Performance Comparison of Activity Recognition Classifiers using Big Dataset

Hameeza Ahmed*, Muhammad Ali Ismail†
Department of Computer & Information Systems Engineering,
NED University of Engineering & Technology University Road, Karachi-75270, Pakistan*†

Abstract

Human activity recognition is a promising concept of pervasive computing. Multiple number of on body sensors is employed to achieve this task. Activity Recognition Chain (ARC) makes the process of activity recognition possible. ARC includes various stages namely, data acquisition, preprocessing, segmentation, feature extraction, classification, and decision fusion. Amongst these, classification is the most critical stage. The paper deals with classifying human activities on a big dataset. The classifiers include Naive Bayes, HMM, DA, and k-NN. The paper shows which classifier is best suited in big data environment for classifying the activities.

Keywords—Activity Recognition Chain; Big Data; Classifier; Dataset; Pervasive Computing.

I. INTRODUCTION

Big data refers to datasets whose size is beyond the ability of traditional data processing applications to manage, capture, analyze, and store. Advanced technologies and applications have been introduced to handle increasing Volumes of data. With the data being accumulated from a large number of sources, the techniques also required to be advanced. The main objective of this research is human activity recognition on big data. In order to accomplish this task, comprehensive big data is required [1], [2]. Pervasive computing means "existing everywhere". It refers to the graceful integration of technology including mobile devices, wireless sensors, wearable computing devices, etc in such a manner, that no user is aware of the embedded environment. Activity-aware systems have new applications in military missions, smart environments, emergency response, and surveillance [3], [4]. Activity recognition is intended to recognize human activities in real life scenarios. The activity recognition system keeps tracking the user behavior. The users can get proactive assistance while carrying out several tasks. Activity Recognition Chain (ARC) comprises of several stages like data acquisition, preprocessing, segmentation, feature extraction, classification and decision fusion.

Classification is the most critical phase of ARC. It is possibly the most popular predictive data mining technique and a discrete supervised machine learning method. The classification algorithm is composed of 2 phases namely training and testing. In the training phase the system is trained with a huge set of sample inputs. The testing phase predicts the values of new test data based on the training [5]. A number of classifier algorithms exist in machine learning namely k-Nearest Neighbors (k-NN), Naive Bayes, Support Vector Machines (SVM), Joint Boosting, Hidden Markov Models (HMM), Discriminative Analysis (DA), Decision Trees, Neural Networks, Logistic Regression and many more.

This paper deals with studying the impacts caused by various classifiers in the process of recognizing an activity. The classifiers include Naive Bayes, k-Nearest Neighbors (k-NN), Hidden Markov Models (HMM) and Discriminative Analysis (DA). The experiment is conducted on Educational Activity Recognition Framework which is freely available. The comparison is made by appropriate performance metrics. A huge data set is required to accomplish the specified task. There are various datasets freely available for human activity recognition. But, many of them cannot be categorized as big data. After extensive research a huge dataset known as REALDISP is selected to perform Activity Recognition [6], [2].

Rest of the paper is as follows: Section 2 discusses related work. Activity Recognition Chain is discussed in section 3. A brief introduction of classifiers is provided in section 4. Section 5 gives description about benchmark dataset and performance metrics. Performance results are presented in section 6, followed by section 7 highlighting conclusion.

