A Hybrid Approach for Human Activity Recognition with Support Vector Machine and 1D Convolutional Neural Network

Md Maruf Hossain Shuvo*, Nafis Ahmed*, Koundinya Nouduri*, Kannappan Palaniappan
Department of Electrical Engineering and Computer Science
University of Missouri, Columbia, MO 65211, USA
E-mails: {msp8c, nax35, knfdt}@mail.missouri.edu , palaniappank@missouri.edu

2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) | 978-1-7281-8243-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/AIPR50011.2020.9425332

Abstract*—The Human Activity Recognition (HAR) is a pattern recognition task that learns to identify human physical activities recorded by different sensor modalities. The application areas include human behavior analysis, ambient assistive living, surveillance-based security, gesture recognition, and context-aware computing. The HAR remains challenging as the sensor data is noisy in nature and the activity signal varies from person to person. To recognize different types of activity with a single classifier is often error-prone. To mitigate this problem, we introduced an adaptive human activity recognition model. We present a two-stage learning process to recognize human activity recorded using a waist-mounted accelerometer and gyroscope sensor. In the first step, we classify activity into static and moving, using a Random Forest (RF) binary classifier. In the second step, we adopt a Support Vector Machine (SVM) to identify individual static activity and 1D Convolutional Neural Network (CNN)-based deep learning model for individual moving activity recognition. This makes our approach more robust and adaptive. The static activity has less frequency variation in features compared to dynamic activity waveforms for CNN to learn. On the other hand, SVM demonstrated superior performance to recognize static activities but performs poorly on moving, complex, and uncertain activity recognition. Our method is similarly robust to different motion intensity and can also capture the variation of the same activity effectively. In our hybrid model, the CNN captures local dependencies of activity signals as well as preserves the scale invariance. We achieved 97.71% overall accuracy on six activity classes of widely accepted benchmark UCI-HAR dataset.

***Index Terms*—human activity recognition, support vector machine, convolutional neural network, hybrid method**

I. INTRODUCTION

Human activity recognition (HAR) gained significant importance in research community as it is a challenging time series classification task. In human activity prediction, firstly the sensor data is recorded for activities of specific subjects, then a machine learning model is trained to generalize the model for unseen data. There are lots of applications of activity predictions like behavior analysis, health and workout monitoring, gait analysis, interactive gaming, gesture recognition, video surveillance etc. HAR also has many applications to improve elderly people's living. For example, through continuous monitoring an effective HAR system can ensure the proper medication, physical activity as well as recognizing the diseased conditions.

*contributed equally

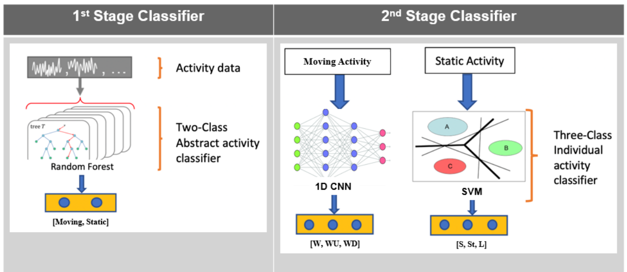


Fig. 1. Overview of the proposed two-stage hybrid method for human activity recognition(HAR). In the first stage we used a Random Forest (RF) classifier to get the high class label. Then each class separately feed into either in 1D CNN or in the SVM classifier depending on the high level class.

One of the main challenges of activity prediction is to generalize the model for different problem, sensors, and activities. Activity signal may vary significantly for different human, even the same person may do the same activity differently in another time. Similarly, different activities may have similar signal pattern which may confuse the learning process. Other challenges include the computational cost to implement in embedded and portable devices, accurate data annotation, variety of complex daily activities, and ensuring privacy of the subjects. Traditional pattern recognition based HAR requires to extract problem specific features to fit a machine learning model. Deep learning (DL) makes the task easy and adoptable by automatically learning the features. In a DL approach, it also can extract high-level features in deep layer that makes it appropriate for complex activity recognition.

