2019 Global Conference for Advancement in Technology (GCAT)
Bangalore, India. Oct 18-20, 2019

Human Activity Recognition with Streaming Smartphone Data

Yanjia Zhang
Department of Mathematics & Statistics
University of South Florida
Florida, USA
yzhang17@mail.usf.edu

Kandethody M. Ramachandran
Department of Mathematics & Statistics
University of South Florida
Florida, USA
ram@mail.usf.edu

***Abstract*—With the widely used smartphones, dynamic data coming from built in sensors, such as human activity data, can be easily obtained. Many applications' developments, such as applications in healthcare, fitness monitoring, and elder monitoring, are based on this kind of dynamic data. Although there are many offline methods that have made a great progress in analyzing these kinds of data, it still has a big challenge to get good results from a streaming data perspective. In this paper, we use an online method called Very Fast Decision Tree (VFDT) to mimic the real scenario. There are two main improvements from the existing models: 1) we train the model online and only use the examples data once for training instead of using them more than once; 2) after building VFDT, the model can be adjusted to identify new activities by adding only small amount of labeled observations. Our experiment on the same existing activities shows that the proposed algorithm achieves an average accuracy of 85.9% for all subjects and single subject accuracy rates are between 60.5% and 99.3%. Moreover, the average accuracy of learning new activity from a different data is 84% and single subject accuracy rate goes to as high as 100%.**

Keywords— Streaming Data, VFDT, Decision Tree, Human Activity Recognition, Smartphone

I. INTRODUCTION

With the development of technology, more and more wearable devices have become available and affordable and the apps with health trackers have become popular. These daily worn devices with applications present a convenient way to record physiological data from users and to provide a basic overview of health status and summary of activities. For example, accelerometer, gyroscope, and magnetometers sensors in the smartphones provide the 3-axis (x, y, z) data, which can be used to track motions, such as walking, standing, and jumping, called Human Activity Recognition (HAR). Because of these advantages, daily activity data is frequently used for health and fitness monitoring or recreational activities. However, most of these devices are not suitable for the medical monitoring of high-risk patients [1]. Meanwhile, there are several challenges and bottlenecks for these data from wearable devices to be more useful and reliable in medical purposes [1]. First, an IoT platform with simple and secure connectivity is required, including data collection, transmission, storage and observation in a medical station. Second, the power needs to be easily managed and monitored long-term without significant power loss. Finally, the data quality should be preserved. From the statistical perspective, these challenges are related to the collection, storage, and compression of the original data, effective ways of selecting data features, and good algorithms using the least information to build the precise models for prediction and classification. Based on these purposes and based on the fact that a truly streaming data is not publicly available, we proposed pseudo streaming methods of identifying human activities of smartphone-based data with high speed classification and efficient data usage. We use the data from the UCI Machine Learning website [2] as the case study. We also use WISDM lab data [3] to explore the adaptive power of this model.

A. Previous Works

The work of human activity recognition based on the sensors can be traced back to 1990s [5]. Sharma et al. [6] applied neural networks (ANN) for a chest worn wireless sensor dataset and achieved 83.95% accuracy. Kwapisz [7] performed the J48 decision tree and multi-layers perceptron's method to the HAR data from a smartphone with only one accelerometer. They point out that these two methods have higher accuracy than other data mining methods. However, both lack the ability to efficiently identify similar activities, for example, walking upstairs vs. downstairs and sitting vs. standing. He and Jin [8] combined Principle Components Analysis (PCA) and Support Vector Machine (SVM) to classify four activities and got 97.5% average accuracy. Sohn and Khan [9] also used PCA but they combined it with Linear Discriminant Analysis (LDA) and Artificial Neural Net (ANN) to detect if activities are abnormal. The highest accuracy rate they got is 78%. Wanmin Wu et al. [10] used K Nearest Neighbors (KNN) as the best classifier with iPod Touch data, but the results show that it fails to effectively classify similar activities as well. Anguita et al. [11] used 561 transformed features to classify six different activities using a one vs. all SVM and obtained as high as 96% accuracy. Fergani [12] used PCA based multi-classfier to get 96.9% average accuracy for daily activities. Zhang, Wu and Luo [13] point out that the combination of the Hidden Markov Model and the Deep Neural Network (HMM-DNN) has a higher accuracy compared with Gaussian mixture method, Random Forest, and their combination with HMM. The accuracy of HMM-DNN is 93.5%. Guo et al. [14] performed a two layer and multi-strategy frame work for sensor smartphone data and the result shows 95.71% average accuracy. Besides, Charissa Ann Ronao and Sung-Bae Cho [15] applied deep learning neural networks (DNN) to both raw sensor data and FFT smartphone data. Their work shows an overall 94.79% accuracy with raw sensor data and 95.75% with additional FFT information. Nakano and Chakraborty [16] point out that the convolutional neural network (CNN) has better performance in identifying dynamic activities than other methods. The average accuracy is 98% with classifying walking, walking upstairs and walking downstairs. Andrey Ignatov [17] used CNN for the

