October 2021

Online and Adjusted Human Activities Recognition with Statistical Learning

Yanjia Zhang
University of South Florida

Online and Adjusted Human Activities Recognition with Statistical Learning

by

Yanjia Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
with a concentration in Statistics
Department of Mathematics and Statistics
College of Arts and Sciences
University of South Florida

Major Professor: Kandethody Ramachandran, Ph.D.
Chris P. Tsokos, Ph.D.
Sherwin Kouchekian, Ph.D.
Lu Lu, Ph.D.
Yicheng Tu, Ph.D.

Date of Approval:
October 26, 2021

Keywords: HAR, Online Analysis, Adaptive Learning, Smart Device, Machine Learning

Dedication

This doctoral dissertation is dedicated to my family, especially

To my father: Zhongjun Zhang

To my mother: Wanzhen Qian

To my sister: Jiaheng Zhang

Acknowledgments

Above all, I offer my genuine gratitude to my major advisor Professor Kanadethody M. Ramachandran, who inspired me with his support and guidance. Without his advice and help, I could not have developed the knowledge and accomplished this dissertation. He has not only guided me into the statistical learning field but also taught me the way to learn new things which will allow me to continue my work and embrace new knowledge.

I am also truly grateful to the members of my dissertation committee: Professors Chris P. Tsokos, Lu Lu, and Yicheng Tu for their time and kind assistance during my dissertation preparation. Specifically, Dr. Lu's courses taught me how to apply statistical models and experiments from the theoretical to the real world. And her well prepared and organized classroom presenting added positively to my educational experience. Also I am truly grateful to my defense chair, Professor Manish Agrawal for his time and support for my defense.

My heartiest appreciation also goes to all the faculties and staff of the department of Mathematics Statistics who contributed in diverse ways to make my experience a token of memory at University of South Florida.

I would like to express my thanks to my father, mother, and sister for their many manifold blessings and supports to me. Without them I would not have been able to complete the pursuit of my doctoral studies.

Table of Contents

List of Tables

List of Figures

Abstract

Wearable human activity recognition (HAR) is a widely application system for our daily life. It has been built in many devices, such as smartphone, smartwatch, activity tracker, and health monitor. Many research try to develop a system which requires less memory space and power, but has fast and accurate classification results. Moreover, the objective of adjusting the classifier by the system self is also a study direction. In the present study, we introduced the machine learning methods to both smartphone data and smartwatch data and an adjusted model with the continuous generating data. Further, we also proposed a new HAR system which could adjust it by customer's personal input.

First, we present a comparison with several popular machine learning methods using smartphone data and try to find the most effective model for identifying activities, such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA), Decision Tree (DT), and so forth. Two datasets with different transition methods from smartphone are used. Also, we used grid search, multi-fold cross validation, and dimensional reduction method to improve the performance. Meanwhile, a two-layer method for activity identification is proposed. This method is more flexible of choosing classifiers for activities.

Then, in order to avoid the fixed built in HAR system, we used an online method called Very Fast Decision Tree (VFDT) to mimic the real scenario, since we do not have data collected in a streaming basis. There are two main improvements from the existing models: 1) we train the model online and use the training data for training and adjusting then delete the previous data; 2) after building VFDT, the model can be adjusted to identify new activities by adding only small amount of labeled observations.

Finally, we proposed a personalized HAR system with interactive function. With this system, customers could build their personal HAR system by inputting their data. The system includes two steps, unsupervised method and supervised method. The unsupervised step is used for identifying if the new input data has any new activities. K-cluster is applied. The supervised step is used for identifying the specific activities and update the activity classes if there are new activities. Quadratic discriminant analysis (QDA) is applied.

Chapter 1
Introduction

With decades of development on technology in sensors and accelerometers, many different types of devices in various fields are able to have one or more sensors with low cost and less power consumption. Human Activity Recognition (HAR) is one of these fields. Human activity recognition system is built for recognizing the actions for one or more person from a series of sensor observations on the person's movements and/or the movements associated with the surrounding environmental conditions. From some time, researchers in computer science noticed that HAR had the strength of providing personalized support for many different applications, such as in medical assistance, smart home system, healthcare monitoring, and indoor and outdoor activity surveillance. These systems has different methods to collect information based on visual, non-visual, and multi-modal sensor technologies. For example, use the visual videos for living assistance smart homes, accelerometers sensors for healthcare monitoring applications, and multi-modal sensors for indoor and outdoor activity surveillance systems.

The smart home applications are quite useful in the daily life. As Demiris et al. [26] stated that a smart home is an environment equipped with sensors that enhance the safety of residents and monitor their health conditions. As we all know that the continuous improvement of the human's living conditions and medical support cause the increasing of life expectancy. But according to the article [79], the birth rates over 50 years decreased substantially. The human societal structure, such as age pyramid, has been changed a lot, more older person but less young. This results in less people to support the elderly. By monitoring person's behavior, an HAR system is possible to track the health condition of an elderly people or patient, to provide some basic supports, to secure their safety and well-being, and to notify the health personnel in case of an urgent need. Thus, this type of support system is necessary and essential for elderly people or people who like to live independently as long as possible. Recently, HAR system uses the combination of video and wearable sensors and sensors with environment, which will collect sequence of images, audio, accelerometers data, associate with many environmental data, such as light, humidity, temperature, presence, electrical sensors. GER'Home project is one of such smart home system for elderly people [105]. This project aimed to improve elderly life conditions at home and to reduce the costs of long hospitalizations.

By recognizing a person's activities, this system could detect the emergency accidents, e.g. Falls, and even detect the early stage of deteriorated health status and some early illness diagnosis. HERMES is another project for assisting the user in performing everyday tasks and to support independent living. It combines a home-based and mobile device to support the user's cognitive state and prevent age-related cognitive decline, which includes images, audio, and GPS data. It was designed for people who are suffering from mild memory problems and tried to assist them when necessary. Kasteren et al. [85] mentioned in their research about a multi-sensor smart home system, which is used for monitoring and assisting elderly people in their home. This wireless sensor network includes reed switches to measure open–close states of doors and cupboards, pressure mats to measure sitting on a couch or lying in bed, mercury contacts for movement of objects (e.g., drawers), passive infrared (PIR) to detect motion in a specific area; float sensors to measure the toilet being flushed, temperature sensors to measure the use of the stove or shower. There are many different smart home systems with different types of sensors, such as Radio-frequency identification (RFID) [92], [28], and [15], Wireless sensor networks (WSNs) [85], [75], and [84], bio-sensors [23], and so on. All these types of smart home provide assistance to elderly, disorder, and independently living person to improve the life quality.

Another popular HAR application is healthcare monitoring system. Comparing to smart home system, healthcare monitoring system is much simpler and less costly, and it is usually with body-worn sensors, such as accelerometers, gyroscope, and bi-axial sensors. Basically, healthcare monitoring system are designed for medical purposes, such as fall detection, human movement tracking, cognitive assistance, and activity tracking. These are noninvasive, privacy preserving systems. Thus usually, it requires patients to wear the small devices. Once the patient need help, the system could notify to the data center for quick help. For example, Zhu et al. [104] developed a smart assisted living system (SAIL) aimed to help elders and sole seniors. This system combined a body sensor network (BSN), a companion robot, a smartphone, and a remote health provider. The BSN collected the vital signs and sent them to the robot, which inferred the human intentions and health conditions and responded back. Then, if needed, the smartphone served as a gateway to access the remote health-care providers. Another example is for fall detection. Gannapathy et al. [37] stated in their paper that the National Institutes of Health found that 67% of elderly who fall and fail to seek help within 72 hours are unlikely to survive, thus, the personal emergency response system is essential for the elderly. In the article, they also purposed a system used zigbee-based wearable sensor system for an E-safe system for fall detection and notification. It could automatically detect customer's fall with the situation and then notify the correspondents via zigbee technology. Then the external correspondents could get notified via message and email. Similarly, an European Union founded system called CAALYX [16] was also designed for fall detection events and notifying the providers with emergency situation with a wearable device. Better thing is that this system could also report the current medical status and the current location to the emergency team for immediate help. All these small but efficient devices are meaningful for future society. Based on Goldstone's research [40] there are 30% of population in Americans, Canadians, Chinese, and Europeans will be over the age of 60 years in 2050, thus, the individual healthcare could be a big challenge for the whole society. And the healthcare monitoring might be a good solution.