There are lots of new innovative ideas, and experiments are ongoing with model architecture to predict human activity more accurately [1]. The deep learning architectures used for HAR can be categorized into six major types: Deep Neural Network (DNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Deep Belief Network (DBN), Stacked Autoencoder (SAE), and Hybrid Models. In [2], they worked on 6 class activity of UCI-HAR dataset [3] and they stacked the raw signal row by row to form an image. Then applied the 2D Discrete Fourier Transform (DFT) to signal image and choose magnitude to form the activity image. They showed that the images vary for different activity and this visual difference indicates their potential for Deep Convolutional Neural Network (DCNN) to extract discriminate image features. In [4], authors first divided activities into dynamic and static categories. Then used CNN models to discriminate between activities of a particular category. Their limitations is, for dynamic activity the CNN model performed better compared to the static activity. In [5], they worked with CNN models and tried to get an optimized model by adopting the number of layers, filter size, pooling size, as well as tuned various hyperparameters. For complex activity recognition it is difficult for a single model to detect all activities accurately and effectively [6]. Thus, several hybrid approaches have been proposed over the last few years [7] [8]. One of the very first hybrid approach was the combination of Linear Discriminant Analysis (LDA) with the Artificial Neural Network proposed in [9]. This method first distinguishes between the static and dynamic state at lower level and then recognize more specific activities. In [10] a Hierarchical Hidden Markov Model (HMM) with two-stage was proposed to recognize five types (stand, walk, stair up, stair down and run) of human action and three type of human activities (shopping, taking bus, and moving by walking). Another approach is the two-step method using two-level Continuous HMMs (CHMMs) in [11]. In the first level two CHMM was used to identify the stationary and moving activities, and in the second-level six CHMMs used to classify six specific activities of UCI-HAR dataset. In [12] author proposed a hybrid model combining the SVM and HMM and showed better performance than signle SVM or neural network.

Although these hybrid methods sometimes produce better results but most of the methods are complex and computationally inefficient. In this work, we combined the power of classical machine learning approach: Random Forest(RF) and the Support Vector Machine (SVM) along with the deep Convolutional Neural Networks (CNN). Our method works well for both stationary activities and complex moving activities. To take the full advantage of SVM and CNN, we first divide the static and moving activities using Random Forest and then used SVM for static activity and CNN for moving activity recognition as illustrated in Fig. 1. To get better accuracy, adapting the hyperparameters to fit data well is imperative. That's why we tuned the number of layers, architecture, and other hyperparameters. To compare the performance of different methods, we use recognition accuracy as a comparison metric as it is the only common metric used in all HAR methods.

Our major contributions are (a) we present a hybrid approach for human activity prediction that effectively utilizes the power of Support Vector Machine and deep Convolutional Neural Network (CNN), (b) by separating static and moving activities and using appropriate methods for different abstract class, we are getting higher accuracy for all type of activities, (c) we demonstrate that such approach can be robust and the average inference time (below 100 μs) indicates that our method can be adapted for real-time applications.

For our implementation we have used Python as programming language and TensorFlow as the deep learning framework. To evaluate the time complexity, we tested all our models in the same Windows machine having Intel Core i5 processor with 8 GB ram and 64-bit operating systems.

The rest of the paper is organized as follows: in section II the hybrid method for human activity recognition as well as our adapted models are presented. In section III, we presented our results and analysis, and in section IV conclusions are drawn and future directions are discussed.

II. Human Acitivity Recognition using Hybrid Method

A. Support Vector Machine (SVM) for Static Human Activity Recognition

The support vector machine (SVM) finds the optimal hyperplane that sufficiently separates the data originally developed for binary classification [13]. The SVM has some advantages over other machine learning methods like it is effective in high dimensional space, uses a subset of training points making it memory efficient, and can effectively work where the dimensions are greater than number of samples. In the training stage, the hyper-planes are calculated to find different labels of the training data. There are two approach to utilize SVM for multi-class classification problem: one-vs-all, and one-vs-one. The first method considers all classes in a single optimization problem and is computationally more expensive. This method requires number of SVM equal to the number of classes. One the other hand, one-vs-all is constructing and combining binary classifier in a way to make it work as a multi-class classifier. Besides improving the classification accuracy, this approach also reduces the computational cost. Thus, in our implementation we adopted one-vs-one approach to classify the static activities.

