	90.3	84.9	77.0	92.7	93.8	94.1
HAPT Dataset	89.2	80.2	74.7	91.0	91.7	92.6

Table 6. Comparative analysis of Cohen's Kappa Score as compared with proposed model

COHEN'S KAPPA SCORE				
Datasets	KNN	CART	BAYES	RF
HAR	0.88	0.78	0.72	0.94
HAPT	0.88	0.78	0.72	0.94

Table 7 shows comparative analysis of training and testing time of proposed approach with existing approaches using HAR and HAPT datasets.

MODEL TRAINING TIME (SS.ms)						
Classifier	KNN	CART	BAYES	RF	ET	Proposed Method
HAR Dataset	00.91	03.79	00.15	18.93	04.25	04.09
HAPT Dataset	00.23	04.55	00.087	20.41	04.16	04.11
MODEL TESTING TIME (SS.ms)						
HAR Dataset	14.20	00.0079	00.19	00.10	00.13	00.11
HAPT Dataset	19.14	00.0069	00.24	00.09	00.14	00.12

Figure 4 and 5 show comparative analysis of recognition accuracies of proposed approach with existing techniques using HAR and HAPT datasets.

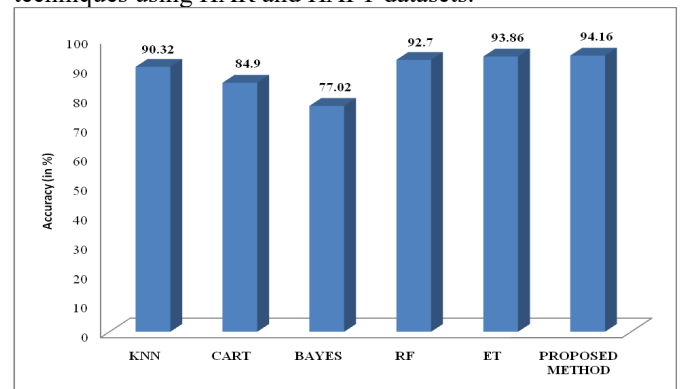


Figure 4. Recognition accuracies of proposed approach with existing techniques using HAR dataset

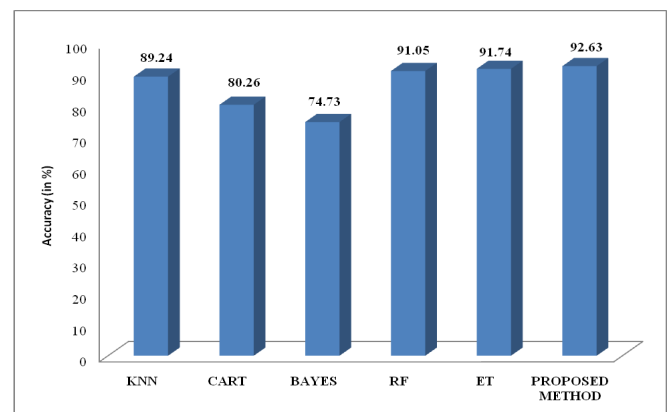


Figure 5. Recognition accuracies of proposed approach with existing techniques using HAPT dataset

The proposed approach predicts the class labels of each activity with a reasonable accuracy of 94.16% for HAR dataset and 92.63% for HAPT dataset. From the results it is observed that our proposed approach shows better accuracy in comparison with existing classifiers used for activity recognition.

5. Conclusion

Activity recognition system plays a vital role in many applications such as virtual education, gaming, entertainment, sport injury detection, elderly care and rehabilitation, and smart home environment monitoring. AR systems can be used as an intelligent system in healthcare to monitor the behaviour of healthy people by recognizing their daily activities. This system also helps in performing a long-term analysis of early detection of some chronic diseases and to improve the quality of life. In this work, a random_split_point procedure is devised for human activity recognition using two public datasets HAR and HAPT from UCI machine learning repository. This approach utilizes the existing extremely randomized (EXTRA) trees ensemble method with a new procedure for random_split_point selection for building trees. Experimental results and comparative analysis show that the proposed method outperform over other existing techniques with an accuracy of 94.16% for HAR dataset and 92.63% for HAPT dataset. The proposed approach takes less computational time in finding optimal split-points and less model building time. In this an attempt is made to improve the activity recognition accuracy over some significant methods available in literature.