A new application of HAR is about the indoor and outdoor surveillance, especially for the security firms. The basic idea is to automatically identify the human anomalies via camera based technologies. As we all know, the traditional surveillance system are monitored by person through the screen. This becomes more difficult as the increasing number of cameras and complex outdoor environment. For example, it is hard to detect the anomaly in a metro station or an airport with crowed people. The HAR surveillance system could be a solution with non-stop monitor. Brémond [17] proposed a video-based activity-monitoring framework named VSIP. This system combined several motion detectors and trackers in the cameras, based on the environment situations and identification methods, the anomalous behavior, such as fighting and vandalism events, could be recognized and the system will send the alarms to the security. similarly, Chang et al. [20] conducted a research on multiple cameras with complex social interactions, such as schools, public parks, city centers, prisons, and other crowed places. They tried to detect and predict suspicious and aggressive behaviors among groups. There are other specific behaviors surveillance system, such as Fusier et al. [34] proposed a system used in airport, which is able to recognize 50 types of events including activities such as baggage unloading, aircraft arrival preparation, and refueling operation. All these surveillance systems reduce the work and stress under a robust approach for real-world conditions.

Our research is only focused on the daily human action and activity recognition, such as walking, jogging, walking upstairs, walking downstairs, sitting, standing, and so on. Although there are many devices, such as smartwatch from Google, Fitbit, and Apple, have many different activities and body conditions, like swimming, kicking, heat beats, and so on, we only include these 7 simple activities as an example for current research. And later, we will analysis another dataset with 18 activities from smartwatches. This research topic could track back to decades ago and continues to be increasing, Moeslund et al. [67] summarised more than 350 publications from 2000 to 2006. It divided into four research areas: Initialization, Tracking, Pose estimation, and Recognition. And there were 83 publications in Recognition with 3-axes sensors. Ahad et al. [3] mentioned various methods from 2001 to 2008 in activity recognition from 3D and 2D version, including Hidden Markov Model (HMM)-based, Discriminative Conditional Random Field (CRF), Maximum Entropy Markov Models (MEMM), and principal Components Analysis (PCA) as a way for dimensional reduction.

It also mentioned several challenges, such as view-dependent methods might fail from different angles, outdoor activity recognition lacked the prior environment information, and the larger and larger size of dataset was required for recognize increasing activities. Moreover, there are many difficulties with privacy issues in video camera based research [45]. Because of these reasons, the personal-device sensor based human activity representation research might become another approach for avoiding these challenges and issues. With the exceptional development of microelectronics and computer systems, as well as the ability of sensors and mobile devices with unprecedented characteristics, the high computational power, small size, and low cost allow people to interact with the devices as part of their daily living. All of these made the personal-device sensor possible for collecting daily life data [59]. Particularly, the recognition of human activities has become a task of high interest within the field, especially for medical, military, and security applications. For instance, patients with diabetes, obesity, or heart disease are often required to follow a well defined exercise routine as part of their treatments [3]. Therefore, recognizing activities such as walking, running, or cycling becomes quite useful to provide feedback to the caregiver about the patient's behavior. Likewise, it could also be used as a activity alarm, such as patients with dementia and other mental pathologies could be monitored to detect abnormal activities and thereby prevent undesirable consequences.

Chapter 2

Literature Survey

In order to introduce the research on HAR systems with wearable sensors, a taxonomy is used. The definition is from Lara et al. [59]. In the survey, the HAR system with wearable sensors was divided into two levels, the first level was about the response time, either immediate feedback or need more time to identify activities, which could be named online or offline, The second level was about the learning methods, supervised or semi-supervised, respectively. These categories was divided according to the different purposes and methods used in the HAR system. They have different system designs, feature extraction approaches, learning algorithms, and challenges. In this chapter, we summarize several research on the online activity system. The online activity recognition system asks for real-time response with the data, this means the system requires high data processing ability. Some of the research even consider the adjustment with the new labeled data, either new activities or new same activities data. The application of this category system is usually in healthcare, which may need for continuously monitoring with physical or metal pathologies safety, recovery, or protection. some of the online human activity recognition research works are introduced below with Supervised Learning and Semi-supervised or Unsupervised learning. Besides, since the feature selection step is also crucial for getting a decent classification result, such as Incel [52] stated that by utilizing information from orientation and rotation changes could improve the recognition accuracy with the smartphone datasets, this chapter also includes some articles on the methods of selecting features.

2.1 Supervised Learning

In HAR system, labeled sensor data is not a big challenge. Some researchers use cameras to take the video record for the whole process while subjects are performing specific activities, and then manually label activities with time stamps [83]. Some researchers employ a system that ask the participants to select the activity to be performed [58]. In this way, the label will be automatically inputted with the time stamp. With these methods, supervised learning methods could play a great role in activity recognition. Among these research, there are two big groups with different classification methods, one group uses machine learning, such as Decision Tree (DT), Support Vector Machines (SVM), Linear discriminant analysis (LDA), and so on, another group uses neural network, such as Deep Neural Network (DNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM). However, these methods are not antithetical to others, many researchers combine several methods together.

2.1.1 Supervised Learning Methods

Maurer et al. [64] introduced a HAR system named eWatch. This system was built in a sport watch with four sensors, including accelerometer, light sensor, thermometer, and microphone. The data processing and identification were executed in the system, thus, no wireless communication is needed. The decision tree C4.5, LDA transformation with a four second window time-domain, and feature extraction were included in the system. The best overall accuracy was up to 92% for totally six activities: descending, running, sitting, standing, ascending, and walking, while ascending and descending the stairs was confused with walking. The response time for feature extraction and identification is less than 1ms, which is very suitable for a small device with quick feedback function.

Lara et al. [58] introduced a mobile based real-time HAR system under Android platform. This system had three sensors devices, two in phone, namely the GPS and accelerometer, and one in a chest trap measuring the heart rate, respiration rate, skin temperature and so on. The Bluetooth, TCP, and HTTP servlets were required to receive, send, and transform the raw data. C4.5 decision tree was applied after the feature extraction, and the overall average from subject-independent (the subjects in test set was not in the training set) was up to 92.64% for three activities, running, walking, and sitting. The response time for each 5-second-long overlapped window totally with preprocessing, feature extraction, and classification could be reduced to 66.5ms, which made the system responsive. However, with the variety activities in our daily life, three activities in the system might not satisfy the daily activity requirement.

Attal et al. [11] presented a HAR for remote monitoring of elderly or dependent subjects with daily activities and detecting unpredictable events, such as standing, walking, stair decent, sitting down, sitting on the ground. The inertial sensors were placed in three places on chest, right thigh, and left ankle. Bluetooth was required to pass the raw data to PC. 6 subjects participated in their experiments. The overall accuracy with extracted feature from subject-dependent (the subjects in test set was the same in the training set) was up to 99% from k-nearest neighbor and 98.95% from Random Forest. The result showed that the eleven time-domain features, such as mean, variance, median, and range, were useful to identify for these daily activities. However, 6 healthy subjects with average of 26 years old, might not be sufficient to provide a reliable conclusion for elderly or dependent people. Besides, no response time was mentioned in the paper.