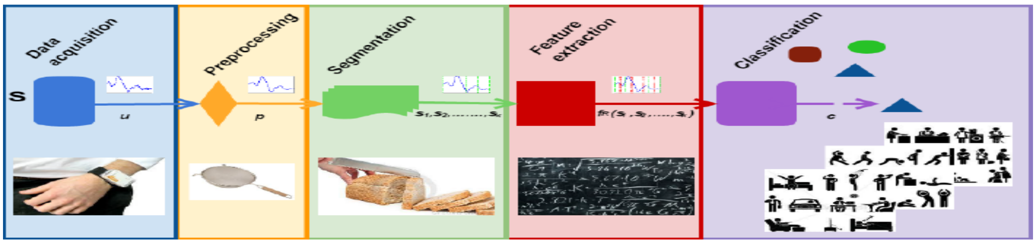
II. RELATED WORK

A huge amount of research is being done in the field of Activity Recognition, which is the latest development these days. In [7] each stage of ARC is discussed in detail. The experiment is conducted on sample dataset. The paper reports the results achieved using multiple classifiers. According to the results, SVM is superior to other classifiers with highest precision and recall values. Naive Bayes and k-NN have lowest recall. An Online Human Activity Recognition on Smart Phones is presented in [8]. It shows the classifier clustered k-NN performs much better than Naive Bayes classifier with respect to accuracy. Another important study is shown in [9]. This paper compares the performance of base level and meta level classifiers for the process of activity recognition. It shows how plural voting is more consistent than other techniques.

III. ACTIVITY RECOGNITION CHAIN

The process of recognizing an activity is termed as activity recognition chain (ARC). It is a series of signal processing, pattern recognition and machine learning techniques. An ARC involves several steps such as data acquisition, preprocessing, segmentation, feature extraction, classification, and decision fusion. The activity recognition chain is shown in Fig. 1.

The streams of sensor data obtained through multiple on Activity Recognition Chainbody sensors are passed as an input to ARC. The acquired data is then preprocessed to filter out artifacts. This entire process of data acquisition and filtering is known as sensor data acquisition and preprocessing. The second stage is data segmentation, which divides the data into multiple sections containing a gesture or activity.



Various methods are available to perform segmentation such as sliding window, energy based segmentation, and rest position segmentation. The segmented data is then passed through the process of feature extraction and selection which extracts the features containing the activity characteristics from the signals within each segment. There are a wide range of choices available in this process such as signal based features, event based features, body model features, and multilevel features. In the considered experiment, signal based features are used. These include popular statistical features mean and variance. These features are simple showing high performance results in various activity recognition tasks. The fourth phase is the process of training and classification. In the training part, a classifier model is trained by the extracted features and class labels. In the classification part, a score for each activity class is calculated by the features and the trained model. A number of classification techniques exist such as k-Nearest Neighbors, Support Vector Machines, Hidden Markov Models, Naive Bayes, Discriminant Analysis and many more. The final stage of ARC is decision fusion which takes place to fuse output of several classifiers into a single decision. Decision fusion can either fuse features called early fusion or it can fuse classifiers known as late fusion [7].

IV. CLASSIFIERS

Classification process is used to categorize an unknown observation into a set of categories, by getting trained from a set of training data comprising of ample observations with known category membership. It is supervised machine learning technique. A large number of classifier techniques exist; some of them are discussed below.

A. k-NN

k-Nearest Neighbors (k-NN) is the most commonly used machine learning technique for activity recognition. The k-Nearest Neighbors comprises of training and classification phases. In the first phase the feature vectors and class labels of the training samples are stored. In the second phase, the test samples are classified by assigning the class labels having the highest similarity to training set by comparing shortest distance metric. Euclidean distance, Manhattan distance, and Minkowski distance are commonly used metrics [10], [11].

B. Naive Bayes

A Naive Bayes classifier is a fast, simple, and easy to implement probabilistic classifier, which is based on Bayes' theorem. Naive Bayes classifiers can be efficiently trained, in a supervised learning scenario. They perform quite well in many complex real world situations, in spite of their simplified design. They require small amount of training data for parameters estimation [12], [13], [14].

C. HMM

The Hidden Markov Model (HMM) is a powerful statistical learning technique for generative sequences modeling. Basically, they are a form of stochastic finite state machine well suited to activity recognition. It can be completely defined by the static state transition probability distribution, number of hidden states, the initial state distribution, and the observation symbol probability distribution. It is assumed that the first order Markov property is followed [15], [16], [17].

D. Discriminative Analysis

It is a multivariate statistical technique which is used to perform the classification of each observation into multiple groups. DA constructs a descriptive group discrimination model which is based on predictor variables. DA predicts group membership based on a linear grouping of the interval variables. The process starts with a set of observations where both the values of the interval variables and group membership are identified. A model is achieved at the end of the process. It allows the prediction of group membership when the interval variables are identified [18], [19].