B. Convolutional Neural Network (CNN) for Moving Human Activity Recognition

The Convolutional Neural Network (CNN) is widely used in image analysis but it demonstrated promising results for other type of signals like speech recognition, text analysis, and human activity prediction [14]. In HAR, CNN offers two major benefits over other models: local dependency and scale invariance [15]. In this work we have used 1D CNN as the CNN layer to exploit this advantage. The deep CNN for human activity recognition generally follows a similar architecture having multiple convolution layer, non-linear transformation, pooling, and finally a fully connected layer [16]. The convolutional layers uses a set of kernels to automatically generate the feature maps that are the input to the subsequent layers. There is some approximation function that further purify the feature maps such as rectified linear unit (ReLU), sigmoid, hyperbolic tangent etc. The pooling operation keeps the strong features in a kernel, thus removes the weak features and reduces the feature dimension.In [5] authors illustrated that

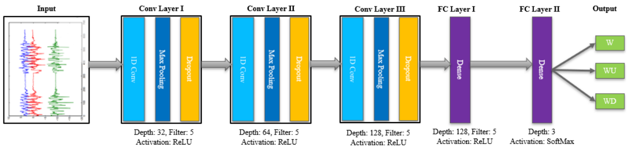


Fig. 2. 1D CNN architecture for moving activity classification. We have used three 1D convolution layer (each consisting of a convolution, max-pooling, and dropout), and two fully connected layer with ReLU activation in all hidden layers and softmax activation at the output layer.

more layers increase the model complexity and after third convolution layer adding more layers decrease performance rather than improving. Our deep 1D CNN architecture is presented in Fig. 2. The fully connected layer at the last layer with softmax activation gives the prediction probability for our three moving activity classes (walking, walking upstairs, and walking downstairs). As we are dealing with time series data, therefore, all our convolutions are 1D CNN. We tuned the model architecture and hyper-parameters to get the best performance. We have used three convolution layers, because three layers performed best and adding more layer increases the model complexity with little or no improvement in performance.

C. Hybrid Approach for Human Activity Recognition

In our work, we developed a hybrid model with one deep 1D CNN and one SVM to predict the human activity as shown in Fig. 3. This hybrid model is a good choice because the CNN captures the spatial relation between signals and the SVM captures the spatio-temporal relationship. Together it enhances the ability to recognize different activities that have varied signal distributions. Our model first identifies the static and moving activity using a Random Forest (RF) binary classifier. The RF combines many decision trees into a single model and uses the average of multiple trees or compute the majority votes to make a prediction in the terminal leaf. In our problem, RF provides better result (100% accuracy) than other binary classifier thus we selected RF for high-level classification. After identifying the abstract level of static or moving activities, we send static activities to SVM and moving activities to the CNN model for appropriate classification. Our

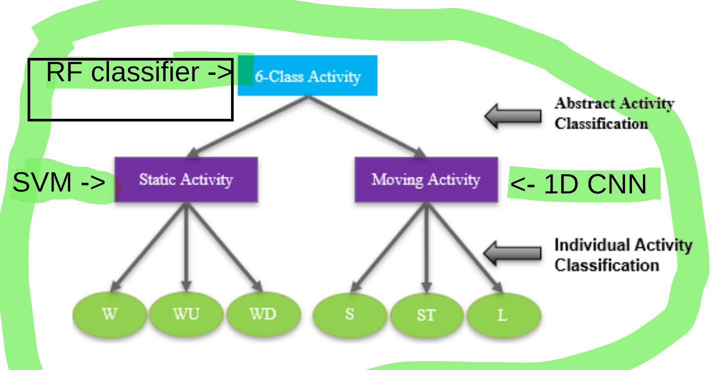


Fig. 3. Illustration of hybrid method for human activity classification. From six total activity we classify the type of activity using a binary classification. Features of each class are fed into separate classifiers to identify individual activity.