And Abidine et al. [2] also introduced a combined method with PCA, conditional random filed (CRF), LDA, HMM, and SVM to classify the activities. The significant components from PCA were added to the set of feature extraction from LDA, and then the modified weighted SVM (WSVM) was applied. Four similar datasets with four state-change sensors located in different environments were collected and evaluated.The data were TK26M, TK57M, Tap80F, and OrdonezA, with 8, 16, 10, and 10 activities, respectively. These daily activities included Leaving, Toileting, Showering, Sleeping, Shaving, Relax, and so on. The results regarded to the accuracy rates were from 22% to 95% and did not improve significantly. However, the idea of combining different dimensional reduction methods was useful, and the weighted SVM for unbalanced data was also a good try.

Cheng et al. [22] used the dataset from W.Ugulino's team, which was collected from 4 person with 5 activities, sitting, sitting down, standing up, standing, and walking, by wearing 4 sensors in different places. They applied SVM, HMM, and multi-layer neural network (MNN) to the data with both subject-dependent and subject-independent experiments. The results from the subject-dependent were above 90% overall accuracy, and with the HMM, the accuracy for activities were close to 100%. While the overall accuracy for one person from subject-independent were less than 50% and some activities could not be identified, such as sitting was confused with walking. It could indicate that the activity pattern is different from person to person, and personalized system should be built. However, it is not clear that which body location the sensor has to be placed that could have a better classification results.

Jiang et al. [53] compared the performance of Deep Convolutional Neural Network (CNN) from three different data sets, UCI, SHO, and USI, which all based on wearable sensors, first two were from smartphone and recorded 6 and 11 daily activities, the rest one was from MotionNode and had 7 activities. All these three data collection platforms were not involved in feature extraction and classification. With different number of activities, sensor locations, time windows, and number of subjects, the overall accuracy rates were between 97.59% to 99.93% with CNN+ (the CNN with bi-class SVM). They then converted the signal data into image data in order to use the 2D CNN algorithm. It also represented that with less than 3% of improvement, the computational time for CNN+ (3.85ms) was doubled than that for DCNN (1.56ms). Since the computational time was only for the learning part, the image preparation part was not mentioned, the total process time was uncertain. Also, without the information of training and test data split, it was not clear either subject-dependent or subject independent. Meanwhile, the device memory would be another issue, since image data might need more space to be stored. Because of these reasons, this method would be more suitable for an offline system, which would provide a large database for both storage and classification, instead of online.

Ronao and Cho [78] also used the CNN for identifying the activities. Instead of transferring the sensor data into window based features, they use the data directly. However, by applying CNN with different layers and units to the UCI data (6 activities) for subject-independent experiment, the result showed around 90% of overall accuracy rate. Later, they used the deep CNN to transfer the sensor raw data combined with the temporal fast Fourier transform method [77]. They achieved an overall accuracy rate of 95.8%. However, this accuracy rate was used with a large filter size, with more than 100 units and 4 layers. The computational time was longer than the previous one. Moreover, Liu et, al. [61] also used CNN for identifying human activities. The data they used was collected by binary PIR sensors (PIR), which might reduce the data size. They collected data in three groups and 3 activities in each group. The overall accuracy rate was up to 97.7%. However, there was no specific activity mentioned in the research, the applicability of this methods to the real-world was not certain.

Besides the research on the daily activity recognition, Kautz et al. [54] also used deep covolutional neural network to the HAR in the beach volleyball players' actions. The researchers collected the data by putting a tri-axis sensor on players' wrist of the dominant hand, and divided the actions into ten classes, Underhand Serve, Overhead Serve, Jump Serve, Underarm Set, Overhead Set, Shot Attack, Spike, Block, Dig, and Null Class. With 39 window-fixed transformed features, the overall accuracy rate for subject-independent was up to 83%, which was 16% better than others, such as KNN, SVM, NB, and decision tree. The required training computing time was also more than others. It needed more than 2.5hrs. But consider this as a built-in system, the classification with feature calculation and prediction was 0.06ms on average.

Li et al. [60] applied several deep learning algorithms, including Long-Short-Term-Memory (LSTM), CNN, Multi-Layer-Perceptron (MLP), and Hybrid Covolutional and Recurrent, to two datasets, namely OPPORTUNITY and UniMiB-SHAR. The former one required 7 wirless body-worn inertial sensors with 12 additional 3D accelerometers in different body locations. It collected 9 classes, including one Null class, which was 72.3% of the total data. This data was about opening or closing doors, fridges, and dishwasher. The later one was based on the smartphone 3D accelerometer, placed on either left or right pocket. The activities it collected were some daily activities, such as standing up from sitting, standing up from laying, walking, running, going up, jumping, going down, lying down from standing, and sitting down. The biggest challenge from these two datasets was the imbalance samples among classes. The best overall accuracy rate from subject-dependent for the first dataset was up to 91.1% with hybrid model and 77% for the second data from a soft assignment variant. The split of training and test was not clear, as well as the computational time. Besides, a large number of sensors were used in the first dataset, which was inapplicable in the daily life.

Wang [89] proposed a deep belief network (DBN) with continuous autoencoder (CAE) and fast stochastic gradient descent (FSGD) to identify 19 activities sitting, standing on the ground, lying on the back, etc. The data (Altun) was collected by Altun et al. [9] from 5 sensors with different location on the body, one each on the arms and legs, and one on the body torso. The research showed that the frequency-domain features had better results than the time-domain had and the Accelerometer with Magnetometer gave the best overall accuracy rate for a 10 fold cross validation. After the built-in the DBN in the HAR system, the overall accuracy rate for a single person was around 82% with 0.11 millisecond per 5-s segment. While with an overall accuracy rate of 99.3% in the training data, 82% for the test data, the model might be overfitting. However, sensor in five different locations might be uncomfortable and invasive.

He et al.[46] applied the generalized discriminant analysis (GDA) and relevance vector machines (RVM) with a public dataset named WARD. 20 subjects performed 13 activities wearing 5 sensors on wrists, waist, and ankles. The data included several transition activities, such as turn right, turn left, and walk left circles. It showed that with the GDA, the overall accuracy from the 3-fold cross validation was 99.2%, which was 23% higher that from PCA and 59% higher than from LDA. Meanwhile, for single activities, the RVM also had a good performance, with all above 97%, which is better than SVM and KNN. Since the process time was not given and with 5 sensors, it was not clear if this was suitable for a single wearable device.

Berchtold et al. [13] proposed an online HAR system named ActiServ. This system included data collection, data storage, data transmission, classifier sets, personal training, and feedback interface. The data had 10 classes, such as sitting, standing, lying, walking, climbing stairs, cycling, holding phone, talking on the phone, typing text, and no movement. It used fuzzy inference system and Activity Classification Module Set (ACMS) combined Personal Trainer Service (PTS) to identify several activities from the phone accelerometer. The overall accuracy rates were up to 99.3% even for the activities such as holding phone, talking on the phone, and typing text message. However, the delay on the feedback was quite long, the running time duration was up to hours for one activity for the personalization part. When with the built-in model, which could be considered as offline part, the overall accuracy rates dropped to 70%. Besides, the whole system was based on the server, which would cause some connection issues.