V. Benchmark Dataset and Performance Metrics

There exists a large number of datasets freely available for the process of activity recognition. But, very few of them could be considered as big data. For this work, REALDISP Dataset has been selected. It fulfills all the major requirements of the considered experiment.

A. *The REALDISP Dataset*

The REALDISP (REAListic sensor DISPlacement) benchmark dataset lend itself for benchmarking activity recognition techniques. The dataset includes a wide range of physical activities (warm up, cool down and fitness exercises), sensor modalities (acceleration, rate of turn, magnetic field and quaternions) and participants (17 subjects). There are 3 sensor displacement scenarios in the considered dataset namely ideal, self, and mutual. At, specific timestamp readings from 9 sensors is obtained. The sensors are IMU's (Inertial Measurement Unit). Each sensor provides 3D acceleration (accX,accY,accZ), 3D gyro (gyrX,gyrY,gyrZ), 3D magnetic field orientation (magX,magY,magZ) and 4D quaternions (Q1,Q2,Q3,Q4). The activity set consists of 33 activities [20].

B. *Performance Metrics*

A number of standard performance metrics are used to better evaluate the performance of ARC.

1) Confusion Matrix: *It is used to evaluate the performance of a multi way classifier. In this, the actual class of instances is plotted against the predicted class. A 2- way classifier, confusion matrix is represented in Table 1.*

2) Recall: *The proportion of positive labelled instances that are correctly identified as positive. It is also known as True Positive Rate (TPR) or hit rate.*

$$TPR = \frac{TP}{TP + FN} \qquad (1)$$

3) Precision: *The proportion of predicted positive instances that are correct. It is also known as Positive Predicted Value (PPV).*

$$PPV = \frac{TP}{TP + FP} \qquad (2)$$

4) False Positive Rate: *The proportion of negative instances that are incorrectly recognized as positive.*

$$FPR = \frac{FP}{FP + TN} \qquad (3)$$

5) Average Training Time: *The time which is required to train the labels for each training set. It is expressed in seconds. The average is performed on the time of each training set.*

6) Average Testing Time: *The time which is required to test the labels for each testing set. It is expressed in seconds. The average is performed on the time of each testing set.*

VI. Experimental Results

A. *Experimental Setup*

The REALDISP dataset has 17 subjects which are further divided into an ideal and self. We have limited the input samples by selecting only first 9 class labels of each subject's ideal files. We further divided the activities into two sets. Set 1 contains basic foot activities and set 2 contains jump activities. Null activity is included in both the sets. The data is then sent as an input to Activity Recognition Framework [21]. The framework is responsible to perform all the steps of an ARC. The run time parameters are shown in Table 2.

B. *Performance Results*

The experiment of recognizing an activity in the big dataset has been conducted by MATLAB activity recognition framework using the parameters mentioned in Table 2. By varying the classifier type, performance significantly varies. The performance comparison can be made using the above mentioned metrics. The confusion matrices of each 5 classifiers for both the activity sets are shown in Figures 2-9.

In the Activity Recognition process, both the precision and recall metrics are of significant importance. When an algorithm returns considerably more appropriate results, it is termed as having high precision. Whereas when an algorithm produced results are mostly appropriate, it is termed as recall. Precision refers to quality or exactness and recall refers to quantity or completeness. Fig. 10 and 11 depicts the recall and precision results of each 4 classifiers for basic foot activities set. The recall and precision outcome of each 4 classifiers for jump activities are shown in Fig. 12 and 13. Table 3 depicts the performance results for basic foot and jump activities set in tabular form.