SVM is a one-vs-one three class classification model with degree 7 ploy svc kernel. The multi-class SVM model is trained with our 561 feature vectors extracted from the raw activity signals. We developed the deep learning CNN model to recognize the moving activities. Our CNN model has three convolution layers, each convolution layer 1D filter size is 5×5 with ReLU activation followed by max-pooling and 20% dropout.

TABLE I
SUMMARY OF TRAINING AND TEST SAMPLE SPLIT USED IN OUR EXPERIMENTS. W: WALKING, WU: WALKING UPSTAIRS, WD: WALKING DOWNSTAIRS, S: SITTING, ST: STANDING, L: LAYING

Class	Moving			Static			Total
	W	**WU**	**WD**	**S**	**St**	**L**	
Train	1226	1073	986	1286	1374	1407	7352
Test	496	471	420	491	532	537	2947

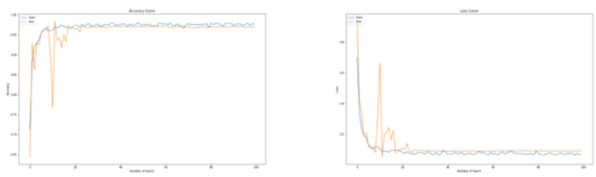
TABLE II
CONFUSION MATRIX FOR ABSTRACT ACTIVITY CLASSIFICATION

Confusion Matrix		**Predicted Label**	
		Static	Moving
True Label	Static	1560	0
	Moving	0	1387

III. EXPERIMENTAL RESULTS

A. Datasets

In this research, we used the UCI-HAR dataset [3]; the recording of 30 subjects having an age range from 19 to 48 years. The dataset consists of the six activity signals of daily living obtained by a waist mounted smartphone following the activity protocol. The data acquisition uses the smartphone's accelerometer and gyroscope. The tri-axial (x, y, z) data of activities are: walking, walking-upstairs, walking-downstairs, sitting, standing, and laying. After acquisition, the data were sampled at $50Hz$ and separated into 128 windows with 50% overlapping. There are total 9 channels of gyroscope and accelerometer: (i) 3 channel body accelerometer, (ii) 3 channel total accelerometer, and (iii) 3 channel body gyroscope. Each channel recorded 128 real valued vectors to depict an activity. The dataset also includes the accurate labels. The sensor data are noisy in nature. Thus, a noise removal filter removed the noise and a low pass Butterworth filter



(a) Learning accuracy plot (b) Learning loss plot

Fig. 4. Training performance curve for 1D CNN: training and validation (a) accuracy, and (b) loss with number of epochs.

TABLE III
CONFUSION MATRIX FOR HUMAN ACTIVITY RECOGNITION ON UCI-HAR DATASET USING PROPOSED HYBRID METHOD

Confusion Matrix		**Predicted Label**						
	Activities	Sitting	Standing	Laying	Walking	Walking Upstairs	Walking Downstairs	**Recall(%)**
True Label	Sitting	449	41	1	0	0	0	91.4
	Standing	8	524	0	0	0	0	98.5
	Laying	0	0	537	0	0	0	1.0
	Walking	0	0	0	490	4	2	98.8
	Walking Upstairs	0	0	0	2	469	0	99.6
	Walking Downstairs	0	0	0	2	9	409	97.4
Precision(%)		98.2	92.7	99.8	99.2	97.3	99.5	
F1 Score(%)		94.7	95.5	99.9	99.0	98.4	98.4	

separated the gravitational and body motion components from the acceleration signal. Along with the raw sensor signals, the dataset also includes a very well-engineered set of 561 features calculated from 128 reading of each channel. Our training set consists of 7352 samples and we used 20% of the training samples for model validation. During training we used early stopping and model checkpoint to get the best learned model at minimum validation loss and maximum validation accuracy. We evaluated the model performance on our test set that consists of 2947 samples. Table I shows the summary of our training and test samples split.

