Journal of Theoretical and Applied Information Technology
15th October 2023. Vol.101. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A DATA DRIVEN EMBEDDED FEATURE SELECTION AND ENSEMBLE CLASSIFIER FOR HUMAN ACTIVITY RECOGNITION

S. ANTHONISAMY[1], P. PRABHU[2]

[1] Research Scholar (Part-Time), Department of Computer Applications, Alagappa University, Karaikudi, India.
[2] Associate Professor in Information Technology, Centre for Distance and Online Education, Alagappa University, Karaikudi.630003, Tamil Nadu, India;
Email: [1]antony2006reegan@gmail.com; [2]prabhup@alagappauniversity.ac.in

ABSTRACT

Human Activity Recognition (HAR) is a complex yet important to assess or monitor humans in the area of healthcare. Itrecognizes activities from a series of observations on the actions of subjects and environmental conditions. HAR research is the basis of many applications including video surveillance, health care, and human-computer interactions because it provides information about the identity of a person, their personality, and psychological state which is difficult to extract. The human ability to recognize another person's activities is one of the main subjects of study of the scientific areas of computer vision and machine learning. For learning about human activities, HAR can benefit from the usage of Machine Learning Techniques (MLTs) where accuracy is crucial. Individuals have been interested in recording human activities for the past decade but important issues need to be addressed in order to fully utilize technology in human activity information. Though many studies have investigated HARs, there is a need for higher accuracy in classifications and an impending need to find most suitable HAR features. Hence, this work proposes the A Data Driven Embedded Feature Selection and Ensemble (ADDEFE) Machine Learning Model for Human Activity Recognition technique. This research work uses an embedded feature selection and ensemble classifier for recognizing human activity from sensor-based data as these kinds of classifiers have achieved better performances with the use of weighted combinations. The proposed ADDEFE(Grid+SVM) method gives accuracy of 93.4 and 96.4 respectively for WISDN and UCI datasets. Theproposed ADDEFE(Grid+Random forest) method gives accuracy of 91.6 % accuracy for WISDN dataset.Hence the experimental results of the proposed work outperform with traditional methods.

Keywords: *Decision Tree, Embedded Method, Ensemble Classification, Human Activity Recognition, Prediction, Sensors.*

1. INTRODUCTION

Human Activity Recognition (HAR) plays a significant role in human-to-human interaction and interpersonal relations. Because it provides information about the identity of a person, their personality, and psychological state, it is difficult to extract. The human ability to recognize another person's activities is one of the main subjects of study of the scientific areas of computer vision and machine learning. When attempting to recognize human activities, the kinetic states of a person have to be determined for a computer to efficiently recognize this activity. Human activities, such as "walking" and "running," arise very naturally in daily life need to be recognizes. Humans execute a variety of activities which are highly diverse. HAR is the study concerned with the ability to recognize intricate human activities and interpret them. HAR can be automated using sensor generated data [1]. HAR is critical to healthcare monitoring applications and is used in a wide range of domains, including ubiquitous computing, HCIs (Human-Computer Interactions), and surveillance. Human/object detections [2], physical sciences [3], computer engineering [4], and natural sciences have all found HAR to be beneficial. WSNs (Wireless Sensor Networks) and the production of wearable body sensors have advanced the use of technology in customized healthcare. By monitoring data provided by wearable or external devices, HARs may detect human behaviours of subjects. This activity recognition is essential to humanity since it records people's behaviors with data that allows computing systems to monitor, analyze, and assist their daily life. There are two mainstreams of human activity recognition systems namely video-based systems and sensor-based systems. Video-based systems use cameras to take images or videos to recognize people's behaviors while Sensor-based systems utilize on-body or ambient sensors to dead reckon people's motion details or log their activity tracks. Considering the privacy issues of installing cameras in our personal space,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

sensor-based systems have dominated the applications of monitoring our daily activities. Besides, sensors take advantage of pervasiveness. Thanks to the proliferation of smart devices and Internet of Things, sensors can be embedded in portable devices such as phones, watches, and nonportable objects like cars, walls, and furniture. Sensors are widely embedded around us, uninterruptedly and non-intrusively logging human's motion information. Accelerometer and gyroscope sensors orientation are depicted in Figure 1.

