International Journal of Engineering & Technology, 7 (3.8) (2018) 63-68

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Novel Hybrid Framework for Optimal Feature Selection and Classification of Human Activity Recognition

Nilam Dhatrak[1], Anil Kumar Dudyala[2*]

[1,2] Department Of Computer Science & Engineering
National Institute Of Technology Patna
Patna, India
Dhatraknilam@Gmail.Com
Anilkumard001@Gmail.Com

Abstract

In today's world individuals health concern has improved a lot with the help of advancement in the technology. To monitor an age old person or a person with disability, now-a-days modern wearable smartphone devices are available in the market which are equipped with good collection of built in sensors that can be used for Human Activity Recognition (HAR). These type of devices generate lot of data with many number of features. When this data is used for classification, the classifier may be over trained or will definitely give high error rate. Hence, in this paper, we propose a two hybrid frameworks which gives us optimal number of features that can be used with different classifiers to recognize the Human Activity accurately. It is observed from our experiments that SVM was able to classify the HAR accurately.

Keywords: *feature selection; classification; Human Activity Recognition; SVM-RFE; BPSO.*

1. Introduction

Alarming growth in the population of age group greater than sixty five brings the researchers new challenge of ensuring health of these people. It's not possible to take care of such large number of patients individually. Here comes an on body sensors that can be used for offsite monitoring of the patient. It provides more accurate and precise measurements and is portable.

Now-a-days modern smartphone devices are equipped with good collection of built in sensors, such as accelerometer, gyroscope, proximity sensor, etc., Smartphone devices can be used as on body sensor as they are portable and can be worn on body using a belt very easily.

We can learn and identify the type of activity performed by the patient with the help of data collected by smartphone sensors. Human Activity Recognition (HAR) is a process of recognizing or identifying activity based upon the attributes derived from motion, location, physiological signals and environmental information [1]. Different approaches can be used for HAR task among which supervised machine learning is the most popular approach.

In supervised machine learning approach a model is trained using a labelled data, where it learns a function that is used for predicting the test data. Different machine learning approaches have been proposed by many researchers to recognize the human activity task. Smartphone sensor signals are processed mathematically, which yield large number of features for the HAR data. As HAR data has large number of features, classifier will require high computational cost and greater time complexity. Also considering all features may result in overfitting of the model. Hence, feature selection has to be done to improve the performance of a classifier with respect to accuracy and computational cost. Feature selection plays the major role in classification as selecting non-redundant and more relevant features improves the performance of the classifier.

Feature selection techniques are broadly classified into filters, wrappers and hybrid methods[1]. Filter technique works in the following way, first it ranks all the features of the data and selects only few features that satisfy the given threshold[3], while the wrappers use predictor performance as an objective function to find the best subset in feature space. As filters evaluate each feature separately, interactions among different features of the dataset cannot be obtained which leads to lower performance than wrappers. While the wrappers do consider feature interactions but require higher time complexity to evaluate 2^n subsets where n is the number of features. To obtain a trade-off between computational cost and performance, we can use hybrid approach in which a filter will be followed by a wrapper.

Here we propose two frameworks which are of hybrid approach. The first framework uses wrapper 1 followed by wrapper 2, whereas the second framework uses filter followed by wrapper.

The remaining part of the paper is organized as follows, section 2 discusses the related works in this area where as the framework proposed is discussed in the section 3. Section 4 describes about the HAR dataset used in the problem and results and discussion is dealt in section 5. Finally, section 6 gives the conclusion of the work.