TABLE IV
PERFORMANCE COMPARISON ON UCI-HAR DATASET

Model	**Accuracy (%)**
SVM [3]	96.37
SVM+HMM [17]	97.60
LSTM-CNN [18]	95.78
Bidirectional-LSTM [19]	92.67
CNN-LSTM [20]	93.40
1D CNN [5]	94.79
DeepCNN [2]	97.59
TFFT + CNN [5]	95.75
DeepConvLSTM [21]	95.80
Stacked Autoencoder [22]	92.16
CHMM [23]	93.18
Proposed hybrid method	**97.66**

B. Implementation Details and Results

To classify the activity at a higher level we used the RF classifier and achieved an accuracy of 100%. The confusion matrix for static and moving activity classification is presented in Table II. Depending on the abstract activity class, in the next step, the static activity features are fed into the SVM, and moving activity features are fed into the 1D CNN classifier. We trained the SVM model with a poly kernel of degree 7. The total training time taken for SVM is below 5 seconds. The overall accuracy of 96.8% has been achieved for the static activity classification. As shown in the confusion matrix in Table III, we achieved a precision of 98.2%, 99.8%, and 92.7% for the sitting, lying, and standing class classification. For moving activity recognition, we designed an optimized deep 1D CNN architecture capable of recognizing the walking, walking upstairs, and walking downstairs activity. We trained the 1D CNN with a Stochastic Gradient Descent (SGD) optimizer with a categorical cross-entropy loss function. The learning performance curve for 1D CNN is presented in Fig. 4 that shows a better learning behavior with no overfitting with increasing accuracy and decreasing loss values with the number of epochs. The total training time taken for 100 epochs is 170 seconds. We are getting an overall accuracy of 98.6% for moving activity classification. Precision for walking, walking-upstairs, and walking-downstairs is 99%, 97.3%, and 98.6% respectively. The recall values for each of the activity classes are also presented in the Table III. Combining both static and moving activity classification, we are achieving an overall accuracy of 97.66% for UCI-HAR data which outperforms or comparable in performance with state-of-the-art machine learning and deep learning methods as illustrated in Table IV. Therefore, we are getting a very robust result with our two-stage hybrid model combining the power of machine learning and deep learning for human activity classification.

IV. CONCLUSIONS

In this paper, we presented a hybrid method to effectively perform human activity recognition. Our method first identifies the abstract activity by using a Random Forest classifier to identify the activities type as static and moving. For static activity's specific recognition we used support vector machine and for moving activities we designed a deep 1D CNN. We achieved an overall accuracy of 97.71% which is comparable to state-of-the-art performance. Our future plan is to deploy our implementations into low-power integrated circuits to make it well-suited for wearable sensors to identify wide variety of activities in real-time.

ACKNOWLEDGMENT

This work is partially supported by awards from U.S. NIH National Institute of Neurological Disorders and Stroke R01NS110915 (KP), and the U.S. Army Research Laboratory project W911NF-18-2-0285. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U. S. Government or agency thereof.

REFERENCES

[1] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," Pattern Recognition Letters, vol. 119, pp.3-11, 2019.