THE CONFUSION MATRIX FOR 2 WAY CLASSIFIER

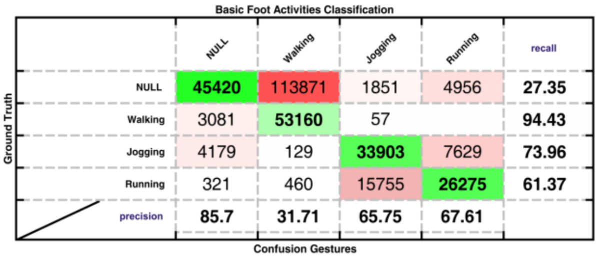
	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

RUN TIME PARAMETERS

Properties	Values
Toolbox	Educational Activity Recognition Framework
Sensors Used	Accelerometer_1
Activity Set	Foot Activity: (Null, Basic Foot Activities Walking, Jogging, Running) Jump Activity: (Null, Jump up, Jump front & back, Jump sideways, Jump leg/arms open/closed, Jump rope)
Segmentation Technique	Sliding Window
Features Selected	Mean, Variance
Classifier Techniques	HMM, k-NN, Naive Bayes, DA
Fusion Type	Early

PERFORMANCE RESULTS OF BASIC FOOT ACTIVITIES AND JUMP ACTIVITIES

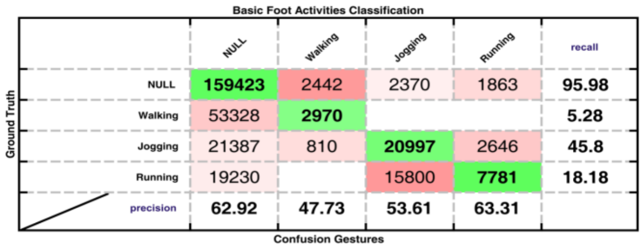
	k-NN	Naive Bayes	HMM	DA
Basic Foot Activities				
Average Training Time in Seconds	0.0019	0.0615	97.9765	0.0045
Average Testing Time in Seconds	0.5868	0.0084	5.5820	0.0998
Overall Precision in Percentage	55.04	43.37	43.92	42.27
Overall Recall in Percentage	21.90	73.54	78.19	73.03
FPR in Percentage	15.61	83.80	87.12	87.05
Jump Activities				
Average Training Time in Seconds	0.0012	0.0381	67.2697	0.0013
Average Testing Time in Seconds	0.3708	0.0069	5.0490	0.0778
Overall Precision in Percentage	46.57	50.18	55.29	52.55
Overall Recall in Percentage	33.34	48.36	50.77	48.45
FPR in Percentage	22.69	28.47	24.35	25.94


Basic Foot Activities Classification

Ground Truth	NULL	Walking	Jogging	Running	recall
NULL	45420	113871	1851	4956	27.35
Walking	3081	53160	57		94.43
Jogging	4179	129	33903	7629	73.96
Running	321	460	15755	26275	61.37
precision	85.7	31.71	65.75	67.61	

Confusion Gestures

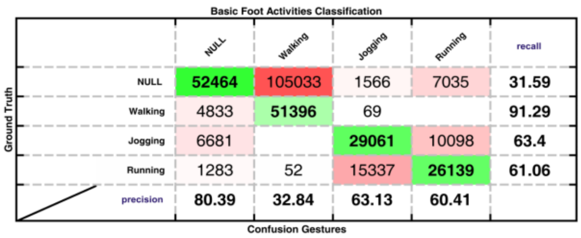
Fig. 2 Classifier HMM Confusion Matrix of Basic Foot Activities


Basic Foot Activities Classification

Ground Truth	NULL	Walking	Jogging	Running	recall
NULL	159423	2442	2370	1863	95.98
Walking	53328	2970			5.28
Jogging	21387	810	20997	2646	45.8
Running	19230		15800	7781	18.18
precision	62.92	47.73	53.61	63.31	

Confusion Gestures

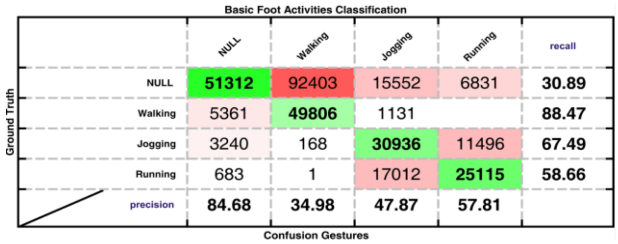
Fig. 3 Classifier k-NN Confusion Matrix of Basic Foot Activities