Wang et al. [88] compared traditional convolutional neural network (CNN) and DeepConvLSTM methods with their own method on two datasets. The methods included the network net with attention levels (Net-att) with dot product for matching the local and global features, operation of parameterized compatibility (pc), and different normalization functions (softmax/tanh). The two datasets were a well labeled UCI HAR dataset and a weakly labeled data collected by authors from smartphone. The weakly labeled data only obtained the information about the activity occurred but without the time. UCI HAR dataset included 6 daily activities, 3 static activities and 3 dynamic activities and it was divided in 70% for training data and 30% for test data. The weakly labeled data only had 5 dynamic activities and 70% was for training, 10% for validation, and 20% for test data. The results showed that proposed method was comparable to the CNN and DeepConvLSTM for UCI HAR data (93.41% compared to 93.21% and 93.54%), and even better than those two for the weakly labeled data (93.83% compared to 89.62% and 90.04%). Meanwhile, the proposed method asked less computation time.

Table 1 gives a brief summary of research, includes the data names, the number of activities in the data, the methods used, the training and test divisions (subject-independe or subject dependent), and the best overall accuracy rates from the methods. As it shows that the deep learning methods, such LSTM, CNN, the combination of two deep learning methods, were applied to these sensor data frequently and other machine learning methods, such as SVM, DT, and KNN, are also used. It is clear that the deep learning methods also could not grantee a high accuracy rate for different datasets and KNN and RF might give a high degree of performance. In other words, the data structure is important for selecting method. Supervised methods use the labeled data to build a model, which could be used for prediction or classification, but it required a lot of labels, which might need lot of work and might have the error when mark the labels. Thus, for big dataset, unsupervised method or auto-label creating might need to be considered.

Table 1 Datasets and Results Summary for HAR Supervised Classification Methods

Data Name	# of Activities	Research	Methods	Subject-independent	Best Overall Acc.(%)
ActiServ	10	Berchtold et al.[13]	ACMS-PTS	Yes	70
Altun	19	Wang [89]	DBN, KNN	Yes	82
eWatch	6	Maurer et al. [64]	DT, LDA	Yes	92
GPS-HAR	3	Lara et al. [58]	DT	Yes	92.6
OPPORTUNITY	9	Li et al. [60]	LSTM, CNN, MLP, Hybrid CNN and RNN	No	91.1
PIR	9	Liu, et, al. [61]	CNN	Not Mentioned	97.7
SHO, USI	11,7	Jiang et al. [53]	CNN, CNN+, SVM	No	97.8, 99.9
TK26M, TK57M, Tap80F, OrdonezA	8,16, 10,10	Abidine et al.[2]	PCA, CRF, LDA, HMM, WSVM	No	95.6,81.4 77.2, 88.4
UCI	6	Jiang et al. [53]	CNN, CNN+, SVM	No	97.6
		Ronao and Cho [78] [77]	CNN	No	90, 95.8
		Wang et al. [88]	CNN, DeepConvLSTM, NET-att2-pc-tanh	No	93.41
UniMiB-SHAR	9	Li et al. [60]	LSTM, CNN, MLP, Hybrid CNN and RNN	No	77
Volleyball	10	Kautz et al. [54]	CNN	Yes	83
W.Ugulino	5	Attal et al. [11]	KNN, RF	No	99
		Cheng et al.[22]	SVM, HMM, MNN	Yes	< 50
WARD	13	He et al.[46]	GDA, RVM	Yes	99.2

F1 represents the best F1 score from methods in the article.

2.1.2 Supervised Learning Feature Methods

Besides generating the appropriate classification algorithms from machine learning and deep neural network to improve the accuracy rate of identifying activities, the features used in the model also play an important role. Thus there are some research that focus on feature selection, such as feature transferring and subset of features selection. Plötz et al [73] discussed several feature selection analysis, and dimensional reduction methods in their research, such as PCA, LDA, empirical cumulative distribution function (ECDF), and independent component analysis (ICA), as well as the Restricted Boltzmann Machine (RBM), which is one of the deep learning methods for autoencoder based feature learning. They compared these methods with four datasets, Ambient Kitchen (AK), Darmstadt Daily Routines (DA), Opp, and Skoda. These four datasets were performed under different scenarios with different activities. AK was collected from twenty participants who prepared either a sandwich or a salad with sensor-equipped kitchen utensils. DA was collected in a living lab-like environment. Two tri-axial accelerometers were placed on the wrist and pocket to record 35 daily activities. Opp was collected in kitchen with 19 worn sensors, which contained open then close the fridge, clean the table, etc. And Skoda contained 46 activities from workers wearing 20 accelerometers on both arms in a car maintenance scenario. With the same classifier, the best accuracy rates were from FFT and the combination of PCA and ECDF for data AK and DA, the combination of PCA and ECDF and the combination of RBM and ECDF for data Opp ,and RBM and the combination of PCA and ECDF for data Skoda. Different datasets had different best feature analysis method, but the combination ones with ECDF might be a better choice for all the four datasets. The best overall accuracies for these four datasets were 88.7%, 89.1%, 81.5%, and 75.9%, respectively.

Vollmer et al. [87] also used the same datasets to evaluate their feature transformation method. They formalized the sensor feature learning as a sparse coding problem. They tried to minimize an energy function on the error between reconstruction and input with a penalty on the activation. Later, by combining a variant of Non-negative Matrix Factorization (NMF), they transferred the sensor data to be a non-negative matrix, named Shift-invariant Sparse Coding (SISC). The contribution of this was they gave a new idea to transfer the sensor data with sparse matrix into the statistical features, such as mean, maximum, minimum, standard deviation, and so on. The overall classification accuracy rates from SISC with KNN for AK, AD, Skoda, and Opp were 91.7%, 86.1%, 84.5%, and 81.6%, respectively. As we could see that the result from AD data was worse than it from Plötz et al [73], while others were better. Then Bulling et al [18] also mentioned several feature selection and dimensional reduction methods in their survey paper. They included some methods which could reduce the computational time and the memory usage, such as minimal redundancy maximal relevance, correlation based features selection method, and ECDF were for dimensional reduction methods. Different features section or extraction method fit for different datasets.

Zhang and Sawchuk [101] purposed a sparse represented based classification method for identifying 9 human activities from MotionNode, such as walk forward, walk right, up stairs, run forward, and sit on a chair. Instead of using PCA and LDA as the linear transformation, they applied the random projection (RP). The data was collected from 14 person wearing a MotionNode with accelerometer, gyroscope, and magnetometer on their right front hip. The overall accuracy rate for left-out subject was up to 95.2% which was more than 1.3% higher from the best of SVM, KNN, and NB. Moreover, the results also showed that the sparse represented based method worked better with more than 40 feature dimension data. It was also suitable for wearable devices which does not require large memory space. However, some activities were confused, such as Sit on the chair from Standing, Jump up from Run forward. These activities were also hard to be identified by SVM, KNN and NB since the similarity of the raw sensor data, such as frequency in each axis.

Wang et al. [91] purposed a feature pre-process method with game theory based feature selection and Ensemble empirical mode decomposition (EEMD). The Data was collected from 5 subjects by wearing triaxial accelerometer on waist and ankle in their laboratory. These subjects preformed 9 activities, including sitting, lying, standing up from lying, standing, walking, running, two different types of watt bicycling, jumping. The results from KNN and SVM all shown that the overall accuracy were better with the EEMD than that without EEMD. However, some of the activities, such as walking, standing up from lying, and bicycling, were confused with other activities. Thus, the overall accuracy rates for single sensor location were in 73% - 81%. However, by increasing the number of subjects and combining these two sensors together, with a subject independent training-test model, the results might be improved.