[2] W. Jiang, and Z. Yin, “Human activity recognition using wearable sensors by deep convolutional neural networks,” ACM international conference on Multimedia, pp. 1307-1310, 2015.
[3] D. Anguita, A. Ghio, L. Oneto, X. Parra, J.L. Reyes-Ortiz, “A public domain dataset for human activity recognition using smartphones,” European Symposium on Artificial Neural Networks, pp. 437-442, 2013.
[4] H. Cho, and Y. SM, “Divide and conquer-based 1D CNN human activity recognition using test data sharpening,” Sensors, vol. 18, no. 4, p.1055, 2018.
[5] C.A. Ronao, and S.B. Cho, “Human activity recognition with smartphone sensors using deep learning neural networks,” Expert systems with applications, vol. 59, pp. 235-244, 2016.
[6] H.H, Hsu, C.T. Chu, Y. Zhou, and Z. Cheng, “Two-phase activity recognition with smartphone sensors,” International Conference on Network-Based Information Systems (NBiS), pp. 611–615, 2015.
[7] J. Okai, and M. Beekman, A. Knobbe, and C. R. de Sá, “Building robust models for human activity recognition from raw accelerometers data using gated recurrent units and long short term memory neural networks,” International Conference of Engineering in Medicine and Biology Society, pp. 2486-2491, 2019.
[8] G. Filios, S. Nikoletseas, C. Pavlopoulou, M. Rapti, and S. Ziegler, “Hierarchical algorithm for daily activity recognition via smartphone sensors,” World Forum on Internet of Things (WF-IoT), pp. 381–386, 2016.
[9] D. K. Vishwakarma, and R. Kapoor, “Hybrid classifier based human activity recognition using the silhouette and cells,” Expert Systems with Applications, vol. 42, no. 20, pp. 6957–6965, 2015.
[10] Y.S. Lee,and S.B Cho, “Activity Recognition Using Hierarchical Hidden Markov Models on a Smartphone with 3D Accelerometer,” International Conference on Hybrid Artificial Intelligent Systems (HAIS’11), pp. 460–467, 2011.
[11] C. Yang,Z Wang, B. Wang, S. Deng, G. Liu, Y. Kang, and H. Men, “CHAR-HMM: An Improved Continuous Human Activity Recognition Algorithm Based on Hidden Markov Model,” International Conference on Mobile Ad-Hoc and Sensor Networks, pp. 271-282, 2017.
[12] A. M. Khan, Y.K Lee, S.Y. Lee,and T.S. Kim, “A Triaxial Accelerometer-Based Physical-Activity Recognition via Augmented-Signal Features and a Hierarchical Recognizer,” IEEE Trans. Inf. Technol. Biomed, vol. 14, pp. 1166–1172, 2010.
[13] J.A. Suykens, and J. Vandewalle, “Least squares support vector machine classifiers,” Neural processing letters, vol. 9, no. 3, pp. 293-300, 1999.
[14] S.M. Lee, and S. M. Yoon, and H. Cho, “Human activity recognition from accelerometer data using convolutional neural network,” International Conference on Big Data and Smart Computing, pp. 131-134, 2017.
[15] M. Zeng et.al., “Convolutional neural networks for human activity recognition using mobile sensors,” International Conference on Mobile Computing, Applications and Services, pp. 197-205, 2014.
[16] A. Ignatov, “Real-time human activity recognition from accelerometer data using convolutional neural networks,” Applied Soft Computing,vol. 62, pp. 915-922, 2018.
[17] K. Davis, E. Owusu, V. Bastani, L. Marcenaro, J. Hu, C. Regazzoni, and L. Feijs, “Activity recognition based on inertial sensors for ambient assisted living,” International conference on information fusion (fusion), 2016, pp. 371–378.
[18] K. Xia, J. Huang, and H. Wang, “LSTM-CNN Architecture for Human Activity Recognition,” IEEE Access, vol. 8, pp. 56855-56866, 2020.
[19] F. Hernández, L. F. Suárez, J. Villamizar, and M. Altuve, “Human activity recognition on smartphones using a bidirectional lstm network”, Symposium on Image Signal Processing and Artificial Vision (STSIVA), pp. 1-5, 2019.
[20] S. Deep, and X. Zheng, “Hybrid Model Featuring CNN and LSTM Architecture for Human Activity Recognition on Smartphone Sensor Data,”International Conference on Parallel and Distributed Computing, Applications and Technologies, pp. 259-264, 2019.
[21] F. J. Ordóñez, and D. Roggen, “Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition,” Sensors, vol. 16, no. 1, p.115, 2016.
[22] Y. Li, D. Shi, B. Ding, and D. Liu, “Unsupervised feature learning for human activity recognition using smartphone sensors,” Mining intelligence and knowledge exploration, pp. 99-107, Springer, 2014.
[23] C.A. Ronao,and S.B. Cho, “Recognizing human activities from smartphone sensors using hierarchical continuous hidden Markov models,” Int. J. Distrib. Sens. Netw., vol. 13, pp. 1–16, 2017.