978-1-7281-3694-3/19/$31.00 ©2019 IEEE

accelerometer data from smartphones. They obtain a 97.63% average accuracy with the statistical features. As we can see, that DNN and CNN give higher average accuracy rates comparing to others, but they are conducted off-line. These manners ignore the characteristic of data generation and cannot update with new activities. Another issue for most of the methods is the difficulty in discriminating between similar activities, especially for sitting and standing, walking upstairs and walking downstairs.

Some methods consider sensor-based data as time series data, but they are still unlikely to be updated with the upcoming new data, which implies that they all assume the data is a random sample from a stationary distribution [18]. In reality, we can only use the training dataset for creating the model. This dataset comes from small sample subjects in a lab and stores on the local devices. However, when the application is activated, there is only one single subject; this means the new pattern might not be recognized well. Further, the system itself should have the ability to identify more activities if the user provides new labeled data. In this case, we need a model which can quickly deal with incoming data, can keep the useful information from the previous examples, and can be updated with these new labeled data. Because of these considerations, the most appropriate way to build the HAR system might be online with a streaming data.

There are some studies that are conducted for online data analysis. In 2009, N. Gyorbiro, A. Fabian, and G. Homanyi [19] proposed an on-line HAR mobile system. Wang, Liang, et al. [20] used a real-time hierarchical model for recognizing complex activities with body sensor data and had an average accuracy of 82.87%. Okeyo, George, et al. [21] applied a dynamic segmentation model using varied time windows. This work shows an average accuracy above 83% for recognizing activities. Considering the necessity of the sequential training in the real world for sensor data, Al Jeroudi, Yazan, et al. [22] used a sequential extreme learning machine method (OSELM) and achieved an average accuracy of 82.05%. Shuang Na [23] used the Online Bayesian Kernel Segmentation method for classifying 6 activities. The result shows a 92% average accuracy rate. The details of these four papers are in Table I. The first two papers use video data and the advantage of this kind of data is obvious. With visualization, we might be able to classify more complex activities and scenarios, such as making coffee, washing hands, and so on. But saving and processing these streaming videos requires large memory storage and complex pre-process data steps. So, smartphones with one or two accelerometer sensors is more suitable for recording daily activities. [22] and [23] are two examples of this. They both use the same data from UCI. Unfortunately, [22] needs a large size window segmentation to train the hidden layers. And [23] only uses the last data window to create a new classifier but forgets all the previous information. Both [22] and [23] lack the ability to adapt the incoming labeled data from single users and might violate the stationary assumption at the very beginning.

To address the above challenges and try to improve the existing methods, we propose an online tree based method with preprocessed feature selection. Very Fast Decision Tree (VFDT) is a tree based online classifier, which was first proposed by Pedro Domingos and Geoff Hulten in 2000 [24]. The purpose of this algorithm is to deal with continuous data streams by building decision trees using constant memory and time per example [24]. This method is used in many streaming fields, including fraud detection [25], [26], and sensor networks [27]–[29]. It can also be applied for handling missing values [30] and implementing in distributed environment [31]. These works provide the evidence that VFDT is a most prevalent learner in streaming data classification problems. In our case, the main reasons for selecting VFDT are as follows: 1) it has small memory space requirement, thus making it suitable for smartphones; 2) its use of subsampling to build decision trees helps in detecting activities changing; 3) it adjusts the previous decision tree to the new coming labeled data; 4) it avoids segmentation, which is another big challenge for streaming data analysis. These advantages make VFDT to be a suitable online classifier for human activities system built for smartphones data.