Zdravevski et al. [97] also focused on the features extraction process. They proposed a 14 steps process to generate sliding windows based features with magnitude time series, delta series, first derivatives series, and Fast Fourier Transformation (FFT). The final feature set included all of the best selected subset features from these groups, which could be considered as an ensemble of features (EF). Then they applied this process to four different datasets, DaliAc, mHealth, FSP, SBHAR, and SBHARPT. These datasets were from different sensors or locations. The DaliAc dataset was collected from 19 subjects performing 13 activities with 4 sensors. It recorded several normal daily activities, such as walking, standing, lying, and so on. mHealth was a dataset from an open framework for agile development of mobile health application. It included 12 activities and some of them were aimed to be special purpose, such as waist bends forward, frontal elevation of arms, knees bending, and jump front and black. FSP was a dataset with 7 daily activities from 5 wearable sensors. SBHAR was also a dataset had 6 daily activities from smartphone sensors, such as accelerometer and gyroscope. The labels were marked by manually from a simultaneously recorded video. SBHARPT was from the same authors of the SBHAR, but included 6 extra transition activities, such as stand to sit, sit to stand, lie to sit, and so on. They then implemented the new features sets with the logistic regression, RF, (extremely randomized trees) ERT, SVM, KNN, and NB. All of these datasets were divided into three groups, training, validation, and test. Each group included different subjects. The subject independent overall accuracy rates were between 91% to 99.8% from SVM. However, sitting, walking downstairs, and walking upstairs were still confused here, the precision rates for these activities were around 70%. Moreover, they did not mention the computing time for transferring the original timestamp series data to the features used in the model. This made it hard to evaluate the suitable for a wearable HAR system.

All of the research mentioned above showed that the dimensional reduction might be depended on the data information collected from initial sensor. For example, in Plötz et al. [73] and Vollmer et al. [87], they applied several dimensional reduction methods with KNN to different datasets, the best feature selection was arbitrary and it was a big challenge to generalize. In other words, there is no universal best method for feature analysis. Because of this, an automatic extraction of feature process would be a great help for HAR classification. The deep learning provide methods for learning important features from the raw sensor data. Zeng et al. [98] used three datasets as Plötz et al [73] in their research, Opp, Skoda, and a new dataset Actitracker, to illustrate that CNN based features with partial weight sharing approach (CNN-PWS) was practical on wearable sensor data. Actitracker contained six daily activities, such as walking, ascending stairs, jogging, etc. It were collected in a controlled laboratory environment using smartphone. All of the three datasets had a better identification accuracy rate (KNN) with the features selected from CNN and partial weight sharing than from Statistical RBM, and PCA-ECDF ($1.2\%, 9.02\%, 4.41\%$, higher than the best algorithm on Opp, Antitracker, Skoda respectively). Comparing with the results from Plötz et al. [73], this method improved the overall accuracy more than 10% for Skoda data (75.9%, 88.9%), but performed worse in Opp data (81.5%, 76.83%).

Later, Ordóñez and Roggen [72] proposed a continuous feature selection method to the same mobile and wearable sensor data (Opp and Skoda). They used a combination of convolutions neural network and LSTM for identifying some static and periodic activities, such as open or closed doors, fridge, or drawers, standing, walking, sitting, and writing on notepad. The overall F1 scores were improved by 6% on average than other machine learning methods, such as LDA, QDA, and CNN for both Opp and Skoda datasets. Thus the overall F1 scores were 91.5% and 95.8% from DeepConvLSTM for Opp and Skoda datasets, respectively. Additionally, they considered the multimodal fusion analysis, which combined several different sensors, such as different types of measurement sensors with different locations. The F1 scores were improved up to 16% from only one accelerometer to the Opp set, which had 19 sensors. However, this was not ubiquitous in daily life. Besides, the computational time was not mentioned. They trained the model on all data from subject 1 and partial data from Subject 2 and 3, and leaved the rest of Subject 2 and 3 for validation and test.

Alsheikh et al. [7] used a new transform way for a combined deep leaning and HMM model (DL-HMM) to identify activities. Instead of using the raw time-series sensor segmentation data or statistical transformed data, they used the spectrogram data. There were three datasets they used to evaluate their model: WISDM, Daphnet freezing of gait (Daphnet), and Skoda. WISDM inclded 6 daily activities, Daphnet was labeled as either "freezing" or "no freezing". This data was designed to detect freezing events of patients with Parkinson's disease. And Skoda had 10 activities. The deep generative model was used for computational intrinsic features and by finding the posterior probability distribution to recognize the activities. Then the HMM model trained by the emission probabilities was considered as a temporal patterns model in activities. The performance showed that this DL-HMM combined model had better outcomes. They improved the accuracy rate from 85.1% (C4.5) to 98.23% with the WISDM data, from 86% (HMMs) to 89.63% with Skoda data, and improved the Sensitivity (TPR) from 82% (C4.5 and KNN) to 91.5% with the Daphnet data. Ravi et al. [76] also used spectrongram features, which they called shallow features, with deep features together in a back-ward propagation deep neural network (CNN). They applied this to five datasets, ActiveMiles, which included 7 daily activities collected by smartphone in uncontrolled environment, WISDM v1.1 and v2.0, which were collected from smartphone in a laboratory and uncontrolled environments respectively, Daphnet, and Skoda. The results showed that the combination features in ActiveMiles, WISDM v1.1 and v2.0 could improve the overall accuracy rates (95.7%, 98.6%, and 92.7%, respectively) up to 1.2% comparing to only use either shallow features or deep features (95%, 97.4%, and 92.5%, respectively). And the Shallow-features only could give a better results for the Skoda and Daphnet (95.9% and 95.8%, respectively) than the combination (95.3% and 95.8%, respectively). However, some activities were still have confusions, such as the standing from the ActiveMiles data with precision rate of 74.19%, lying down from the WISDM v2.0 with precision rate of 88.65%, and freeze from Daphnet with precision rate of 67.89%.

However, there are also some research showing that the deep learning itself could perform better with the original signal data than the data with transferred features or selected features. Chen and Xue [21] used CNN to identify 8 activities from an Android smartphone application data. The results illustrated that without feature selecting, CNN outperformed the DBN and SVM with transformed features, such as FFT, Discrete Cosine Transform (DCT), and Time-domain Features (TF), not only better with the overall accuracy, but also with each activity. The overall accuracy rate was upto 93.8%, and for each activities, the accuracy rates were in the range from $88\% - 97\%$. Erfani et al. [30] showed the proof that the deep belief networks (DBN) could learn robust features and help to reduce the computing time by a factor of 3 and 1000, for training and testing time, respectively. They used the 6 real-life HAR datasets from UCI Machine Learning Repository (UCI), Forest Adult Gas Sensor Array (Gas), Opp, Dailyand Sport Activity (DSA), Human Activity Recognition (HAR), Banana, and Smiley. The DBN with one class out SVM method had a decent performance on human activity recognition data from UCI website. The test time was reduced from longest 0.88 seconds to 0.0099 milliseconds without losing of accuracy. Yao et al. [94] also used three datasets to illustrate that arbitrary length based CNN works better on the raw sensor sequence data than the ones based on sliding window data. The three datasets were the Opp, Hand Gesture, and a new dataset from a wearable sensor worn by hospitalised patients (Hospital). The overall accuracy would be improved up to 4%. The overall accuracy rates were 88%, 89%, and 79% for Opp, Hand Gesture, and new data, respectively. Moreover, utilization of sequence data would help to avoid the multi-class in one window problem and the difficulty of finding the best window size. This made the HAR system be able to correctly capture the fast transmitted activities.