Basic Foot Activities Classification

Ground Truth	NULL	Walking	Jogging	Running	recall
NULL	52464	105033	1566	7035	31.59
Walking	4833	51396	69		91.29
Jogging	6681		29061	10098	63.4
Running	1283	52	15337	26139	61.06
precision	80.39	32.84	63.13	60.41	

Confusion Gestures

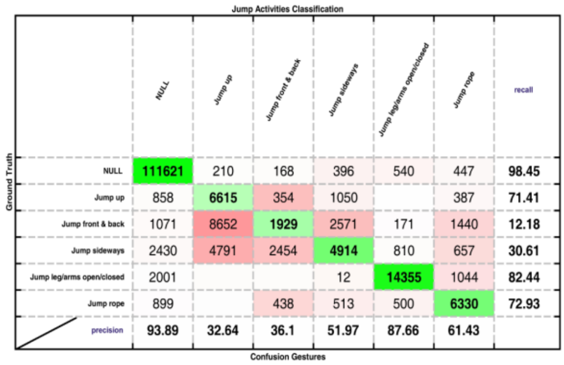
Fig. 4 Classifier Naive Bayes Confusion Matrix of Basic Foot Activities


Basic Foot Activities Classification

Ground Truth	NULL	Walking	Jogging	Running	recall
NULL	51312	92403	15552	6831	30.89
Walking	5361	49806	1131		88.47
Jogging	3240	168	30936	11496	67.49
Running	683	1	17012	25115	58.66
precision	84.68	34.98	47.87	57.81	

Confusion Gestures

Fig. 5 Classifier DA Confusion Matrix of Basic Foot Activities


Jump Activities Classification

Ground Truth	NULL	Jump up	Jump front & back	Jump sideways	Jump leg/arms open/closed	Jump rope	recall
NULL	111621	210	168	396	540	447	98.45
Jump up	858	6615	354	1050		387	71.41
Jump front & back	1071	8652	1929	2571	171	1440	12.18
Jump sideways	2430	4791	2454	4914	810	657	30.61
Jump leg/arms open/closed	2001			12	14355	1044	82.44
Jump rope	899		438	513	500	6330	72.93
precision	93.89	32.64	36.1	51.97	87.66	61.43	

Confusion Gestures

Fig. 6 Classifier HMM Confusion Matrix of Jump Activities


Jump Activities Classification

Ground Truth	NULL	Jump up	Jump front & back	Jump sideways	Jump leg/arms open/closed	Jump rope	recall
NULL	112587	15	78	93	528	81	99.3
Jump up	2001	5916	897	450			63.86
Jump front & back	3333	7950	3642	792	21	96	23
Jump sideways	5796	6567	2568	909	213	3	5.66
Jump leg/arms open/closed	6960		198	327	9309	618	53.46
Jump rope	1802	99	1295	513	2325	2646	30.48
precision	84.98	28.79	41.97	29.47	75.1	76.83	

Confusion Gestures

Fig. 7 Classifier k-NN Confusion Matrix of Jump Activities

Jump Activities Classification

Ground Truth \ Confusion Gestures	NULL	Jump up	Jump front & back	Jump sideways	Jump leg/arms open/closed	Jump rope	recall
NULL	**111837**	84	9	201	756	495	**98.64**
Jump up	705	**7827**	102	234		396	**84.49**
Jump front & back	732	11379	**222**	1458	96	1947	**1.4**
Jump sideways	1272	8616	633	**3390**	594	1551	**21.11**
Jump leg/arms open/closed	708		15	117	**15558**	1014	**89.35**
Jump rope	570	29	583	177	1800	**5521**	**63.61**
precision	**96.56**	**28.02**	**14.19**	**60.79**	**82.74**	**50.54**	

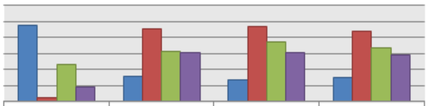
Fig.8 Classifier Naive Bayes Confusion Matrix of Jump Activities