B. *Paper Structure*

In this paper, VFDT is implemented to identify 6 human activities, including walking, waling upstairs, walking downstairs, sitting, standing, and lying down. Our purpose is to build a decision tree-based learner which can update and adjust the previous tree. The contributions of this paper include the following:

1) Selecting features: instead of using principal components analysis, which is used in most of the references above, we use the decision trees to preprocess feature selection from the 561 transformed attributes.
2) Generating streaming data: instead of using all the training data, we use a streaming data generator to release examples at constant times. Thus, we mimic the real data recording process.
3) Updating model: instead of keeping the final model from the lab data, VFDT is capable of implementing new labeled data generated by users. Thus, the model initially built in the system can be considered as the first stage of the training process. During usage, new activities, such as jogging, can be added, then the system can identify the user's personality.

The rest of the paper is organized as follows: Section 2 introduces the data process and structure; Section 3 introduces the proposed method include the feature selection and the VFDT; Section 4 gives the results of the experiment and Section 5 concludes the paper.

II. Data Processing

In this paper, we used the smartphone data from UCI [2] and WISDM Lab [3]. For UCI data, there are 30 volunteers with an age range between 19 to 48 years old. They are randomly divided into training and testing groups, 21 of them are in the training group and the rest 9 are in the testing group. All of them perform 6 activities (walking, walking upstairs, walking downstairs, sitting, standing, and lying down) wearing Samsung Galaxy S II on the waist. The smartphone collected the data in 3-axial linear acceleration and angular velocity. Then the data provider modified the data using a median filter and 3rd order low pass

TABLE I. ONLINE METHODS SUMMARY

Paper	Data Type	Method	Acc. (%)
[20]	Sensory Data	Emerging Pattern Based Algorithm	82.87
[21]	Video Data	Window Approach	83.0
[22]	UCI	Sequential Extreme & One layer network	82.05
[23]	UCI	Online Bayesian Kernel Segmentation	92

Butterworth filter with a corner frequency of 20Hz. Besides, Fast Fourier Transform (FFT) is also applied to the signals. After all of this, we have 561 features from each window of the raw data. In order to mimic the real time online situation, we then leased examples one by one during the training process and discard old observations later to simulate a stream data for which the data points can be used only once, and model is updated gradually. The training data has a total of 7352 examples. The detailed size of each activities in Table II, where W is Walking, WU is Walking Upstairs, WD is Walking Downstairs, ST is Sitting, SD is Standing, and LD is Lying Down. The sizes of each activities are close in number, it is reasonable to consider all the classes as balanced.

We also used WISDM lab data to evaluate capability of our algorithm to recognize new activities without going through extensive training. This data collected from 36 volunteers. They performed 6 activities with an Android-based smartphone in their front pants leg pocket. Every volunteer was asked to walk, walk upstairs, walk downstairs, sit, stand and jog for specific periods of time. Jogging is the new activity. Some of them might not do all the 6 activities. Instead of recording 3 sets of 3-axis data, WISDM data only recorded 2 sets, which means that there are only 6 features in raw data. Besides, WISDM data were transformed in a different way. It calculated some statistics, such as average, standard deviation and difference, instead of using FFT. There are 44 features after transformation, much less than 561 features in UCI data. Since there are several missing values in each feature, we replaced these missing values with 0. To test whether our method can use less examples to identify classes or not, we randomly selected nine volunteers' data as training set. The number of each activity is in Table III. Since there are two volunteers who did not perform Jogging, we ignored these data in out testing. Thus, there are 25 cases.

TABLE II. SIZE OF ACTIVITIES IN UCI TRAINING DATA

Activity	W	WU	WD	ST	SD	LD
Size	1226	1073	986	1286	1374	1407

TABLE III. SIZE OF ACTIVITIES IN WISDM TRAINING DATA

Activity	W	WU	WD	ST	SD	JOG
Size	552	185	153	116	76	425

III. PROPOSED STREAMING METHOD

In this section, we will discuss the proposed method, including the features selection and VFDT algorithm. The big difference here for selecting features from other methods in the literature is using Decision Tree for extracting instead of using Principle Components Analysis (PCA), which is most used in the research, such as [8], [9] and [12].