2. Related Work

Very few researchers had worked on the HAR data. Most of them used different machine learning approaches with this data. JL Reyes-Ortiz et al. has presented compilation of different approaches used for activity recognition including their performance[1]. B Romera-Paredes et al. has worked on HAR data and

concluded that SVM with one class vs one class classification and majority voting ensemble method has given an accuracy of 96.4%[14]. MT Uddin et al. had worked on different types on HAR data and was able to reduce the smartphone based HAR data to 81.35%. Random forest based ensemble method was proposed by Z Feng et al. for the data collected using multiple sensor set, the method resulted into 93.44% accuracy[7]. D Anguita et al. worked on the hardware friendly approach of SVM to classify activities from HAR dataset and achieved average precision of 89%[10]. HAR smartphone data was collected and classified using fusion of different classifiers by A Bayat et al. with performance measure of 91% accuracy[16]. J Usharani et al. used KNN and Clustered KNN classifiers to recognize the activity for the smartphone HAR data, maximum accuracy achieved was 92% using Clustered KNN[17]. DN Tran performed HAR task using SVM, and obtained 89% accuracy.
A plenty feature selection techniques have been proposed and studied in the literature. Survey of feature selection methods was presented by G Chandrashekhar et al. with categorization as filters, wrappers and embedded methods[3]. Incremental feature selection approach was introduced by H Liu to reduce computational cost of the wrapper[21]. Correlation based multivariate feature selection method was proposed by M A Hall[18]. H Peng et al. introduced a new mutual information based filter feature selection method and proposed a hybrid feature selection approach with mutual information based filter and incremental selection based wrapper[20]. Reranking based feature selection method was discovered by P Bermejo[22]. L. Cervante et al. exploited Binary Particle Swarm Optimization (BPSO) technique for feature selection purpose[4]. A hybrid feature selection method was proposed by I Jain et al. with correlation based filter and BPSO based wrapper[8].

3. Proposed Frameworks

The main objective of this work is to build a classification model which can identify the given HAR task and categorize it into one the classes with optimal number of features, high accuracy and less computational cost.
Since the dimensionality of HAR dataset is very large, there is a possibility of irrelevant, redundant or noisy features. Hence, it is necessary to perform feature selection on HAR data to improve prediction performance and avoid overfitting of the classification model.
So, for performing optimal feature selection, two frameworks are proposed. Framework 1 is a composition of wrapper 1 and wrapper 2, while framework 2 comprises of filter followed by wrapper. Both the frameworks contain two stages that are executed in consecutively. The first stage of framework 1 uses SVM-RFE as wrapper 1, whereas the stage 1 of framework 2 uses Random Forest classifier based filter. Stage 2 is common for both frameworks where BPSO is used as wrapper.

3.1. Feature Ranking Using Support Vector Machines

3.1.1. Support Vector Machines

Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection. SVM tries to find optimal hyperplane which can separate input data variables into their respective classes. This is the minimization problem[3] as shown in (1). Decision function D(x) predicts label for unseen data.

Algorithm SVM-train:

Inputs: Training examples $\{x_1, x_2, \ldots x_k, \ldots x_l\}$ and class labels $\{y_1, y_2, \ldots y_k, \ldots y_l\}$ with l number of samples.

Minimize over α_k:

$$J=\sum_{hk} y_h\, y_k \alpha_h \alpha_k (x_h \cdot x_k+\lambda\delta_{hk})-\sum_k \alpha_k \tag{1}$$

subject to:

$$0\le\alpha_k\le C \text{ and } \sum_k \alpha_k\, y_k=0$$

Outputs: Parameters α_k.

The summations run over all training patterns x_k that are n dimensional feature vectors, $x_h \cdot x_k$ denotes the scalar product, y_k denotes the class label, δ_{hk} is the Kronecker symbol ($\delta_{hk} = 1$ if h = k and 0 otherwise), and λ and C are soft margin parameters.
Decision function can be given as:

$$D(x)= w\cdot x + b \tag{2}$$

$$\text{with } w=\sum_k \alpha_k\, y_k \text{ and } b = \langle y_k - w\cdot x_k\rangle$$

Where, w is the weight vector for n dimensional features and b is biased average over all dimensions.