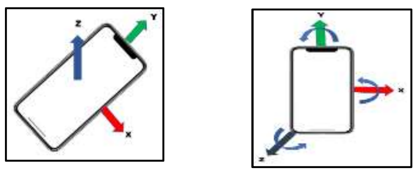

Figure.1. Accelerometer and gyroscope sensors orientation

HAR can take advantage of this sensory technology to monitor HDAs (Human Daily Activities) for cognitive assistance, safety and emergencies non-intrusively [6]. HDAs can be used to assess human physical activities from data generated by accelerometers or gyroscopes or vital sensors [7]. Application of MLTs and DMTs (Data Mining Techniques) [8] on such data is an interesting area of research as they have the ability to learn activities and create models for classifications [9]. HAR exploitations can also benefit pervasive applications and pervasive computing [10]. Elderly people can have proactive assistance where care takers use sensors to track and analyze activities on a day to day basis [11]. HAR can also assist lonely elderly people at homes when it becomes challenging to recognize their HDAs. Thus, HAR is a rapidly expanding topic of research, particularly in the case of smart home health care systems that aid in patient rehabilitation [12]. The growing quantity of wearable sensors and cell phones has prompted more HAR research. Information securing and pre-processing, information division; incorporate extraction and determination, grouping or order, and performance assessment are all segments of a typical HAR framework with various wearable sensors [13]. Many MLTs have been employed in HAR. However, this field still faces many technical challenges. One of the foremost challenge is feature extractions. Activity recognition is a classification task so it shares a common challenge with other classification problems which is feature extraction. For sensor based activity recognition, feature extraction is more difficult because there is inter-activity similarity. Different activities may have similar characteristics (e.g., walking and running). Therefore, it is difficult to produce distinguishable features to represent activities uniquely. Most MLTs proposed rely on feature extraction techniques including time-frequency transformations, statistical approaches and symbolic representations. However, the features extracted are carefully engineered and heuristic. There were no universal or systematical feature extraction approaches to effectively capture distinguishable features for human activities. The use of ensembles have turned the tide in MLT schemes. Ensemble-based MLTs integrate the perspectives of several learners to achieve a superior result. Theoretically, ensemble classifiers perform better than single classifiers [14]. The ensemble classifiers are generated using a separate feature subset selected randomly from the original feature set and rely on the majority vote of a pre-defined number of learners. Strong classifiers, which are extensively utilized in many applications, may be employed with ensemble learning approaches [15]. Though many studies have investigated HARs, there is a need for higher accuracy in classifications and an impending need to find most suitable HAR features and physical activity predictions.

The HAR is essential due to the automatic detection of various human physical activities for human safety and well being in the assistive technologies.

This HAR is a challenging issue still due to the lack of sensor which does not affect the human being especially children and adults, lack of adequate techniques for preprocessing and modeling the data.

This HAR problem can be solved by proposing new methods using machine learning techniques and data collected from various sensor present in the form of smart-phones and wearable sensor.

Hence, this work proposes the Data Driven approach using Embedded Feature Selection and Ensemble learning (ADDEFE) method for HAR Feature selection. In this work, the hypothesis is set to identify relevant physical human activities with high accuracy.

2.HAR RELATED STUDIES

HAR is a vibrant study topic that has attracted a lot of attention in recent years because of the widespread usage of wearable devices in smart settings [16]. Because of the rising expense of healthcare, the ageing population, and the need to support "ageing in place," sensor-based techniques

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

have become common in ubiquitous computing, WSNs, and healthcare. Nonetheless, these techniques have drawbacks to consider, such as ethical issues, comfort, privacy invasion, and obtrusiveness. Many senior residents in AAL situations, for example, have been found to be uncomfortable and unwilling to constantly embrace the usage of body-worn sensors, as well as voicing reservations about the installation of video-based surveillance [17]. As a result, binary sensors deployed in the surrounding environment are becoming increasingly promising for long-term activity monitoring in the ubiquitous computing domain, as these devices eliminate the privacy concerns associated with other approaches to HAR while also being non-invasive to smart home inhabitation. In a recent HAR research [18], binary sensors were utilized to detect nine HDAs done by four smart home residents, such as cleaning, cooking, and sleeping. Motion detectors built inside or connected to smart appliances were among the sensors used. On/off states for cleaning appliances, such as a vacuum, ceiling lights, cooking heaters, TV, and PC, as well as OPEN/CLOSE states for kitchen equipment like the fridge, were included. The selected classifier was a Random Forest model, which obtained 68 percent accuracy; however, the researchers noted that this percentage might be higher if more effective approaches were used Furthermore, binary sensors were used in a home monitoring scenario in [19] to detect four fundamental activity classes: resting, making a meal, eating, and moving from bed to toilet. For activity categorization, DCNNs (Deep Convolution Neural Networks) were developed, in which binary sensor data collected by four door sensors and 31 passive infrared (PIR) motion sensors were transformed into representative activity pictures. The pictures were utilized to train the DCNN model, which had a recognition accuracy of 99.36% for the four HDASs seen in the research. Although this method functioned well, a larger number of activity classes might have been investigated. Another research [20] looked at the possibility of HDAS recognition in a smart home scenario utilizing neural networks. RNNs (Recurrent Neural Networks) and CNNs (Convolution Neural Networks) were used in experiments to recognize behaviours such as cooking, bathing, and sleeping. The data collected from the deployment of binary sensors such as pressure sensors, reed switches, float sensors, and PIR motion sensors was utilized to train the different neural network classifiers, with the RNN and CNN models exhibiting the best results During comparisons, and substantially outperformed other popular classifiers, with accuracies of 89.8% and 88.2%, respectively. In order to describe the activities being detected, HAR requires a feature extraction step in which a set of features is chosen as inputs to a classification model. Various state-of-the-art characteristics for HAR have been identified, however these differ based on the sensors used to collect activity data. Extraction of the maximum, minimum, and range characteristics is useful in distinguishing between activities that include movements of varied ranges in the domains of wearable technology that provide accelerometer data [21]. Calculating the signal magnitude area (SMA) of an accelerometer signal has also been found to be useful in distinguishing between static and dynamic activities [22]. Prabhu et.al, proposed business intelligence methods for e-commerce recommender system using unsupervised classification and supervised learning[27-29]. The general issues in prior proposed schemes can be detailed as: (1) Activity recognitions are classification tasks and hence face the same issues classifiers face in feature extractions. Sensor based activity recognition, feature extractions are more difficult because of inter-activity similarities and producing distinguishable features to represent activities uniquely has been a issue. (2) Training and evaluation of learning techniques require large annotated data samples. However, it is expensive and time-consuming to collect and annotate sensory activity data. Therefore, annotation scarcity is a remarkable challenge for sensor-based activity recognition. The research design is based on machine learning technique and compared with conventional classifiers for performance benchmark.