A. Features Selection

Consider all the 561 features for each observation, there is high dimensional complexity and high correlation between these features. Then, we first selected the most important features. The normal approach is PCA, which sets the eigenvalues of the covariance matrix as the weights for all of features, then uses the linear combinations of these eigenvalues to get the new low dimensional inputs. However, PCA is not a suitable method in online HAR since the activity distribution is changing all the time and hence non-stationary. Lansangan and Barrios said in their paper that PCA of non-stationary time series, the first component will be a linear combination with similar weight for all inputs [32]. Besides, the covariance matrix only based on the training data, it is hard to be updated in a streaming fashion. On the other hand, suppose we ignored the non-stationary aspect and used PCA with 95% of variance explanation in the training and transformed the testing data, result shows that the average accuracy is 76.1% using VFDT, which is lower than proposed feature selection. Also, implementing PCA in algorithm needs more time to compute components than just to use a subset of features. To overcome above mentioned limitations of PCA based methods, we used Decision Tree (DT) to extract important features. When we built a univariate tree, the algorithm only used the necessary variables and selected the most important ones first. This means that the closer to the root, the more important the features are [33]. This method is suitable for non-stationary streaming data, and also from our experiment, this method gives a good preprocess of the data that resulted in 36 features, which in turn results in better classification accuracy. The process is shown in Fig.1.

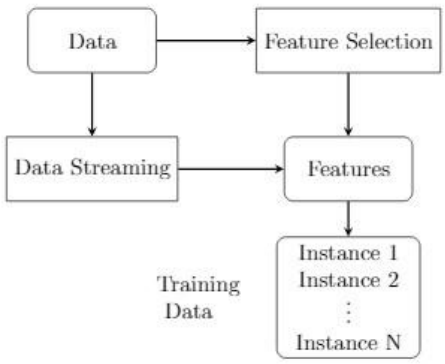

Fig. 1. Feature Selection and Streaming Data Creation.

B. Very Fast Decision Tree (VFDT)

Geoff,etc. [34] introduced a streaming classification method in 2001, namely Very Fast Decision Tree (VFDT). They used the Hoeffding bound to decide the minimum observations needed for each new split and grouped the tree based on the new branch. In other words, VFDT waits for new examples to arrive instead of recruiting previous ones to split the internal nodes. The two main crucial aspects needed to build this tree are deciding when to split a node and which feature is used to split. For the former one, it involves the Hoeffding bounds, which states that with probability (1-δ), the difference between the true mean of a real-valued random variable in range R and the estimated mean will be less than ε after n independent examples, where:

$$\epsilon = \sqrt{\frac{R^2 ln(1/\delta)}{2n}} \tag{1}$$

(1) states that a small part of the sample will be enough to choose an optimal feature for splitting. For the latter one, it needs a heuristic measure. The most popular measures are information gain (IG) which measures the 'purity' of each subset of a split [35], and Gini Index (GI) which estimates the probability of misclassification under the split [36]. For any given potential split, VFDT checks if the difference of heuristic measure of the top two attributes is greater than ε^2 under a given δ, if so, the winning attribute will be picked and tested. Thus, this algorithm can determine the smallest number n of examples needed with a high probability. Moreover, it is easy to estimate learning time since it uses constant time per example. The pseudo-code for VFDT after our tree-based feature selection is shown below. The novelty of the VFDT used in this work lies in using the pre-training examples to build a DT first instead of building the Hoeffding Tree from root. The whole process including feature selection is given in Fig. 2.