Jump Activities Classification

Ground Truth \ Confusion Gestures	NULL	Jump up	Jump front & back	Jump sideways	Jump leg/arms open/closed	Jump rope	recall
NULL	**112020**	24		324	714	300	**98.8**
Jump up	1296	**6525**	840	321		282	**70.43**
Jump front & back	1701	9336	**1155**	1701	81	1860	**7.29**
Jump sideways	2385	7404	819	**3690**	153	1605	**22.98**
Jump leg/arms open/closed	684	195	18	219	**14898**	1398	**85.56**
Jump rope	549	50	529	423	819	**6310**	**72.7**
precision	**94.42**	**27.73**	**34.36**	**55.26**	**89.4**	**53.68**	

Fig. 9. Classifier DA Confusion Matrix of Jump Activities

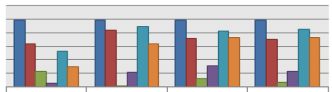
Recall Value of Basic Foot Activity



	k-NN	Naive Bayes	HMM	DA
Null	95.98	31.59	27.35	30.89
Walking	5.28	91.29	94.43	88.47
Jogging	45.8	63.4	73.96	67.49
Running	18.18	61.06	61.37	58.66

Fig.10 Recall Value of Basic Foot Activity for Multiple Classifiers

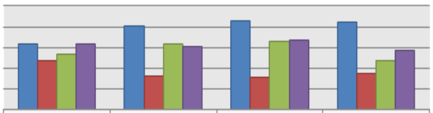
Recall Value of Jump Activity



	k-NN	Naive Bayes	HMM	DA
Null	99.3	98.64	98.45	98.8
Jump Up	63.86	84.49	71.41	70.43
Jump Front and Back	23	1.4	12.18	7.29
Jump Sideways	5.66	21.11	30.61	22.98
Jump leg/arms open closed	53.46	89.35	82.44	85.56
Jump rope	30.48	63.61	72.93	72.7

Fig. 11 Recall Value of Jump Activity for Multiple Classifiers

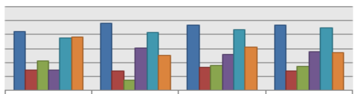
Precision Value of Basic Foot Activity



	k-NN	Naive Bayes	HMM	DA
Null	62.92	80.39	85.7	84.68
Walking	47.73	32.84	31.71	34.98
Jogging	53.61	63.13	65.75	47.87
Running	63.31	60.41	67.61	57.81

Fig. 12 Precision Value of Basic Foot Activity for Multiple Classifiers

Precision Value of Jump Activity



	k-NN	Naive Bayes	HMM	DA
Null	84.98	96.56	93.89	94.42
Jump Up	28.79	28.02	32.64	27.73
Jump Front and Back	41.97	14.19	36.1	34.36
Jump Sideways	29.47	60.79	51.97	55.26
Jump leg/arms open closed	75.1	82.74	87.66	89.4
Jump rope	76.83	50.54	61.43	53.68

Fig. 13 Precision Value of Jump Activity for Multiple Classifiers

By observing the charts in Fig 10-13, we can see how each classifier performs differently with different activities. For Null activity in case of activity set 1, k-NN performs best in terms of recall value. Naive Bayes, DA, HMM have low recall value. In terms of precision of Null activity, HMM leads all other classifiers. DA, Naive Bayes and k-NN are ranked in ascending order of precision. Although, k-NN has higher recall value but it is less precise for Null activity of activity set 1. HMM is the most successful classifier having highest recall value for walking activity, then ranked are Naive Bayes, DA and k-NN respectively. In terms of precision, k-NN is most precise in walking activity, then ranked DA, Naive Bayes and HMM, each having nearly the same precision. Activity Jogging is most accurately recognized by HMM classifier which has highest recall value. DA comes after HMM, it also shows significant outcome. After this, Naive Bayes is ranked. k-NN performs least amongst all. Classifier HMM is most precise for this activity. Naive Bayes comes after this, showing significant outcome. After this, k-NN is ranked. DA is least precise amongst all. For the running activity, classifier HMM shows the highest recall value. Then, ranked are Naive Bayes and DA respectively. k-NN performs least amongst all for the recall value. In terms of precision, HMM is superior to others. k-NN, Naive Bayes, and DA are ranked after HMM for precision value.