3. PROPOSED ADDEFE METHODOLOGY

The main goal of this proposed method is to recognize human physical activities data driven smart phone sensor acquired data with high accuracy. This analysis has found potential applicability in surveillance tasks, gaming, assistive living and improvement general improvement in the quality of life [23].HAR systems typically follow a set of sequential actions namely sensor engineering, and use of MLTs for data analysis. The accuracy of any analysis is dependent on the right combination of features. While more efforts have been spent on automating model selections or tuning, less work has been done on automating feature engineering processes and in spite of the fact that features have an important role in the MLT

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

analysis. The features used in MLT model relates to the risk of over fitting, the model performance and the computational complexity [24]. For a model to be functional on an edge device it can therefore be relevant to minimize the number of features used by the model. At the same time, extracting features is often a time-consuming process that usually requires technical skills, domain knowledge and creativity [25]. For these reasons an automation of the process could be beneficial.

Feature engineering is the craft of extracting relevant patterns or variables from raw data to make it easier for a classifier to perform its task. Feature extraction is the process of extracting features from some input that may consist of either features or raw data. New features can be extracted from existing features by combining them into new representations. Automatic feature extraction often includes a feature selection algorithm.When using the term feature selection, the most common interpretation is the process of selecting a subset from a set of features or filtering where methods utilize the characteristics found in the data and independent of the MLT to be applied subsequently[26]. This work uses random forest feature selection for extracting relevant HAR features from raw data. Figure 2 depicts the architecture of the proposed ADDEFE Design methodology.

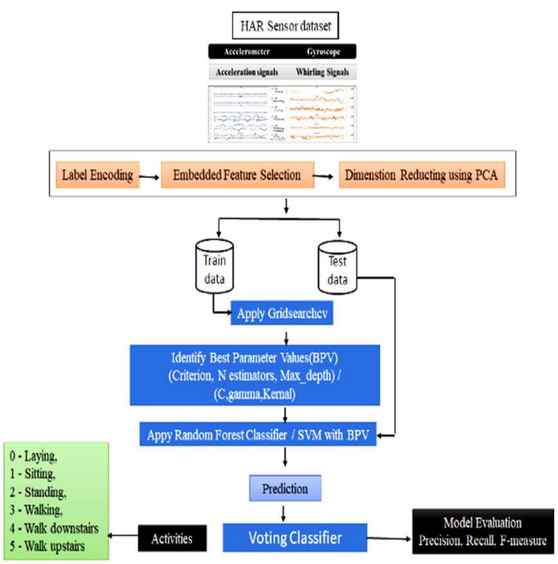

Figure. 2. ADDEFE Design

Support vector machines (SVMs) have been intensively studied and applied to practical problems in many fields of science and engineering and are one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, they are used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. SVM algorithm can be used for Face detection, image classification, text categorization, etc. SVMs have many merits that distinguish them from many other MLTs, including the nonexistence of local minima, the speed of calculation, and the use of only two tuning parameters. SVMs are used in cross validations as they are favorable when determining the tuning parameters. SVMs can estimate the bias of the excess errors in predictions.

A. ADDEFE Data Preparation

The initial operations executed on the HAR dataset's data aims to make the data suitable for further analysis where various techniques are used. This step uses normalization, eliminating non-numeric values and cleaning of corrupted data generally caused by possible hardware failures or problems while capturing data.