Besides these classification and feature transformation research, there are also some people focused on the multisensor area. This is the trend for the human activity recognition system because of the technologies. The system could put different sensors together, for example record the environment conditions, such as humidity and temperature, as well as customer's health conditions, such as heart rate. Yao et al. [95] first proposed to use DeepSense architecture to model temporal relationships among the sequence data and fuse multimodal sensor inputs, such as motion sensors and movement sensors (CarTrack). The DeepSense involved both convolutional neural network and the recurrent neural networks. Both of these parts had the ability to automatically learn features. The data HHAR was collected from 9 users wore smartwatch and smartphone who executed 6 types of activities (biking, sitting, standing, walking, climbstair-up, and climbStair-down). The overall accuracy was up to 93.8% and much better than the results from single GRU or CNN. the training process was run on GPU, and the trained model was built on mobile with CPU. However, there was no computational time mentioned in the paper and this system needed a stable connection between the devices and the CPU. As Bulling [18] mentioned in their survey, the automatic feature representation from deep learning helped to capture both intra-class ad inter-class variables. Table 2 and Table 3 give the summary of the datasets and feature methods.

Table 2 Datasets and Results Summary for HAR Supervised Feature Selection Methods 1

Data Name	# of Activities	Research	Methods	Subject-independent	Best Overall Acc.(%)
ActiveMiles	7	Ravi et al. [76]	shallow features (CNN)	Not mentioned	95.7
Actitracker	6	Zeng et al. [98]	CNN-PWS	No	96.88
AD	35	Plötz et al [73]	PCA, LDA, ECDF, ICA	No	89.1
		Vollmer et al.[87]	SISC	No	86.1
AK,	2	Plötz et al [73]	PCA, LDA, ECDF, ICA	No	88.7
		Vollmer et al.[87]	SISC	No	91.7
Android App	8	Chen and Xue [21]	CNN	No	93.8
DaliAc, mHealth, FSP, SBHAR, SBHARPT	13, 12, 5, 6, 12	Zdravevski et al. [97]	EF	Yes	93.4, 99.9, 99.8, 95.9, 95.8
Daphnet	2	Alsheikh et al.[7]	DL-HMM	Not mentioned	91.5 (TPR)
		Ravi et al. [76]	shallow features (CNN)	Not mentioned	95.8
HHAR	6	Yao et al.[95]	DeepSense	Yes	93.8
IMU-based	9	Wang et al. [91]	Game theory based EEMD	Yes	81
MotionNode	9	Zhang and Sawchuk[101]	RP	Yes	95.2

F1 represents the best F1 score from methods in the article.

Table 3 Datasets and Results Summary for HAR Supervised Feature Selection Methods 2

Data Name	# of Activities	Research	Methods	Subject-independent	Best Overall Acc.(%)
Opp	19	Plötz et al [73]	PCA, LDA, ECDF, ICA	No	81.5
		Vollmer et al.[87]	SISC	No	81.6
		Zeng et al. [98]	CNN-PWS	No	76.8
		Ordóñez and Roggen [72]	DeepConvLSTM	No	93 (F1)
		Erfani et al. [30]	DBN		
		Yao et al. [94]	CNN	No	88
Skoda	46	Plötz et al [73]	PCA, LDA, ECDF, ICA	No	75.9
		Vollmer et al.[87]	SISC	No	84.5
		Zeng et al. [98]	CNN-PWS	No	88.2
		Ordóñez and Roggen [72]	DeepConvLSTM	No	95.8 (F1)
		Alsheikh et al.[7]	Not mentioned	89.6	
		Ravi et al. [76]	shallow features (CNN)	Not mentioned	95.9
WISDM	6	Alsheikh et al.[7]	DL-HMM	Not mentioned	89.6
		Ravi et al. [76]	shallow features (CNN)	Not mentioned	98.6 (92.7 for v2.0)

F1 represents the best F1 score from methods in the article.

2.2 Semi-supervised or Unsupervised Learning

The semi-supervised system is suitable on the system which only has a few activity labels. This makes the learning process more difficult to learn pattern or activity characteristics from the training data. However, labeling all cases might be hard or expensive in real world. For instance, for the data collection, the participation need to follow a strict illustration to avoid for mislabeling with the transition activities. Sometime, researchers also have to have the video recorder to minimize the mislabeled cases. This makes the semi-supervised or even unsupervised learning methods more valuable.

The early semi-supervised methods applied on HAR was the ensemble methods from Guan et al. [41]. It modified one of the semi-supervised method called Co-training to avoid strict requirement of the labeled dataset. The new method was called En-co-training, which was the combination of the semi-supervised with the ensemble method by majority voting. Instead of using two partial labeled data for training two classifiers in Co-training, they applied all the labeled data to three different machine learning algorithms, such as DT, NB, and DNN. The data (Leg40) they used was from an experiment in 2004 [86]. The participants wore 40 sensors on their legs, (20 each) during the performance of 10 basic activities, lying, kneeling, sitting, standing, walking, running, climbing stairs, descending stairs, bicycling, and jumping. Thus, instead of training leg data (right and left) separately with Co-training, they used all the 40-sensor data for three different algorithms (DT, KNN, and NB). The experiments showed that the overall average rates for 10 test subjects from En-Co-training (between 80% to 86%) were comparable to these from Co-training (around 83%), when training on 10% of the labeled data. While the increasing the labeled data in the learning process, the En-Co-training method performed better than others for each subjects data. The overall accuracy rates were around 90% while others are around 85%. However, both Co-training and En-Co-training did not mention if the classifiers were updated with the results from the unlabeled data. Also, 40 sensors was not a good choice for wearable devices.

Cardoso et al. [19] also used the ensemble methods in their research. They selected three learning methods, NB, Very Fast Decision Tree (VFDT), and KNN, and created two ensemble classifiers, one is Democratic Ensemble Classifier, also called Most Voted Classifier, another one is Confidence Ensemble Classifiers, which was used the certainty percentage of the prediction to decide the agreed output. They also represented a comparison with two different ways of ensemble supervised and semi-supervised methods. Leave-one-out of the person data was used for the semi-supervised update and test. They used the ensemble results to re-train the model and applied the new model to the test. The Smartlab data was used,which was including 6 basic daily activities and 6 transitions between the static postures. Their results implied that with the fixed window size sensor data, the semi-supervised learning was better than the supervised learning, even with small amount of extra unlabeled data updating. Besides, with the perfect segmentation of the window size, which would give the best prediction later, it was able to identify all the activities with almost 100% certain. Meanwhile, they also proved that classifiers were confused by the different placed sensor data. However, this experiment gave the best average accuracy rate was 84.55% from Confidence Ensemble classifier. The system required 15 minutes of data inputting for training. It might be considered as a long time practice for a customer.

Garcia et al. [38] used methods mentioned in above [19] to present the hyperparameter's effects in their experiment. They applied semi-supervised ensemble method to a public dataset, PAMAP2, which including 18 different activities. They also found that the accuracy of this ensemble method was sensitive to the different user's data than to the window size and overlapping percentage. The overall accuracy rates were between 75% to 90%. Additionally, they found that feature extraction process was time and energy consuming. The total time, including the time required to access the data and transfer each data window as an instance, the time for feature extraction, and the time with the classification, was up to 50s for a single user. This might cause some delay for an online system and not be efficiently working for long periods of time. Thus, an extension work of this was proposed [39]. Instead of using ensemble of learning methods, they used Auto-Encoders (AE) to reconstruction every activities individually and labeled the data with the minim reconstruction error from the AE predictions. And each AE was updated with the new data with the minim error less than a threshold. The method was applied to three datasets, WISDM, mHealth, and PAMAP2. The average accuracy rates were up to 82% for both WISDM and MHealth data, but only got 62% for PAMAP2, which was worse than that from [38]. This ensemble method was not as good as expected neither in time and the classification, especially in the average accuracy. But the experiment proposed a new idea for the possibility to identify new activities which were not included in the training, with the minim error greater than the threshold.