References

- [1] Davide Anguita, Alessandro Ghio, Luca Oneto (2013), "Energy Efficient Smartphone-Based Activity Recognition using Fixed-Point Arithmetic," in *Journal of Universal Computer Science*, 19(9), pp. 1295-1314, DOI:10.3217/jucs-019-09-1295.
- [2] Shoujiang Xu, Qingfeng Tang, Linpeng Jin, Zhigeng Pan (2019), "A Cascade Ensemble Learning Model for Human Activity Recognition with Smartphones," in *Sensors*, MDPI, 19(10), pp. 1-17, DOI:10.3390/s19102307.
- [3] Charissa Ann Ronao and Sung-Bae Cho (2017), "Recognizing Human Activities from Smart Phone Sensors using Hierarchical Continuous Hidden Markov Models," in *International journal of Distributed Sensor Networks*, 13(1), pp. 1-16, DOI: 10.1177/1550147716683687.
- [4] Sadiq, Fatai Idowu, Ali Selamata, Roliana Ibrahim (2015), "Performance Evaluation of Classifiers on Activity Recognition for Disasters Mitigation using Smartphone Sensing," in *Jurnal Teknologi Science and Engineering*, 77(13), pp. 11-19, DOI: 10.11113/jt.v77.6320.
- [5] Yufei Chen, Chao Shen (2017), "Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition," in *Special Section on Intelligent Sensing on Mobile and Social Media Analytics*, IEEE Access, 5(1), pp. 3095-3110, DOI: 10.1109/ACCESS.2017.2676168.
- [6] Zhenghua Chen, Qingchang Zhu, Yeng Chai Soh, Le Zhang (2017), "Robust Human Activity Recognition using Smartphone Sensors via CT-PCA and Online SVM," in *IEEE Transactions on Industrial Informatics*, 13(6), pp. 3070-3080, DOI: 10.1109/TII.2017.2712746.
- [7] Bandar Almaslukh, Abdel Monim Artoli, Jalal Al-Muhtadi (2018), "A Robust Deep Learning Approach for Position-Independent Smartphone-Based Human Activity Recognition," in *Sensors*, 18(11):3726, pp. 1- 17, DOI: 10.3390/s18113726.
- [8] Charissa Ann Ronao, Sung-Bae Cho (2016), "Human Activity Recognition with Smartphone Sensors using Deep Learning Neural Networks," in *Expert Systems with Applications*, Elsevier, 59, 235–244. DOI:10.1016/j.eswa.2016.04.032.
- [9] Liang Cao, Yufeng Wang, Bo Zhang, Qun Jin, Athanasios V. Vasilakos (2018), "GCHAR: An efficient Group-based Context-aware human activity recognition on smart phone," in *Journal of Parallel and Distributed Computing*, Elsevier, 118(1), pp. 67-80, DOI:10.1016/j.jpdc.2017.05.007.
- [10] Eiman Kanjo, Eman M.G. Younis, Nasser Sherkat (2018), "Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach," in *Journal of Information Fusion*, Elsevier, 40, pp. 18-31, DOI: 10.1016/j.inffus.2017.05.005
- [11] Jian Sun, Yongling Fu, Shengguang Li, Jie He, Cheng Xu, Lin Tan (2018), "Sequential Human Activity Recognition Based on Deep Convolutional Network and Extreme Learning Machine Using Wearable Sensors," in *Journal of Sensors*, Hindawi, 2018, pp. 1-10, DOI: 10.1155/2018/8580959.
- [12] Abdulhamit Subasi, Dalia H. Dammas, Rahaf D. Alghamdi, Raghad A. Makawi, Eman A. Albiety, Tayeb Brahimi, Akila Sarirete (2018), "Sensor Based Human Activity Recognition Using Adaboost Ensemble