3.1.2. SVM Recursive Feature Elimination (SVM-RFE)

SVM-RFE is application of Recursive Feature Elimination (RFE) [6] approach for feature selection. In SVM-RFE, low ranked features are recursively eliminated based upon the weight magnitude given by SVM as ranking criterion.

Let s be selected features subset and we need to find top k features, s is initialized to set of complete features.

Algorithm follows following steps:

1. Train the SVM using input samples and evaluate weight vector with feature subset s.
2. Compute score of the features according to the evaluation criteria and rank them according to the score. Evaluation function for SVM-RFE can be given as,

$$c_i=(w_i)^2, \text{ for all } i \tag{3}$$

Here w_i is weight magnitude of feature i and c_i is the score for feature i.

3. Find out one or more features having least score and remove those features from s.

Algorithm continues iterating until top k features are retained in s. Output of the algorithm is ranked feature list r consisting of top k features.

3.2. Feature selection using Binary Particle Swarm Optimization (BPSO)

BPSO is binary extended version of Particle Swarm Optimization (PSO)[5], which is able to deal with search space having discrete values.
BPSO uses equations (4) and (5) to update value of velocity[4]. BPSO differs in type of values generated for the next position of the particle, it generates only binary values. It uses sigmoid function to restrict values to 0 or 1. BPSO finds optimal solution by updating current position depending upon velocity value as per the following equations:

$$x_{id}^{t+1}=x_{id}^{t}+v_{id}^{t+1} \tag{4}$$

$$v_{id}^{t+1}=w*v_{id}^{t}+c_1 * r_{1i}*\left(p_{id}-x_{id}^{t}\right)+c_2 * r_{2i}*\left(p_{gd}-x_{id}^{t}\right) \tag{5}$$

Where, current position of the particle i is represented by x_{id}^{t}, d∈D denotes the d^{th} dimension in the search space, dimensionality of features is represented as D, t denotes the iteration number in the search process. c_1 and c_2 are acceleration constants. w is inertia weight. r_{1i} and r_{2i} are random values uniformly distributed in

[0, 1]. p_{id} and p_{gd} represents the elements of local best and global best in the d^{th} dimension.
BPSO restricts values of x_{id}, p_{id} and p_{gd} to 0 or 1 using sigmoid function as given below:

$$x_{id}=\begin{cases} 1, & \text{if rand()} < s(v_{id}) \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Where, $s(v_{id}) = \frac{1}{1+e^{-v_{id}}}$

Sigmoid function $s(v_{id})$ is used to distribute velocity v_{id} value in range (0, 1). rand() is random number generated selected from a uniform distribution in [0,1][4].
BPSO can be used for feature selection problem, as feature selection problem is discrete in nature. Position of the particle is represented in binary where, 1 represents feature considered and 0 represents not considered. For feature selection problem, p_{id} of particle i is evaluated using predictor performance of the particle as the evaluation measure. In proposed framework we have used hold out method of validation to compute predictor performance.
In each iteration t, particle i is evaluated using evaluation function and local best p_{id} and global best p_{gd} values are updated using (4) and (5). Equation (6) and (7) restrict position value to binary using sigmoid function. After completion of t iterations, particle associated with p_{gd} value gives near optimal feature subset as output.

3.3. Feature importance by Random Forest

Random forest (RF)[11] classifiers are very popular learning algorithms because of their performance for high dimensional data. The randomly split training samples (with replacement) are trained using different decision trees for each split, that results into forest of decision trees called Random Forest.
An important advantage of Random Forest is it can evaluate importance of each feature (i.e., how much it is contributing while predicting particular class[12]). Random Forest uses gini index to decide or split on particular node. Each node represents a feature, hence importance score of each feature can be computed based upon gini index[13]. Gini index Gini(v) at node v gives impurity at node v as:

$$\text{Gini}(v)=\sum_{i=1}^{I} f_i(1-f_i) \tag{8}$$

Where f_i is the fraction of class i records at node v.
Information gain of feature X_i needs to be calculated to decide whether or not to split on node v, it can be calculated as:

$$\text{gain}(X_i,v)=\text{Gini}(X_i,v)-\left(W_L \text{Gini}(X_i,v^L)+W_R \text{Gini}(X_i,v^R)\right) \tag{9}$$

where $\text{Gini}(X_i,v)$ represents impurity at node v, v^L and v^R represents left and right child of node v respectively. W_L and W_R are fractions of samples assigned to v^L and v^R respectively. Feature which gives highest reduction in impurity is considered for splitting finally.
Next importance score of the feature will be calculated using gain as shown below:

$$\text{Imp}_i=\frac{1}{n_{tree}}\sum_{k\epsilon S_{X_i}} \text{gain}(X_i,v) \tag{10}$$

Where n_{tree} represents number of trees in the forest and $k\epsilon S_{X_i}$ represents set of split nodes[13].

The features are ranked according to the importance score to filter out top k features.

3.4. Framework 1: SVM-RFE BPSO

The Framework 1 is a wrapper followed by wrapper hybrid model where SVM-RFE is used as wrapper 1 and BPSO based wrapper is used as wrapper 2.
When a set of features are supplied as input the framework, SVM-RFE ranks all the features based upon SVM weightage score and the features which are having a score above certain threshold t are selected and supplied as input to the iterative BPSO based wrapper for further optimization.
The framework's execution is described in the following steps:

1) Step 1: A dataset with D features is supplied as input to stage 1 of the framework.
2) Step 2(**stage 1- SVM-RFE**):
SVM-RFE algorithm is used as wrapper 1 whose functionality is as follows,
a) A supplied input is taken and the SVM weight score for all the features is computed by SVM-RFE.
b) In the next step, the SVM-RFE eliminates the features having least SVM weight score (i.e. less than specified threshold) that is computed using (3) recursively.
c) Thus obtained k features are supplied as input to the stage 2.
3) Step 3(**stage 2- iterative BPSO**):
Stage 2 uses iterative BPSO as wrapper 2 which takes the k number of features supplied from stage 1 as it's particles.
Particle is random set of features selected that can be represented as $p=\{x_1, x_2, x_j, \ldots, x_k\}$,where x_j represents j^{th} feature among k features.. BPSO uses a collection of particles called as swarm of particles, which can be represented as $S=\{p_1, p_2, p_j, \ldots, p_k\}$.

repeat until (*condition 1* or ***condition 2*** **)** **//repeat 1**
/*
***condition 1*:** fitness value of p_{gd} of i^{th} iteration is less than $(i-1)^{th}$ iteration
***condition 2*:** number of features considered by p_{gd} of i^{th} iteration is greater than $(i-1)^{th}$ iteration for the same fitness value of p_{gd}
***/**
repeat until (number of iterations is equal to t**)** **// repeat 2**
for all the particles in the swarm **do**
1. Construct dataset with reduced dimension k', where k' is the number of features considered by the particle p_i
2. Train the classifier with reduced data using hold out method to find out performance of the particle (i.e. fitness value of the particle)
3. Update values of x_{id}, v_{id}, p_{id} and p_{gd} according to equation (4), (5), (6) and (7).
end for
end repeat **// repeat 2**
return (p_{gd} and it's fitness value**)**
end repeat **// repeat 1**
return (optimal feature subset)

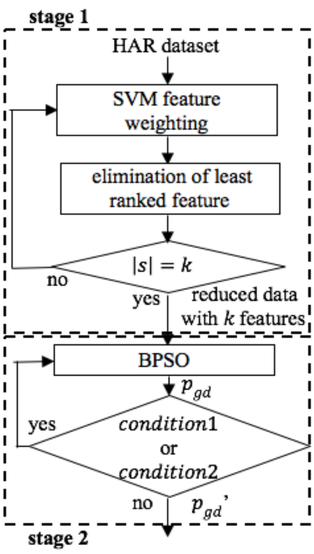

Fig. 1. framework 1: SVM-RFE BPSO

4)Step 4:

Return optimal feature subset

Fig. 1 visualizes the flow of the framework 1. In stage 1, supplied features are stored in feature subset s. Least ranked features in feature subset *s* are recursively eliminated using SVM-RFE till k number of features are retained in s.