Later, some research also used deep learning, such as CNN in the HAR identification. Zeng et al. [99] combined the supervised CNN and CNN-Encoder-Decoder (unsupervised learning) to create a semi-supervised CNN-encoder-decoder, and with lateral connections to create a semi-supervised Convolutional Ladder Network (CNN-Ladder) semi-supervised CNN-Ladder. By getting the parameters from both supervised and unsupervised CNN, they applied the semi-supervised CNN to three datasets, ActiTracker, PAMAP2, and mHealth. ActiTracker and mHealth were the same datasets introduced above. PAMAP2 was a new dataset used here. It was collected from 9 participants with 12 lifestyle activities, such as walking, knees bending, lying, etc. The most unique characteristic in PAMAP2 was the records with the heart rate and the temperature data. The results showed that the CNN-Encoder-Decoder and CNN-Ladder performed better than the results from supervised logistic Regression and CNN, especially, CNN-Ladder improved the mean F1-scores (F_m) with by 17.64%, 3.59%, and 9.65% on ActiTracker, PAMAP2, and mHealth, respectively. And they were also better than other semi-supervised methods, such as self-training and pseudo-label. The improvements were 16.46%, 4.11%, and 8.5% for three datasets, respectively. However, the method was based on empirical labeled and unlabeled data, the rule of dividing them were not clear. The computational time was not shown in the article.

Balabka [12] also purposed a combined methods with CNN and auto-encoders to predict the activities in the test data which not included in the training set but only in the validation set. The Sussex-Huawei Locomotion-Transportation (SHL) dataset was collected from the participants who wore the mobile devices at different places, including Bag, Hips, Torso, and Hand. They performed 8 different activities. The special part was the Hand data was only included in the validation and test set, meaning it did not been trained in the model. And the validation and test set were divided into labeled and unlabeled data. The results showed that the F1 scores for 8 activities were in the range of 75% to 97%, the activity 3 had 97%, while activity 8 had 75%. The overall average F1 score on the validation test set was 90.2%. However, the algorithm was hard to apply to specific activities since there was no information for the types of activities. In the research, it was not clear that how to use the unlabeled data for the new model. Beside, the total computational time for the validation training set was more than 6 days, it was a very long time for getting the model.

Also, some researchers tried non-parametric methods to HAR, such as Ma et al. [63]. They thought that for most semi-supervised methods neither sensitive to the outliers and hyper-parameters, or easily be misleading by mislabeled samples and assumptions. So they introduced a graph-based non-parametric method, named LabelForest (LF) to avoid the weakness. It involved greedy spanning forest, Silhouette-based sample filtering, and a SVM classifier. In the research, it showed that the LF algorithm improved the accuracy rates up to 57% and 175% on the balanced and unbalanced data compared to other common algorithms, such as k-NN, ϵ-NB, DT, regression (LR), SVM, and sequential k-means. These results were from three sensor datasets. The datasets included HART, Smartsock, and Phone. The first one included 6 daily activities from 30 subjects, the second one was from 10 subjects with 12 activities, and the last one contained 6 activities from 9 subjects wore a variety of smartphones. It also showed that the quality of the labeling results was more important than the quantity. With the lower rate of the labeling rate, LF had the most highest precision rates, which was up to 86%. Moreover, LF had a steady improvement with the increasing number of seed data samples, while others were fluctuated. Also, LF had a good performance with small amount of seed training data for all three datasets, 97.1%, 96.7%, and 86.3% for HART, SmartSock, and Phone, respectively, while others required more training data to achieve the similar accuracy rate. The semi-supervised method could result a good perform without lots of labeled observations.

Oh et al. [70] proposed a combination method for semi-supervised UCI data set and mHealth data. They used the active learning (AL) and transfer learning (TL) separately instead of using active transfer learning (ATL) together. And they set a small part of the sample as labeled while leave others unlabeled. They first applied DNN to the labeled training data as a basic model. Secondly, with the transfer learning to create a correctly predict set, which had a correct high probability and was used for semi-supervised learning, and false predict set, which had a incorrect high probability and was used for retraining the basic model. The result from UCI showed that by using 198 queries, the accuracy rate was up to 95.5%, which was better than the rate from the DNN with random sampling (92.9%, with 1000 queries) and a little worse than that from the active transfer learning (95.8%). Also, the results from mHealth data showed that the proposed method had the best accuracy rate (95.9%) with 693 queries, which the accuracy rate from ATL was 94.9%. This proposed method reduced the size of labeled data (only around 24% of the total data was labeled). Since it required retraining the basic model with the unlabeled data, the memory of the device might be an issue. Besides, the whole process might also require a long computational time with the labeled data training and unlabeled data traning.

Bettini et al. [14] proposed a communicated HAR system between customer and the device, such as smartphone and smartwatch. The system was a combination of active learning (AL) and transfer learning (TL) with semi-supervised Federated Learning (FedHAR). They assumed that the system was builtin with a classifier which was pre-trained with the global data, and provided the options with the results from the classifier. With the labels provided by the customer, the system delivered these labeled data to a server, added the them into the sample data, and had a new classifier, (personalized classifier). They then applied this process to the dataset MobiAct and WISDM. The MobiAct was a labeled dataset from 60 subjects who holding a smartphone in the pocket. The dataset included 4 activities: standing, walking, jogging, and sitting. The results showed that FedHAR with active learning and label propagation improved the F1 score significantly (from 3% to 20%). They divided datasets into 3 parts, 15% of data contained data from users that only for the pre-training, 65% contained the data for federated learning, and the rest 20% for test only. And the average F1 scores for two WISDM and MobiAct were up to 94% and 85% after 30 rounds of updating, respectively. The advantages of this system was the communication process and low requirement of memory. The communication process was a good idea for collecting labeled data and the requirement of memory for a device was not a big challenge since the classifier parameters were the only information needed to keep in the system. However, during the personality transfer process, the data had to be delivered to a server to have the new updated classifier, this might cause delay and might also have issues when customer switched several activities in a short time.

Table 4 gives the summary of the datasets and feature methods. Unlike the supervised learning, semi-supervised learning provides more flexibility to modify the HAR system, for example adding the interaction between the system and customers in Bettini et al. [14]. They designed a step to confirm if the activity prediction was correct. This not only helped the system to collect new data but also update the model. However, a method which could implement new activities to the model and adjust the model to personal preferred will be a new idea. Based on this, we proposed a new streaming system which called as adjusted HAR. This system could update the built-in model with the new input. It was not only for improving the classification results but also making the system adaptive to different customers. Moreover, another advantage of this adjusted system is that it could adjust the personal HAR model with time goes. It is designed for elder people who might change the activity pattern after unexpected accidents.

Table 4 Datasets and Results Summary for HAR Semi-Supervised Methods

Data Name	# of Activities	Research	Methods	Subject-independent	Best Overall Acc.(%)
Actitracker	6	Zeng et al.[99]	CNN-encoder-decoder, CNN-Ladder	No	66.32 (Fm)
HART	6	Ma et al. [63]	LF	No	97.1
Leg40	10	Guan et al.[41]	En-co-training	Yes	90
mHealth	12	Oh et al. [70]	AL, TL, ATL	No	95.9
		Zeng et al.[99]	CNN-encoder-decoder, CNN-Ladder	No	54.9 (Fm)
		Garcia et al.[39]	AE	Yes	82
MobiAct	4	Bettini et al. [14]	FedHAR	Yes	85 (F1)
PAMAP2	12	Zeng et al.[99]	CNN-encoder-decoder, CNN-Ladder	No	69.38 (Fm)
		Garcia et al. [38]	Ensemble	Yes	75-90
		Garcia et al.[39]	AE	Yes	62
Phone	6	Ma et al. [63]	LF	No	86.3
SHL	8	Balabka [12]	CNN-auto-encoders	No	90.2 (F1)
Smartsock	12	Ma et al. [63]	LF	No	96.7
Smartlab	12	Cardoso et al. [19]	Ensemble	Yes	84.5
UCI	6	Oh et al. [70]	AL, TL, ATL	No	95.5
WISDM	6	Bettini et al. [14]	FedHAR	Yes	94 (F1)
		Garcia et al.[39]	AE	Yes	82

Fm represesnts the mean of F1 scores and F1 is the best F1 score from methods in the article.