- Classifier,” in *Procedia Computer Science*, Elsevier, 140(2018), pp. 104-111, DOI:10.1016/j.procs.2018.10.298.
- [13] Ole M. Brastein, Roland Olsson, Nils-Olav Skeie, Thomas Lindblad (2017), “Human Activity Recognition by Machine Learning Methods,” in *Norsk Informatikkonferanse*, Norway, 25, pp. 1-12.
- [14] Min-Cheol Kwon, Sunwoong Choi (2018), “Recognition of Daily Human Activity Using an Artificial Neural Network and Smart watch,” in *Journal of Wireless Communications and Mobile Computing*, 2018, pp. 1-9, DOI: 10.1155/2018/2618045.
- [15] Artur Jordao, Antonio C. Nazare Jr., Jessica Sena, William Robson Schwartz (2019), “Human Activity Recognition Based on Wearable Sensor Data: A Standardization of the State-of-the-Art,” in *Computer Vision and Pattern Recognition*, arXiv:1806.05226 .
- [16] Ku Nurhanim Ku Abd. Rahim, I. Elamvazuthi, Lila Iznita Izhar, Genci Capi (2018), “Classification of Human Daily Activities Using Ensemble Methods Based on Smartphone Inertial Sensors,” in *Sensors*, MDPI, 18(12):4132, pp. 1-19, DOI: 10.3390/s18124132.
- [17] Heeryon Cho ID and Sang Min Yoon (2018), “Divide and Conquer-Based 1D CNN Human Activity Recognition Using Test Data Sharpening,” in *Sensors*, MDPI, 18(4), pp. 1-24, DOI: 10.3390/s18041055.
- [18] Münzner S, Schmidt P, Reiss A, Hanselmann M, Stiefelhagen R, Dürichen R (2017) CNN-based Sensor Fusion Techniques for Multimodal Human Activity Recognition. In: *Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, Miami, USA, pp 158–165.
- [19] Nurhanim K, Elamvazuthi I, Izhar LI, Ganesan T (2017), “Classification of Human Activity based on Smartphone Inertial Sensor using Support Vector Machine,” in *IEEE 3rd International Symposium on Robotics and Manufacturing Automation (ROMA)*. IEEE, Kuala Lumpur, Malaysia, pp 1–5.
- [20] Luštrek M, Cvetkovic B, Mirchevska V, Kafalı Ö, Romero A, Stathis K (2015), “Recognising lifestyle activities of diabetic patients with a smart phone,” in *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*. IEEE, Istanbul, Turkey, pp 317–324.
- [21] Ole M. Brastein, Roland Olsson, Nils-Olav Skeie and Thomas Lindblad, Human Activity Recognition by machine learning methods, <https://ojs.bibsys.no/index.php/NIK/article/view/425>, (2017).
- [22] Amandeep Kaur, Sahil Sharma, Human Activity Recognition Using Ensemble Modelling, *SmartCom 2016*, CCIS 628, pp. 292–298, 2016, Springer Nature Singapore Pte Ltd., DOI: 10.1007/978-981-10-3433-6_35.
- [23] B Padmaja, V V Rama Prasad, K V N Sunitha (2019), “Distributed and Parallel Decision Forest for Human Activities Prediction: Experimental Analysis on HAR-Smartphones Dataset,” *Journal of Computer Science*, 15 (5), pp. 673-680, DOI: 10.3844/jcssp.2019.673.680.
- [24] B Padmaja, V V Rama Prasad, K V N Sunitha (2016), “TreeNet analysis of human stress behavior using socio-mobile data,” in *Journal of Big Data*, Springer, 3(24), 2016.
- [25] B Padmaja, Madhu Bala Myneni, E Krishna Rao Patro (2020), “A comparison on visual prediction models for MAMO (multi activity-multi object) recognition using deep learning,” in *Journal of Big Data*, Springer, 7(24), March 2020.