In stage 2, reduced data with k features (given by stage 1) are provided as input to BPSO. BPSO outputs the particle which is having maximum fitness value (i.e. p_{gd}). Unless and until condition 1 or condition 2 is satisfied for p_{gd} element, BPSO continuously optimizes the p_{gd} value to achieve p_{gd}'.

Once the loop is terminated, p_{gd}' will have the best particle given by stage 2, which can be considered as optimal feature subset.

3.5. Framework 2: Random Forest BPSO

Framework 2 is another hybrid model which also contains two stages where filter based technique is used in stage 1 and wrapper based technique is used in stage 2.

The working strategy of framework can be described as follows,

1) Step 1:Complete dataset with all features is supplied as input to the stage 1

2) Step 2(**stage 1- Random Forest**):

Random Forest based filter technique is used as filter over here in the stage 1, whose functionality is described as follows,

a. Random Forest computes the scores of all features and ranks them based upon the average gini index given by different decision trees of the random forest.

b. It has been found from our experiments that k number of optimal features can be used as input to stage 2 (where k=300).

3) Step 3(**stage 2- iterative BPSO**):

Stage 2 of framework 2 is similar to stage 2 of framework 1 as discussed in the earlier subsection.

In figure 2, framework 2 is depicted with stage 1 containing Random Forest based filter and stage 2 as BPSO based wrapper. Stage 2 is same for both the frameworks.

Stage 1 of figure 2 takes the HAR data and with the help of Random Forest it ranks all the features of the data based upon gini index score. From thus obtained ranked features, least ranked features are eliminated and top *k* features having maximum score are provided to stage 2.

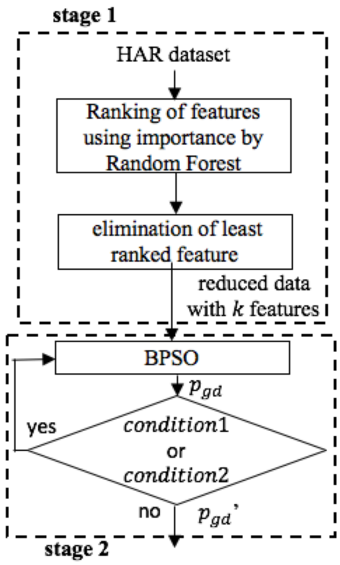

Fig. 2. framework 2: Random Forest BPSO

Stage 2 of framework 2 is same as that of framework 1 and will give p_{gd}' element with optimal feature subset as an output.

The figure 3 depicts the block diagram of the prediction model which uses the optimal features obtained from either of the framework.

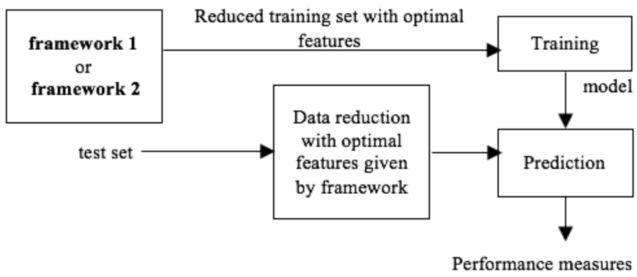

Fig. 3. Block diagram of Prediction model with optimized features

First, one of the proposed hybrid framework is chosen and executed by supplying the complete HAR dataset to obtain the reduced optimal feature dataset. Using thus obtained dataset, a classification algorithm is trained and a model is built.

From the given test set, optimal feature reduction is done manually based upon the optimal features given by one of the framework that is used to achieve reduced feature dataset.