Algorithm 1 The VFDT Algorithm

Input: *S : a streaming of example*
X: a set of selected features
IG: Information Gain
δ: probibility of misclassification
τ:a tie threshold
n_{pre}*: # of examples used in pre-training*
n_{min}*: # of examples for checking new split*
Output: VFDT

Procedure: *VFDT* $\{S, X, IG, \delta, \tau, n_{pre}, n_{min}\}$
Let DT be a tree from the n_{pre} *examples using 36 features*
Let $n_{ijk}(l)$ *be # of examples in leaf l for* i^{th} *feature* j^{th} *value in class k*
Updating:
Let $X_1 = X \cup \{X_\emptyset\}$
Let $IG(X_\emptyset)$ *be the most frequent predicted class in S*
for each (x, y) in S **do**
  Sort (x, y) into leaf l using DT
  for each x_{ij} in x such that $X_i \in X_1$ **do**
    Increment $n_{ijy}(l)$
    Label l with the majority class among $n_{ijk}(l)$
  end for
  if (examples at l are not in the same class and $n_{ijy}(l)$ mod $n_{min} = 0$) **then**
    Compute $IG(X_i)$ *for each feature using* $n_{ijk}(l)$
    Select the highest two $IG(X_{i1})$ *and* $IG(X_{i2})$*:*
    $\triangle IG = IG(X_{i1}) - IG(X_{i2})$
    Compute $\epsilon = \sqrt{\frac{R^2 \ln(1/\delta)}{2n_{ijk}(l)}}$
    if *(*$\triangle IG > \epsilon$ *or* $\triangle IG \leq \epsilon < \tau$ *and* $X_{i1} \neq X_\emptyset$*)* ***then***
      Add new split to l with X_{i1} *and have a new leaf* l_m
      Let $n_{ijk}(l_m) = 0$
    end if
  end if
end for
Return VFDT

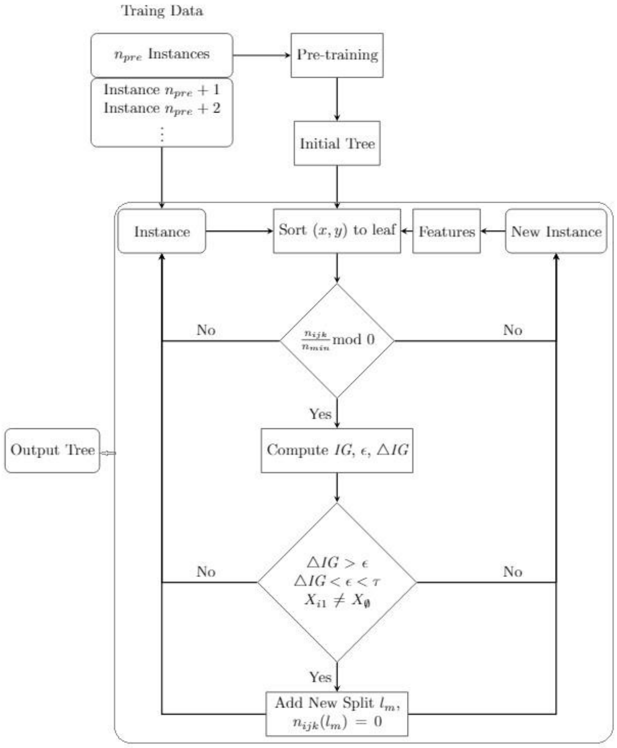

Fig. 2. VFDT with DT Pre-training Diagram.

IV. RESULTS

We used the training observations to extract the features by Decision Tree. By selecting the best depth among 2 to 10 and using the 10-folds cross validation to avoid the overfitting, we got the best tree with an average validation accuracy of 87.36% from all the 7352 observations, with maximum depth at 7 and with 36 features. Then we used these 36 features to create an online tree. After preparing the training data as the streaming data, we fixed the minimum number of checking if a new splitting is needed, n_{min}=20. As time goes by, the tree will be more and more deep until it runs out of the lab-data or the threshold of the information gain. In our experiments, the tree will be paused after reading 7352 records. We call this as lab step, which prepares the model and system. The result of the model will be built into the single device. Next, testing data from 9 new volunteers will be used. This step generates two types of data: with labeled activities and without labels. We used the labeled records to continually update the tree model to be more personal and used the unlabeled records to evaluate model performance. The finally results we got from VFDT with an overall average accuracy for 9 subjects together is 85.9% (without personality). While for single subject self, some of them have lower average accuracy, such as Subject 4 only has 60.5%, the main problem for recognizing the right activities is Walking Upstairs. It only has 8% of the accuracy. The accuracy for Subject 7 with Walking Downstairs is even worse. Some of them performed much better than the overall average, such as Subject 6, it achieves 99.4% of accuracy. The details are shown in Table IV.