B. ADDEFE Feature Extraction

The proposed ADDEFE technique extracts features from raw input data without the need for repeated human intervention. A random RF (Random Forest) selection has been chosen as the embedding technique. At training, RF, an ensemble approach, creates a large number of individual DTs (Decision Trees). To create the final forecast, the mode of the classes for classification or the mean prediction for regression, the predictions from all trees are combined. The decrease in node impurity weighted by the chance of accessing that node is used to determine feature significance in RF. The number of samples that reach the node divided by the total number of samples is the node probability. For each DT, a node's importance using Gini Importance for only two child nodes (binary tree)based on Equation (1)

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \quad (1)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Where ni sub(j) - node j's importance, w sub(j) - weighted samples count which reach node j, C sub(j)- node j's impurity value, left(j) -child node from left split on node j and right(j)- child node from right split on node j. DTs compute feature's importance using Equation (2).

$$fi_i = \sum_{j:\text{node j splits on feature i}} ni_j / \sum_{k\in\text{all nodes}} ni_k \quad (2)$$

Where fi sub(i)- feature I's importance and ni sub(j)-node's importance. The values get normalized in the interval [0, 1] by dividing sum of feature importance values and based on Equation (3).

$$\text{norm } fi_i = fi_i / \sum_{j\in\text{all features}} fi_j \quad (3)$$

The total resulting feature importance for RFs is the average of all the trees where each trees's sum of feature importance is computed and divided by total trees count as depicted in Equation (4).

$$RFfi_i = \sum_{j\in\text{all trees}} \text{norm } fi_{ij} / T \quad (4)$$

Where RFfi sub(i)- I's importance of feature computed from RF trees, normfi sub(ij)-i's normalized feature importance in tree j and T -trees count. The proposed ADDEFE Framework uses three main steps namely data preparation followed by dimensionality reduction and extraction of important features as a part of its feature engineering process.

C. ADDEFE Dimensionality Reduction

In dimensionality reduction, the features which contain important HAR information are identified. This step is important for two reasons. It filters the data by eliminating sections that are unrelated to the recognition problem, while also decreasing the amount of data that has to be processed in later steps by extracting small segments. This is an essential element of ADDEFE since it helps overcome hardware restrictions by restricting the amount of data that can be stored and data taken for processing at each time step.

D. ADDEFE Classification model

In this work, Support Vector Machine classifier is used to test the performance. Gridsearch optimization is used to tune the hyper parameters of classifier. The voting classifier can be used to form ensemble classifier.

E. Algorithm ADDEFE

Input:
HAR dataset T;
Gyroscope data; Accelerometer data
Output:
Predicted human physical activities
Evaluation Measures :(P,R.F1 measure)

Begin
1. Handling missing values (if any)
2. Train and test split of data T
3. Feature selection using Random Forest
4. Dimension reduction using PCA
5. Apply optimization (Grid search/Random)
a. Identify best classification parameter values (BPV)
6. Apply BPV in Train data.
7. Prediction with best parameter values
8. Apply Voting classifier
9. Find the Best Voting Value
10.Evaluate prediction (Precision,Recall,F1-measure)
11.Draw Confusion Matrix
End

4. RESULTS AND DISCUSSION

This section displays stage wise experimental results of the proposed scheme executed using Python 3.9 on an AMD Athlon processor with 4 GB memory. The experiments were coded for the HAR Smart Phone Data Set obtained from Irvine's repository (University of California)[29]. The data set includes conventional activities like general walking, walking upstairs/downstairs, sitting, standing, and lying down, trapped at 50 Hz with the help of sensors of people between the ages of nineteen and forty eight. The x, y, and z accelerometer data (linear acceleration) and gyroscopic data (angular velocity) were recorded. Each participant executed the exercises two times with sensors attached to the left and right.

It takes a little more work to load the raw data. The three major signal types in the raw data are total acceleration, body acceleration, and body gyroscope. There are three data axes in each time step, indicating a total of nine variables. Furthermore, overlapping windows of 2.65 seconds of data were used to split each batch of data into 128 time steps. Each data row comprises 128 * 9 or 1,152 pieces, and these data windows correspond to engineering feature windows. This is a bit less than double the size of the 561 element vectors with superfluous data. As part of the loading data preparation, the windows and features were flattened into a single long vector. Figure 3 depicts the activities of subjects in a HAR dataset.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

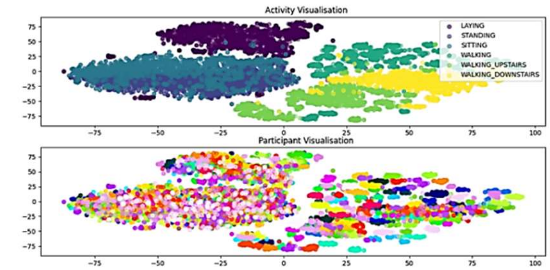

Fig. 3. HAR Subjects and their Activity Visualization

The Activity label distribution of the datasetare depicted in Figure 4.The 6 human physical activities are (laying, sitting, standing, walking, walking_down and walking_up) numbered 0,1,2,3,4,5 respectively.