Thus obtained reduced test data is supplied to the trained model and its performance is measured with the help of F score and recall.

The experiments have been conducted exhaustively by using different classification algorithms and by tuning their various parameters.

4. Dataset Description

The dataset used in these experiments is Human Activity Recognition (HAR) dataset which is publicly available at UCI Machine Repository[15].

- The data was collected using a smart phone device Samsung S II, which is having inbuilt gyroscope and accelerometer tools**.** An experiment was performed on a group of 30 volunteer who were wearing this smart phone on their west. Experiments were video recorded and labelled manually.

- HAR dataset consists of 10299 samples that are divided into two files. The training file consists of 7352 samples and a separate test file consisting of 2947 samples. The ratio between train and test data is 71:29. There are total six classes in the whole data which are shown in Table 2
- Training and test data was obtained by different volunteer groups.
- Features of this dataset are obtained from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ. Other signals were derived by splitting or modifying the raw signals.
- All the derived signals were used to estimate variables of feature vector such as mean, standard deviation, signal magnitude area, etc.
- Dataset has total 561 features which are of numerical type.

Table 1: class labels and their respective activities

Class label	Activity
A	Standing
B	Sitting
C	Laying
D	Walking
E	Walking downstairs
F	Walking upstairs

5. Results and Discussion

The HAR data used for the experiments had the 561 features which has been reduced to 79 features by our proposed framework-1 without compromising on the recall and f-score of the used classifiers. The feature reduction is represented is represented in Table 2.
In fact, it would be quite interesting to know that, with the help these optimal features given by the proposed hybrid framework the results of almost all classes has improved to desirable extent. As we can see from the Table 2, though the CFS method has given very less features of 57, it is not able to give good f-score and recall value for most of the classes of the HAR dataset. On the other side, the two proposed hybrid frameworks has given significantly less features without reducing the recall and f-score of all the classifiers. Though there isn't much difference between the results of the proposed frameworks, the framework-1 has given slightly high values compared to the framework-2 across all the classes of the dataset, by using only two additional features compared to framework-2, which can be considered as negligible compared to the total number of features reduced.
Vast experiments have been conducted using different classifiers, with all the five types of datasets (i.e., full dataset, reduced feature dataset obtained from CFS method, reduced feature dataset obtained from BP method and using reduced optimal number of feature set obtained from both the proposed frameworks. Different classifiers were also used for the experiments. i.e., Linear SVM, Random Forest, Decision Tree and AdaBoost technique.
All the classifiers were trained using different data sets individually and a model is built. Using this model the learning capability of the model is assessed using relevant test sets (i.e., it is ensured that the total number of features used for training are there in the test data also). The results of various classifiers for different classes is shown from Table 3 to Table 6.
It could be observed from the Table 3 to Table 6 that among all the classifiers, the linear SVM has given good measure of score with respect to recall and f-score. Though Random Forest performed the second best, the difference between the two scores is significantly high for the framework -1.
The other techniques which are used, like Adaboost and decision tree, nowhere stand near the linear SVM score. Hence, it is observed that Linear SVM has given a good result of above 91% for all the classes of the HAR dataset. Which is a benchmark achievement for this dataset with optimal number of features and for all classes.
It was found from the literature survey that, nobody has achieved an accuracy of above 90% for all the classes of the HAR data till now. Hence, in this perspective the result which we obtained using least number of features given by our proposed frameworks can be considered as a significant contribution in this domain.