These results indicate that the activities are varied from person to person, and it is necessary to import personal activity pattern at the beginning and update to the personal model from the general case. Take four activities sequence plots for examples. In Fig. 5, we can visualize that for Static activities, Sitting, Standing, and Lying Down, the 3-axis of total acceleration gave enough information for identifying them. But the Sitting and Standing do not have many differences for most of the volunteers, such as in Fig.4. The rest of 3 activities are more complex as the changes between them are tiny, such as in Fig. 4 and Fig. 6.

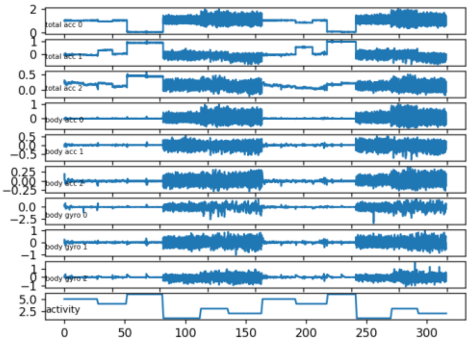

Fig. 3. Example Sequence for Subject 2.

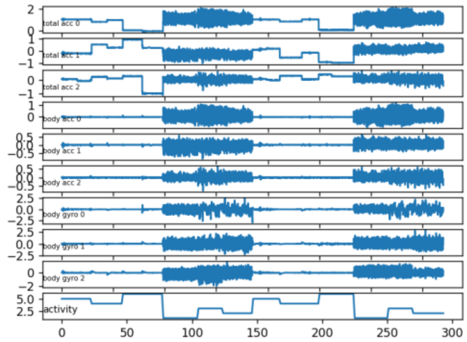

Fig. 4. Example Sequence for Subject 4.

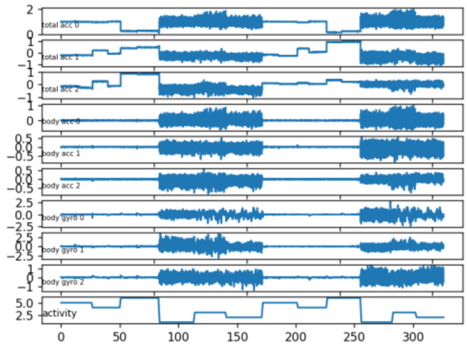

Fig. 5. Example Sequence for Subject 6.

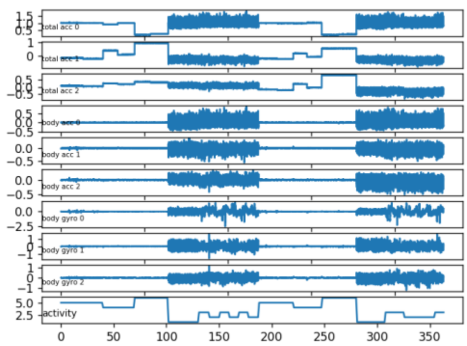

Fig. 6. Example Sequence for Subject 7.

To show the ability of updating our model to new activities, we use another data set from WISDM Lab [4]. Although these two data types are different, it can roughly show the power. This data has 36 volunteers who performed a new activity Jogging instead of Lying Down. Moreover, the data transform method is different, thus the data only has 44 features including the single axis. To keep the same number of attributes, we selected the last 36 ones since the decision tree method shows that the most important attributes are the last ones. By randomly selecting only 9 of all the volunteers as the training, we evaluated our model with Jogging. The average accuracy of all the 25 test subjects for Jogging is 84%. The accuracy for one single person can go up to 100% and 16 out of 25 accuracy rates are higher than 90%. More details can be found in Table V. This proves that our model can learn new activities which are not present in the training dataset. This is one of the big differences from all the other models so far.