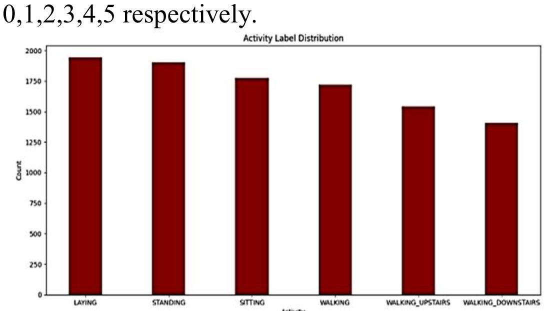

Figure.4. Activity label distribution of the dataset

The proposed scheme's selected features were evaluated after splitting the dataset into test and train datasets and a 30/70 split and encoding therm.

A. ADDEFE Feature Selection Evaluation

Finding a high-level representation is important to ensure the generalization capacity of the HAR system on large-scale data. The feature extractions were done based on Gini importance of features. Figure 5 depicts the computed importance of features.

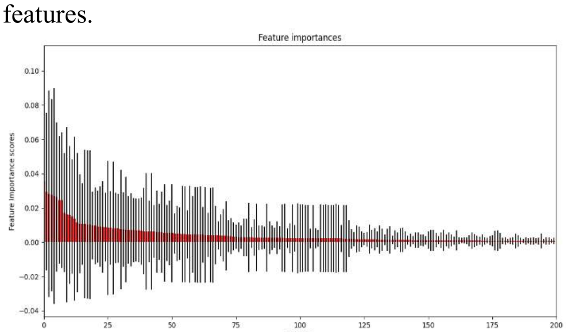

Figure.5. HAR Dataset Feature Importance

The proposed ADDEFE was used to select optimal features and 113 features resulted from the feature selection based on their importance.

B. Dimension Reduction using PCA

Dimensionality of data can always be a curse in predictive modeling as increasing number of input variables generates greater complexities. Hence, high-dimensionality statistics aim to reduce dimensionalities while visualizing data. Similar techniques can also be used machine learning to decrease classification or regression dataset's dimensions for achieving better training/prediction models. Thus, dimensionality reductions result in reducing the count of input characteristics where feature selections, use of linear algebra, and projections are the general approaches used. PCAs (Principle Component Analysis) are used in the proposed ADDEFE to reduce the dataset's dimensionality. Figure 6 depicts the ADDEFE PCA output based on cumulative variance.

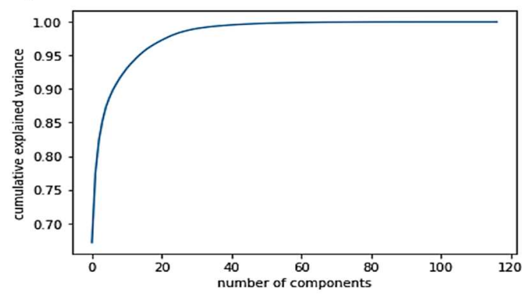

Fig.6. ADDEFE PCA output

C. ADDEFE Classification model results

The dataset was also classified with the proposed model's chosen feature sets from HAR Smart phone data set were classified by SVM. Figure 7 depicts the the classifications on the derived optimal feature subset of HAR Smart phone Data Set in terms of their cross validations.

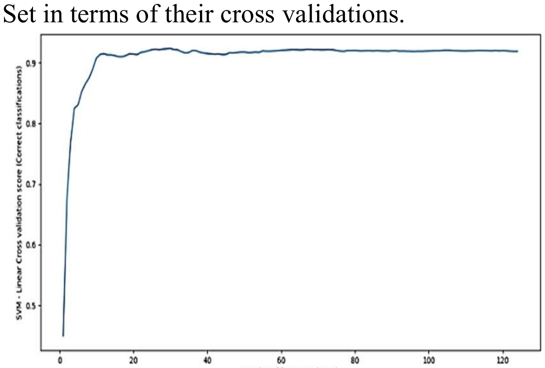

Figure.7. Cross Validation of ADDEFE Features with SVM

Table I showsthe results of the SVM classifiers along with other techniques on the Smartphone dataset.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table I. Accuracy of SVM classifier

Activity	Precision	Recall	F1-score	support
0	1.0000	1.0000	1.0000	537
1	0.9379	0.8921	0.9144	491
2	0.9079	0.9455	0.9263	532
3	0.9897	0.9143	0.9505	420
4	0.9439	0.9839	0.9635	496
5	0.9318	0.9575	0.9445	471
Measure	**Precision**	**Recall**	**F1-score**	**support**
Accuracy			0.9505	2947
Macro Average	0.9519	0.9489	0.9499	2947
Weighted Average	0.9512	0.9505	0.9504	2947

A confusion matrix, a performance measurement for machine learning classification problems was also used as performance measure. It is a table with different combinations of predicted and actual values where output can be two or more classes. The figures show the confusion matrix with and without normalization by class support size (number of elements in each class).Higher the diagonal values of the confusion matrix the better, indicating many correct predictions.
Figure 8 shows the confusion matrix of SVM classifier for smartphone dataset.