For activity set of jump activities, the classifier k-NN gives highest recall value for the null activity. DA, Naive Bayes, and HMM also shows significant recall values. In terms of precision Naive Bayes leads other classifiers then ranked DA, HMM and k-NN respectively. Classifier Naive Bayes gives best recall result for Jump up activity. Then ranked are HMM, DA, and k-NN respectively. HMM shows highest precision here. k-NN, Naive Bayes, and DA are ranked after HMM respectively. In case of activity Jump front and back, classifier k-NN gives the best performance with highest recall and precision values.

HMM, DA and Naive Bayes are ranked after k-NN respectively in both the recall and precision values. Naive Bayes performs worst here. Classifier HMM gives the best recall results for the activity Jump sideways. DA, Naive Bayes, and k-NN classifiers are ranked after HMM respectively. With respect to precision, Naive Bayes is most precise here. DA, HMM, and k-NN come after Naive Bayes respectively. For the activity Jump legs/arms open closed, classifier Naive Bayes gives the best recall results. DA is ranked second here. Then come HMM and k-NN respectively. k-NN performs least amongst all. In terms of precision, classifier DA shows the best precision results. HMM is ranked second, and then come Naive Bayes, and k-NN respectively. For the activity Jump rope, HMM gives the best recall result. Then ranked are DA, Naive Bayes and k-NN respectively. k-NN shows least performance. In terms of precision, k-NN gives the best results. Then, ranked are HMM, DA and Naive Bayes respectively.

By performing the above comparison, we can observe that for some activities the recall value is not so higher. But still one classifier performs better than the other though it might have less overall recall value. By observing the above charts, it is clearly seen how HMM dominates other classifiers in basic foot activity set both in terms of recall and precision. Similarly, for the jump activity set each HMM, k-NN and Naive Bayes show significant performance but DA is weak and there is no single ruler here. In terms of precision as well, Naive Bayes and k-NN shows higher precision for 2 activities whereas HMM and DA shows higher precision for 1 activity only. In the overall it is observed that HMM dominates other classifiers. But, the training and testing time of HMM is higher than other techniques. k-NN is ranked after HMM as it shows higher recall for 1 basic foot activity and 2 jump activities. Plus, it has good precision and low training and testing time. k-NN is less complex as compared to other techniques. Naive Bayes is ranked third as it shows higher recall and precision for 2 activities only. The training and testing time of Naive Bayes is also lower as compared to other techniques. Although DA, is unable to give highest recall value for any particular activity but it has good overall recall. DA gives high precision for 1 activity only, but the training and testing time of DA is less.

VII. Conclusion

The paper presents an approach of identifying activities of a huge dataset known as REALDISP using MATLAB Activity Recognition Framework. Classification is much important as compared to other stages of an ARC as it is the intermediate stage between preprocessing and final fusion. A good classifier can certainly recognize the activities more accurately. The classification results can be better observed by the confusion matrix. A tradeoff is seen in the results which are presented in this paper, HMM which identifies greater number of activities with higher recall and precision requires more time for both the training and testing. k-NN, Naive Bayes, and DA are trained in less time but there recognized activities are less as compared to HMM. k-NN is a reasonable classifier as it recognizes 3 activities superiorly. Few classifiers give good recall and precision results for some activities while other show good performance for different activities. Thus, the results of multiple classification techniques can be fused together to better perform the process of activity recognition.

References

[1] H. Ahmed, M. A. Ismail, M. F. Hyder, "Performance Optimization of Hadoop Cluster Using Linux Services," *International Multi Topic Conference (INMIC), 2014 IEEE 17th International Conference on*, December 8-10, 2014, Karachi, Pakistan.

[2] M. A. Ismail, H. Ahmed, "Performance Comparison of Classifiers for Activity Recognition using Big Dataset," Accepted for *International Multi Topic Conference (IMTIC 2015).*

[3] Satyanarayanan, M., "Pervasive computing: vision and challenges," *Personal Communications, IEEE* , vol.8, no.4, pp.10,17, Aug 2001 doi: 10.1109/98.943998.