TABLE IV. ACCURACY FROM VFDT WITH 36 FEATURES

Subject	W	WU	WD	ST	SD	LD	Average
Sub 1	1.0	1.0	1.0	0.45	0.91	1.0	0.903
Sub 2	0.97	1.0	1.0	0.62	0.93	1.0	0.915
Sub 3	1.0	0.96	1.0	0.36	0.65	1.0	0.826
Sub 4	0.69	0.08	1.0	0.60	0.24	1.0	0.605
Sub 5	0.78	0.88	1.0	0.39	0.94	1.0	0.849
Sub 6	0.96	1.0	1.0	1.0	1.0	1.0	0.994
Sub 7	0.96	1.0	0.00	0.48	1.0	1.0	0.761
Sub 8	0.29	1.0	0.87	0.37	0.66	1.0	0.707
Sub 9	1.0	0.8	0.89	1.0	0.97	1.0	0.948
Average[a]	0.92	0.87	0.77	0.65	0.91	1.0	0.859

[a] Average means the average acc. we got by testing all the 9 subjects together.

TABLE V. ACCURACY FOR JOGGING WITH WISDM DATA

Sub.	Acc.	Sub.	Acc.	Sub.	Acc.	Sub.	Acc.
Sub 1	0.98	Sub 8	0.79	Sub 17	0.98	Sub 24	0.97
Sub 2	0.98	Sub 9	0.66	Sub 18	0.95	Sub 25	0.13
Sub 3	0.46	Sub 11	0.39	Sub 19	0.94	Sub 26	0.97
Sub 4	0.98	Sub 12	0.97	Sub 20	0.36	Sub 27	0.98
Sub 5	0.93	Sub 13	0.84	Sub 21	0.99		
Sub 6	0.98	Sub 14	0.98	Sub 22	0.80		
Sub 7	1.00	Sub 15	0.96	Sub 23	0.96	Average[b]	0.84

[b] Average means the average acc. we got by testing all the 25 subjects together.

V. CONCLUSION

To provide a human activity recognition system with automatic updating and adjusting, an online system is required. Most of the methods in the literature are offline, while other online methods do not have this ability. In this paper, we proposed and evaluated the VFDT to identify existing activities online and to recognize new activities when new labeled data available.

The results show that the average accuracy is 85.9% for identifying 6 activities, and 4 out of 9 accuracy rates for single person are above 90%. It can recognize Lying Down with 100% accuracy. For a new activity, VFDT gives an average of 84% accuracy rate and above 90% accuracy for 64% of the testing people.

REFERENCES

[1] M. Haghi, K. Thurow, and R. Stoll, "Wearable devices in medical internet of things: scientific research and commercially available devices, " Healthcare informatics research 23.1, 2017, pp. 4-15.