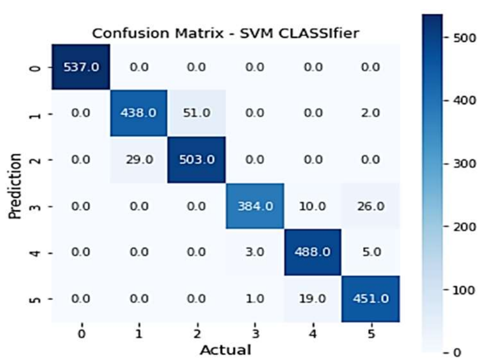

Figure 8. Confusion Matrix of SVM classifier

The original features in UCI smartphone dataset are 561. After applying feature selection technique the reduced number of features are 113. After applying Dimensionality Reduction using PCA, the reduced number of features became 82. Now the SVM classfier is applied to test the performance of the feature selection. The obtained train accuracy and test accuracy are 0.9392 and 0.9562 respectively.

Table II showsthe results of the ADDEFE+PCA+SVM classifiers along with other techniques on the Smartphone dataset.

Table II. Accuracy of ADDEFE+PCA+SVM classifier using smartphone dataset.

Activity	Precision	Recall	F1-score	Support
0	1.0000	1.0000	1.0000	537
1	0.9569	0.9043	0.9298	491
2	0.9192	0.9624	0.9403	532
3	0.9872	0.9167	0.9506	420
4	0.9602	0.9738	0.9670	496
5	0.9214	0.9703	0.9452	471
Accuracy			0.9562	2947
Macro Average	0.9575	0.9546	0.9555	2947
Weighted Average	0.9571	0.9562	0.9562	2947

Figure 9 shows the Confusion matrix of ADDEFE PCA+SVM classifier for smart phone dataset.

CONFUSION MATRIX PCA+SVM

Activities	Laying	Sitting	Standing	Walking	Walking Down	Walking Up
Laying	537	0	0	0	0	0
Sitting	0	444	45	0	0	2
Standing	0	20	512	0	0	0
Walking	0	0	0	385	6	29
Walking Down	0	0	0	5	483	8
Walking UP	0	0	0	0	14	457

Figure 9. Confusion matrix of ADDEFE PCA+SVM classifier for smartphone dataset

D.Evaluation of Proposed ADDEFE Method

The ADDEFE method selects features using random forest feature selection, reduce the dimension using PCA and optimize the SVM classifier using Grid Search optimization.

The Table.III shows the results of feature selection,dimension reduction and hyper parameter tuning.

Table.III The results of feature selection,dimension reduction and hyper parameter tuning.

No of Selected features using feature selection	113
Dimension reduced	82

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

features	
Best Parameters and Values	{'C': 200, 'class_weight': None, 'gamma': 0.001, 'kernel': 'rbf'}
Best Estimator	SVC(C=200, gamma=0.001)

Table IV shows the accuracy of of classfication of various physical activities using ADDEFE- PCA with grid search CV and SVM.

Table.IV Classfication accuracy of various physical activities using ADDEFE+ PCA with grid search CV and SVM for smartphone dataset.

Classifiers	Precision	Recall	F-Measure
K-NN	0.827	0.842	0.834
SVM	0.951	0.950	0.950
Decision Tree	0.839	0.903	0.870
Random Forest	0.907	0.906	0.906
ANN	0.915	0.914	0.914
Naive Bayes	0.893	0.904	0.898
ADDEFE+Grid+SVM	0.965	0.964	0.964

Table V shows the comparison of training and test accuracy of proposed method with other methods.

Table.V Comparison of training and test accuracy of proposed method with other methods.

Activity	Precision	Recall	F1-score	Support
0	1.0000	1.0000	1.0000	537
1	0.9667	0.8880	0.9257	491
2	0.9070	0.9718	0.9383	532
3	0.9831	0.9667	0.9748	420
4	0.9647	0.9919	0.9781	496
5	0.9742	0.9639	0.9691	471
Measure	**Precision**	**Recall**	**F1-score**	**Support**
Accuracy			0.9644	2947
Macro Average	0.9660	0.9637	0.9643	2947
Weighted Average	0.9652	0.9644	0.9643	2947

Figure 10shows the confusion matrix of ADDEFE (PCA+Gridsearch+SVM) classifier using UCI smartphone dataset.