Table 2: Number of features selected by different feature selection approaches

Feature selection method	All	F1	F2	CFS	BP
Number of features selected	561	79	77	57	282

Table 3: Recall and F1 score of all the classes for Linear SVM classifier (in %)

Class	Recall					F1 score				
	all	*F1*	*F2*	*CFS*	*BP*	*all*	*F1*	*F2*	*CFS*	*BP*
A	100	**100**	99.5	90.8	99.7	98	**99.3**	99.3	86.3	99
B	96.3	**99.7**	99.7	78.4	99.1	97.2	**99.4**	98.9	83.2	98.9
C	98.3	**98.5**	98	100	98.5	99.1	**99.2**	98.8	100	99.2
D	89.2	**91**	90.4	99	89.6	92.7	**93.6**	93	94.2	93.2
E	96.9	**96.8**	96.2	92.9	97.3	94.2	**94.4**	94.4	93.5	94.2
F	100	**99.6**	100	86.8	100	99.8	**99.8**	99.6	91	100

Table 4: Recall and F1 score of all the classes for Random Forest classifier (in %)

Class	Recall					F1 score				
	all	*F1*	*F2*	*CFS*	*BP*	*all*	*F1*	*F2*	*CFS*	*BP*
A	96.9	95.1	96.3	87	96.5	92.6	88.5	92	84	92.2
B	89.5	83.6	88.5	78	87.8	89	85	88.7	81.2	86.8
C	84	81.9	83.3	100	81.9	89.5	87.8	88	100	87.8
D	90	87.5	89.2	97.6	86.5	90.6	88.1	89.2	92.5	86.8
E	92.1	89.4	90.4	84.5	88.3	91.5	89.2	90.3	90.3	88
F	100	100	99.8	87.9	99.8	100	99.8	99.8	88	99.9

Table 5: Recall and F1 score of all the classes for Decision tree classifier (in %)

Class	Recall					F1_score				
	all	*F1*	*F2*	*CFS*	*BP*	*all*	*F1*	*F2*	*CFS*	*BP*
A	89.1	88.7	87.5	82.1	91.1	86.4	85.1	84.8	81.2	88.3
B	79.6	79.1	78.5	78.2	81.1	80.7	80.3	78.7	79.2	79.5
C	83.8	79.7	82.8	100	81.4	85.5	83.3	85.7	100	86.5
D	76.1	78.8	78.4	90.7	74.5	79.6	77.7	81.6	87.9	79.1
E	86	79.6	87.4	73.3	87.2	82.7	80.1	84.3	79.8	82.7
F	100	99.8	99.8	80.3	100	100	99.8	99.9	77.3	100

Table 6: Recall and F1 score of all the classes for Adaboost M1classifier (in %)

Class	Recall					F1_score				
	all	*F1*	*F2*	*CFS*	*BP*	*all*	*F1*	*F2*	*CFS*	*BP*
A	88.7	87.5	88.9	89.1	91.7	85.2	85	85.5	86.2	87.8
B	78.3	82.3	76.6	80.9	80.4	79.2	81.3	77	83.9	80.1
C	81.6	80.7	79.7	100	82.1	84.5	84.6	82.5	100	87.1
D	76.5	78.2	76.3	98	75.1	79.7	77.7	81.2	94.3	99.1
E	83.7	78.9	89.2	86	86.4	82.6	79.3	84.5	91	82.5
F	100	99.8	99.8	90.9	100	100	99.9	99.9	90.1	100

6. Conclusion

For the datasets like HAR data, which contain huge number of features and takes more time and memory to learn for different classifiers, the proposed hybrid frameworks come needy to serve the classifiers with optimal number of features there by reducing the computational time and memory and also gives significant improvement in the recall and f-score. Hence, it can be concluded that, with the help of the proposed hybrid frameworks, one can achieve an optimal number of features which contribute to the performance enhancement of the classifiers. For the HAR data, it can be observed that Linear SVM has given significantly high result for all the classes.

References

[1] JL Reyes-Ortiz,A Ghio,X Parra, D Anguita,J Cabestany, A Catala, "Human Activity and Motion Disorder Recognition: towards smarter Interactive Cognitive Environments," InESANN Apr 2013.

[2] P CasalE, O Pujol, P Radeva, "Human activity recognition from accelerometer data using a wearable device,". InIberian Conference on Pattern Recognition and Image Analysis, 8 June 2011, Springer, Berlin, Heidelberg, pp. 289-296.