[2] UCI Machine Learning Web, https://archive.ics.uci.edu/ml/datasets
[3] WISDM Lab Data, http://www.cis.fordham.edu/wisdm/dataset.php
[4] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," ACM SigKDD Explorations Newsletter 12.2, 2011, pp. 74-82.
[5] F. Foerster, M. Smeja, and J. Fahrenberg, "Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring," Computers in Human Behavior 15.5, 1999, pp. 571-583.
[6] A. Sharma, Y. Lee, and W. Chung, "High accuracy human activity monitoring using neural network," 2008 Third International Conference on Convergence and Hybrid Information Technology. Vol. 1. IEEE, 2008.
[7] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Cell phone-based biometric identification," 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS). IEEE, 2010.
[8] Z. He and L. Jin, "Activity recognition from acceleration data based on discrete consine transform and SVM," 2009 IEEE International Conference on Systems, Man and Cybernetics. IEEE, 2009.
[9] W. Sohn and Z. A. Khan. "Real Time Human Activity Recognition System based on Radon Transform," IJCA special issue on artificial intelligence techniques novel approaches practical applications," AIT, 2011.
[10] W. Wu, et al., "Classification accuracies of physical activities using smartphone motion sensors," Journal of medical Internet research 14.5 ,2012.
[11] D. Anguita, et al., "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," International workshop on ambient assisted living. Springer, Berlin, Heidelberg, 2012.
[12] B. Fergani, "Evaluating a new classification method using PCA to human activity recognition," 2013 International Conference on Computer Medical Applications (ICCMA). IEEE, 2013.
[13] L. Zhang, X. Wu, and D. Luo, "Human activity recognition with HMM-DNN model," 2015 IEEE 14th International Conference on Cognitive Informatics Cognitive Computing (ICCI* CC). IEEE, 2015.
[14] Q. Guo, B. Liu, and C. W. Chen, "A two-layer and multi-strategy framework for human activity recognition using smartphone," 2016 IEEE International Conference on Communications (Icc). IEEE, 2016.
[15] C. A. Ronao, and S. B. Cho, "Human activity recognition with smart-phone sensors using deep learning neural networks," Expert systems with applications 59, 2016, pp. 235-244.
[16] K. Nakano, and B. Chakraborty, "Effect of dynamic feature for human activity recognition using smartphone sensors," 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST). IEEE, 2017.
[17] A. Ignatov, "Real-time human activity recognition from accelerometer data using Convolutional Neural Networks," Applied Soft Computing 62, 2018, pp. 915-922.
[18] G. Hulten, L. Spencer, and P. Domingos, "Mining time-changing data streams," Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2001.
[19] N. Gyorbiro, A. Fabian, and G. Homanyi, "An activity recognition system for mobile phones," Mobile Networks and Applications 14.1, 2009, pp. 82-91.
[20] L. Wang, et al., "A hierarchical approach to real-time activity recognitionin body sensor networks," Pervasive and Mobile Computing 8.1, 2012, pp. 115-130.
[21] G. Okeyo, et al., "Dynamic sensor data segmentation for real-time knowledge-driven activity recognition," Pervasive and Mobile Computing 10, 2014, pp. 155-172.
[22] Y. Al Jeroudi, et al., "Online sequential extreme learning machine algorithm based human activity recognition using inertial data," 2015 10th Asian Control Conference (ASCC). IEEE, 2015.
[23] S. Na, Kandethody M. Ramachandran, and M. Ji. "Online Bayesian Kernel Segmentation," unpublished.
[24] P. Domingos, and G. Hulten, "Mining high-speed data streams," Kdd. Vol. 2. 2000.
[25] T. Minegishi, and A. Niimi, "Detection of fraud use of credit card by extendedVF DT," 2011 World Congress on Internet Security (WorldCIS 2011). IEEE, 2011.
[26] T. Minegishi, and A. Niimi, "Proposal of credit card fraudulent use detection by online-type decision tree construction and verification of generality," International Journal for Information Security Research (IJISR) 1.4, 2011, pp. 229-235.
[27] L. Cohen, et al., "Real-time data mining of non-stationary data streams from sensor networks," Information Fusion 9.3, 2008, pp. 344-353.
[28] M. M. Gaber, S. Krishnaswamy, and A. Zaslavsky, "On-board mining of data streams in sensor networks," Advanced methods for knowledge discovery from complex data. Springer, London, 2005, pp. 307-335.
[29] J. Gama, and R. U. Pedersen, "Predictive learning in sensor networks," Learning from Data Streams. Springer, Berlin, Heidelberg, 2007, pp. 143-164.
[30] M. V. V. S. Subrahmanyam, and R. V. Venkateswara, "VFDT Algorithmfor Decision Tree Generation," International Journal for Development of Computer Science and Technology (IJDCST) 1.7 2013.
[31] S. Desai, et al., "Very fast decision tree (vfdt) algorithm on hadoop," 2016 International Conference on Computing Communication Control and automation (ICCUBEA). IEEE, 2016.
[32] J. R. G. Lansangan, and E. B. Barrios, "Principal components analysis of nonstationary time series data," Statistics and Computing 19.2 2009, pp. 173.
[33] E. Alpaydin, Introduction to machine learning, MIT press, 2014.
[34] G. Hulten, L. Spencer, and P. Domingos, "Mining time-changing data streams," Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2001.
[35] L. E. Raileanu and K. Stoffel, "Theoretical comparison between the gini index and information gain criteria," Annals of Mathematics and Artificial Intelligence 41.1, 2004, pp. 77-93.
[36] L. Breiman, Classification and regression trees, Routledge, 2017.