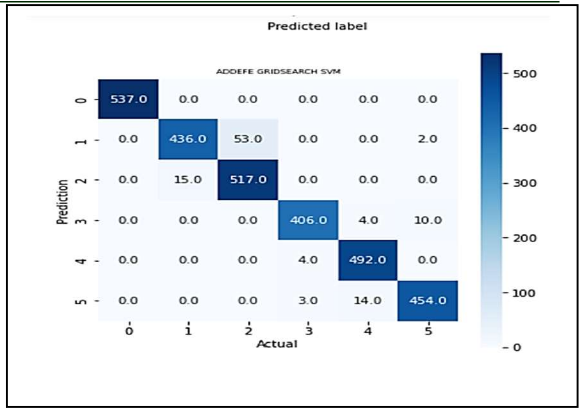

Figure 10. Confusion matrix PCA+Grid search+SVM

Table.VI lists comparative results of the proposed ADDEFE-SVM classifiers along with other techniques on the UCI smartphone dataset.

Table VI- Performance Comparision of proposed ADDEFE method with Conventional classifiers using UCI smartphone dataset

Method	Training Accuracy	Test Accuracy
SVM	0.9317	0.9504
FS+PCA+SVM	0.9392	0.9562
ADDEFE(FS+PCA+ GRSVM)	0.9408	0.9643

It can be seen from the above table the the prposed scheme scores higher values in terms of classification accuracies.

Table VII lists Comparative results of the proposed ADDEFE-SVM classifiers along with other techniques on the WISDN dataset.

Table VII - Performance Comparision of proposed ADDEFE method with Conventional classifiers usingWISDN dataset

Classifiers	Precision	Recall	F-Measure
K-NN	0.832	0.853	0.842
SVM	0.875	0.870	0.872
Decision Tree	0.893	0.913	0.902
Random Forest	0.917	0.916	0.916
ANN	0.925	0.924	0.924
Naive Bayes	0.839	0.914	0.874
ADDEFE(Grid+SVM)	0.942	0.928	0.934

It is evident from Table VI and VII that the proposed ADDEFE+SVM classifies instances of HAR datasets with greater accuracies.

The proposed ADDEFE based methods using Random forest and SVM classifiers gives best performance accuracy upto 96.4%. When compared

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

with conventional methods proposed method gives best performance upto increase of 83.2 (conventional) to 96.4% (proposed) classification accuracy. There is a increase of upto 13.2% classification accuracy obtained using ADDEFE method.

5. CONCLUSIONS AND FUTURE WORK

This paper has highlighted data-driven methods in HAR categorization which rely on high quality data for their training. Highly qualitative and large datasets with proper annotations for HAR datasets are lesser. This proposed work contributes to HAR research in two ways: Improving the quality of HAR dataset and proposing an ensemble feature selection technique for improved classifications of HAR subjects. The proposed technique ADDEFE, created for selecting features required to classify network parameters for implementing HDAs has been proposed in this work. Identifying features that are relevant, minimal and apt have to be chosen either manually or automatically. Moreover, recurrences or duplication of fields while analyzing them increase the dataset size and consume costly processing time in computers. Hence dimensionality reduction techniques are applied for better results. The use of PCA seeks to minimize features reducing dimensionality. Thus, this work's proposed technique is a novel feature selection which executes data preparation, dimensionality reduction and feature selections as a part of its feature selection for selecting the optimal set of parameters for classifications. The proposed work has suggested, implemented and demonstrated with results its proposed ADDEFE technique which can be implemented on HAR smart phone dataset.This work presented an efficient method for predicting human activites. Is there any changce for automatic feature selection and prediction using machine learning is to be addressed. Future work would be using an deep learning model which automates the entire set of processes prior to its execution.

ACKNOWLEDGEMENT

This research work was carried out with thefinancially support of RUSA – Phase 2.0 grantsanctioned vide Letter No. F 24-51 / 2014-U, Policy (TNMulti-Gen) Dept. of Edn. Govt. of India,Dt.09.10.2018 at Alagappa University, Karaikudi, Tamilnadu. India.