[4] Choudhury, T., et al., "The Mobile Sensing Platform: An Embedded Activity Recognition System," *Pervasive Computing, IEEE* , vol.7, no.2, pp.32,41, April-June 2008 doi: 10.1109/MPRV.2008.39.

[5] Pandey, Umesh Kumar, and Saurabh Pal. "Data Mining: A prediction of performer or underperformer using classification." *arXiv preprint arXiv:1104.4163* (2011).

[6} Activity Recognition Dataset, https:// archive.ics.uci.edu/ml/datasets/REALDISP+Activity+Recognition+Dataset.

[7] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. "A tutorial on human activity recognition using body-worn inertial sensors". *ACM Comput. Surv.* 46, 3, Article 33 (January 2014), 33 pages. DOI=10.1145/2499621 http://doi.acm.org/10.1145/2499621.

[8] Kose, Mustafa, Ozlem Durmaz Incel, and Cem Ersoy. "Online human activity recognition on smart phones." *Workshop on Mobile Sensing: From Smartphones and Wearables to Big Data*. 2012.

[9] Ravi, Nishkam, et al. "Activity recognition from accelerometer data." *AAAI*. Vol. 5. 2005.

[10] M.Sharma, S. K. Sharma, " Generalized K-Nearest Neighbour Algorithm- A Predicting Tool", *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 11, 2013, pp 1-4.

[11] Voulgaris, Zacharias, and George D. Magoulas. "Extensions of the k nearest neighbour methods for classification

problems."*Proc. of 26th IASTED International Conference on Artificial Intelligence and Applications, CD Proceedings ISBN*. 2008.

[12] Naive Bayes, Available [online]: https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/algo_nb.htm#i1005770.

[13] Naive Bayes Classifier, Available [online]: http://www.ic.unicamp.br/~rocha/teaching/2011s2/mc906/aulas/naive-bayes-classifier.pdf.

[14] Samuel Odei Danso, "An Exploration of Classification Prediction Techniques in Data Mining: The insurance domain", Available [online]: http://www.comp.leeds.ac.uk/scsod/MSc%20Dissertation.pdf.

[15] Bicego, Manuele, David MJ Tax, and Robert PW Duin. "Component-based discriminative classification for hidden Markov models." *Pattern Recognition* 42.11 (2009): 2637-2648.

[16] Senior, Andrew. "A hidden Markov model fingerprint classifier."*Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on*. Vol. 1. IEEE, 1997.

[17] Blunsom, Phil. "Hidden markov models. " *Lecture notes, August* 15 (2004): 18-19.

[18] Fernandez, George CJ. "Discriminant Analysis, a powerful classification technique in data mining."*Proceedings of the SAS Users International Conference*. 2002.

[19] Discriminant Function Analysis, Available [online]: http://www.psychstat. missouristate.edu/multibook/mlt03.htm.

[20] O. Banos, M. Damas, H. Pomares, and I. Rojas, “Activity recognition based on a multi-sensor meta-classifier,” in Proceedings of the 12th International Conference on Artificial Neural Networks: Advances in Computational Intelligence, pp. 208–215, Springer, Berlin, Germany, 2013.

[21] Educational Activity Recognition Framework, Available [online]: https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/object-recognition-and-scene-understanding/educational-activity-recognition-framework-and-dataset/

Biographies

Hameeza Ahmed received the B.E degree in Computer and Information Systems from the NED University of Engineering and Technology, Pakistan in 2012. She is currently perusing her master's dissertation from the same university. Her research interests include distributed parallel computing, pervasive computing, and computer architecture.

Muhammad Ali Ismail received his B.E degree in Computer and Information Systems from NED University of Engineering & Technology, Pakistan in 2003. He obtained his PhD in the field of High Performance Computing from NED University in 2011. He did his Post Doctorate in ADSE from ULBS in 2014. Currently, he is serving as associate professor in the Department of Computer and Information Systems Engineering, NED University. Also, he is visiting research fellow at ULBS, Romania and chair of High Performance Computing Centre (HPCC) NED University. His areas of interest include serial, parallel and multi-core processor architecture, cloud/grid/cluster/distributed memory architecture and programming, serial/parallel meta-heuristics, automatic design space exploration and multi-objective optimization.

***