[3] G Chandrashekar,F Sahin, "A survey on feature selection methods," Computers & Electrical Engineering, vol. 40, pp. 16-28, 1 Jan 2014.

[4] L Cervante L, B Xue B, M Zhang M, L Shang L, "Binary particle swarm optimisation for feature selection: A filter based approach," InEvolutionary Computation (CEC), IEEE Congress , 10 Jun 2012 , IEEE, pp. 1-8

[5] J Kennedy, "Particle swarm optimization," InEncyclopedia of machine learning 2011, Springer, US, pp. 760-766.

[6] I Guyon,J Weston,S Barnhill,V Vapnik, "Gene selection for cancer classification using support vector machines,"Machine learning vol. 46, pp. 389-422,1 Jan 2002.

[7] Z Feng, L Mo, M Li, "A Random Forest-based ensemble method for activity recognition," InEngineering in Medicine and Biology Society (EMBC), 37th Annual International Conference of the IEEE , IEEE, 25 Aug 2015,pp. 5074-5077.

[8] I Jain, VK Jain, R Jain, "Correlation feature selection based improved-Binary Particle Swarm Optimization for gene selection and cancer classification," Applied Soft Computing vol. 62, 1 Jan 2018, pp. 203-215.

[9] J Wang, G Shan, X Duan, B Wen, "Improved SVM-RFE feature selection method for multi-SVM classifier," InElectrical and Control Engineering (ICECE), IEEE, 16 Sep 2011,pp. 1592-1595.

[10] D Anguita, A Ghio, L Oneto, FX Llanas Parra, JL Reyes Ortiz, "Energy efficient smartphone-based activity recognition using fixed-point arithmetic," Journal of universal computer science vol. 19, 2013, pp. 1295-314.

[11] L Breiman, "Random forests," Machine learning vol. 45, 1 Oct 2001, pp. 5-32.

[12] R Díaz-Uriarte, SA De Andres, "Gene selection and classification of microarray data using random forest," BMC bioinformatics vol. 7, Dec 2006, pp. 3.

[13] MT Uddin, MA Uddiny, "A guided random forest based feature selection approach for activity recognition," InElectrical Engineering and Information Communication Technology (ICEEICT), IEEE, 21 May 2015, pp. 1-6

[14] B Romera-Paredes, MS Aung, N Bianchi-Berthouze, "A one-vs-one classifier ensemble with majority voting for activity recognition," InESANN 2013 proceedings, 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning 11 Nov 2013, pp. 443-448.

[15] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz, "A Public Domain Dataset for Human Activity Recognition Using Smartphones," 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.

[16] A Bayat, M Pomplun, DA Tran, "A study on human activity recognition using accelerometer data from smartphones," Procedia Computer Science vol. 34,1 Jan 2014, pp. 450-457.

[17] J Usharani, U Sakthivel, "Human Activity Recognition using Android Smartphone," International Journal of Advanced Networking & Applications (IJANA) ISSN: 0975-0282.

[18] MA Hall, "Correlation-based feature selection of discrete and numeric class machine learning", In Proceedings of the Seventeenth International Conference on Machine Learning (ICML '00), Pat Langley (Ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2000, pp. 359-366.

[19] DN Tran, DD Phan, "Human activities recognition in android smartphone using support vector machine," InIntelligent Systems, Modelling and Simulation (ISMS), 7th International Conference, IEEE, 25 Jan 2016, pp. 64-68.

[20] H Peng, F Long, C Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," IEEE Transactions on pattern analysis and machine intelligence vol. 27, Aug 2005, pp. 1226-1238.

[21] H Liu, R Setiono, "Incremental feature selection," Applied Intelligence vol. 9, 1 Nov 1998, pp.217-230.

[22] P. Bermejo, L. de la Ossa, J.A. Gámez, J.M. Puerta, "Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking," Knowledge-Based Systems vol. 25, 2012, pp.35-44.