REFERENCES

[1] JK Aggarwal, L Xia, OC Ann and LB Theng. Human activity recognition: A review. In Proceedings of the IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2014),2014, 389–393.
[2] WZhao,LYan and Y Zhang. Geometric-constrained multi-view image matching method based on semi-global optimization. Geo-spatial information science, 2018; 21(2), 115-126.
[3] MM Awad. Forest mapping: a comparison between hyperspectral and multispectral images and technologies. Journal of Forestry Research, 2018; 29(5), 1395-1405.
[4] MS Bakli,MA Sakr and THASoliman. A spatiotemporal algebra in Hadoop for moving objects. Geo-spatial information science, 2018; 21(2), 102-114.
[5] A Jalal,MAK Quaid and ASHasan. Wearable sensor-based human behavior understanding and recognition in daily life for smart environments. In Proceedings of the International Conference on Frontiers of Information Technology (FIT), 2018, 105–110.
[6] S Majumder,E Aghayi,M Noferesti,H Memarzadeh-Tehran,T Mondal,Z Pang and MJ Deen. Smart homes for elderly healthcare—Recent advances and research challenges. Sensors, 2017; 17(11), 2496.
[7] DCook, NCKrish and KDFeuz. Transfer Learning for Activity Recognition: A Survey.Knowledge and Information Systems, 2013;36(3), 537-556.
[8] L Chen,J Hoey,CDNugent, DJCook and ZYu. Sensor-based activity recognition. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2012; 42(6), 790-808.
[9] J Wang,Y Chen,S Hao,X Peng and LHu. Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters, 2019; 119, 3-11.
[10] L Liu,Y Peng,M Liu and ZHuang. Sensor-based human activity recognition system with a multilayered model using time series shapelets. Knowledge-Based Systems, 2015; 90, 138-152.
[11] YLiu, LNie, LLiuand DSRosenblum. From action to activity: sensor-based activity recognition. Neurocomputing, 2016; 181, 108–115.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[12] MM Hassan,MZUddin, A Mohamed and AAlmogren. A robust human activity recognition system using smartphone sensors and deep learning. Future Generation Computer Systems, 2018; 81, 307-313.
[13] HHe,Y Tan and WZhang. A wavelet tensor fuzzy clustering scheme for multi-sensor human activity recognition. Engineering Applications of Artificial Intelligence, 2018; 70, 109-122.
[14] GValentini and TGDietterich. Bias-variance analysis of support vector machines for the development of SVM-based ensemble methods. Journal of Machine Learning Research, 2004; 5, 725–775.
[15] R Das and ASengur. Evaluation of ensemble methods for diagnosing of valvular heart disease. Expert Systems with Applications, 2010; 37(7), 5110-5115.
[16] N Hegde, M Bries, TSwibas, EMelanson and ESazonov. Automatic recognition of activities of daily living utilizing insole-based and wrist-worn wearable sensors. IEEE journal of biomedical and health informatics, 2017; 22(4), 979-988.
[17] N Roy,A Misra and DCook. Ambient and smartphone sensor assisted ADL recognition in multi-inhabitant smart environments. Journal of ambient intelligence and humanized computing, 2016; 7(1), 1-19.
[18] K Moriya,E Nakagawa,M Fujimoto,H Suwa,Y Arakawa,A Kimura and K.Yasumoto. Daily living activity recognition with echonet lite appliances and motion sensors. In IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 2017, 437-442.
[19] M Gochoo,THTan, SCHuang, SHLiu and FSAlnajjar. DCNN-based elderly activity recognition using binary sensors. In International Conference on Electrical and Computing Technologies and Applications (ICECTA), 2017, 1-5.
[20] D Singh,E Merdivan,S Hanke, J Kropf, M Geist and AHolzinger. Convolutional and recurrent neural networks for activity recognition in smart environment. In Towards integrative machine learning and knowledge extraction, 2017, 194-205.
[21] DJ Cook and NCKrishnan. Activity Learning: Discovering, Recognizing and Predicting Human Behavior from Sensor Data, 1st ed.Wiley: Hoboken, NJ, USA, 2015.
[22] A Mannini, M Rosenberger, WLHaskell, AMSabatini and SSIntille. Activity recognition in youth using single accelerometer placed at wrist or ankle. Medicine and science in sports and exercise, 2017; 49(4),801-812.
[23] ODLara and MALabrador. A survey on human activity recognition using wearable sensors. IEEE communications surveys & tutorials, 2012; 15(3), 1192-1209.
[24] ZMHira and DFGillies. A review of feature selection and feature extraction methods applied on microarray data. Advances in bioinformatics, 2015.
[25] PDomingos. A few useful things to know about machine learning. Communications of the ACM, 2012; 55(10), 78-87.
[26] I Guyon and AElisseeff. An introduction to variable and feature selection. Journal of machine learning research, 2003; 3, 1157-1182.
[27] Prabhu P., Anbazhagan N. (2013) FI-FCM Algorithm for Business Intelligence. In: Prasath R., Kathirvalavakumar T. (eds) Mining Intelligence and Knowledge Exploration. Lecture Notes in Computer Science, vol 8284. Springer, Cham, pp 518-528.
[28] Paulraj P., Neelamegam A. (2014) Improving Business Intelligence Based on Frequent Itemsets Using k-Means Clustering Algorithm. In: Meghanathan N., Nagamalai D., Rajasekaran S. (eds) Networks and Communications (NetCom2013). Lecture Notes in Electrical Engineering, vol 284. Springer, Cham, DOI : 10.1007/978-3-319-03692-2_19, pp 243-254.
[29] M. Sivakami, P. Prabhu, Classification ofAlgorithms Supported Factual KnowledgeRecovery from Cardiac Data Set, InternationalJournal of Current Research and Review, vol.13.Issue 6. March 2021 pp161-166.
[30] https://archive.ics.uci.edu/ml/index.php.