Chapter 3

Online Human Activity Recognition

Note to Reader:

3.1 Abstract

Wearable sensors in smart-phone and wrist tracking devices are widely used in the activity tracking and body monitoring with a low cost. Human activity recognition (HAR) is one of the important applications. Activities identification then is the core part of HAR. In this chapter, we present a comparison with several popular online machine learning methods using smartphone data and try to find the most effective model for analyzing these sensor data. The methods include Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA), Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN). Two datasets with different transition methods from smartphone are used. The data includes Walking, Walking Upstairs, Walking Downstairs, Sitting, Standing, Lying, and Jogging. The first dataset has the first 6 activities and shows that the SVM with linear kernel and LDA have the highest test average accuracy of 96.4% and 96.2%, respectively. Decision Tree performs worse than others with test average accuracy of 86.0%. While the second dataset excludes the Lying, but has jogging, and shows that LDA and DT are the most appropriate algorithms, the test average accuracy of these are close to 100%. KNN is the worse one with 74.3% test average accuracy. Based on all the results, LDA might be the best one for these sensors data. Moreover, the transition method used to reduce noise and extra information in the second data might be better than that in the first one. It has lower dimensions and better classification performance. In order to get these improved accuracy rates, in this chapter, we used grid search, multi-fold cross validation, and dimensional reduction method. In addition to just doing the comparison, we also pro-

posed a two-layer method for activity identification. This method is more flexible of choosing classifiers for activities and we expect to have better results with the combination of methods.

3.2 Introduction

Human activity recognition (HAR) with wearable sensor is an area that focuses on automatically identifying human activities based on transmitted sensor data. Since wearable devices present a convenient and noninvasive way to record physiological data from users with reduced manual intervention and a low cost, HAR has been successfully applied to several areas. For example, it is frequently used for health monitoring, sport training, and recreational activities recording. Based on the sensors in the smartphone, such as accelerometer and gyroscope, movements are identified or grouped into the sequences of activities with a fixed time window transition. For example, walking, standing, walking stairs, sitting and lying. Many machine learning algorithms have been used over years, including supervised or semi-supervised ways.

The two main classification methodologies are applied to the data. The first one is parametric, such as the multivariate linear regression, Bayesian methods, etc. Some additional conditions are usually needed to meet the assumptions, such as normality and non-collinearity. These assumptions are often violated with the sensor data, as Figure 1 shows the shape from the sensor coordinates, they are not following the normality assumption. This means that LDA will lose power and robustness when apply to dataset collected from other volunteers. The second one is non-parametric, which is distributions free and has less assumptions, such as Support Vector Machine, K-Nearest Neighbor, Neural Network and non-parametric multiplicative regression. These methods are more suitable for dealing with the data that we do not know the relationship between the response and the independent variables or the shape among variables. Considering the large size of the sensor data, the high dimension and the uncertainty of the relationship between variables, we tend to use non-parametric methods to create models and evaluate the prediction on the test data.

3.2.1 Data Introduction

The data used in this report are downloaded from UCI Machine Learning website [1] and the WISDM dataset [56] which is available in public domain. The UCI data experiments were carried out with 30 volunteers with ages 19 to 48 years old. During the experiment, the volunteers were wearing a Samsung Galaxy S II on the waist. The experiment collected the data in 3-axial linear acceleration and angular velocity at a constant rate of 50HZ. They performed six basic activities, walking, walking-upstairs, walking-downstairs, sitting, standing, and lying. The first three are called dynamic activities and the rest are called static activities.

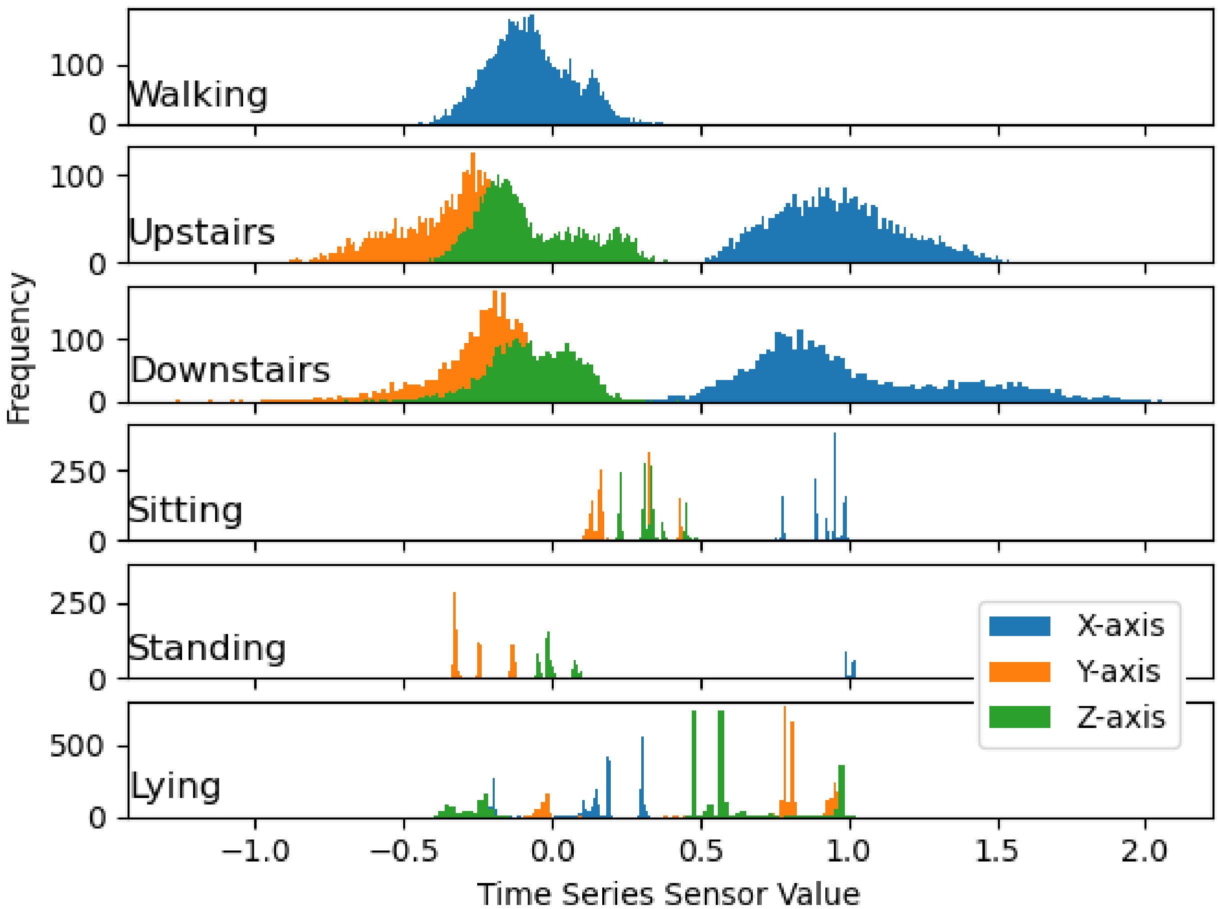


Figure 1. Histogram for Activities

Before we use this dataset, we plot the raw sensor data use the box-plot in Figure 2 and run the Runs randomness test. The plot shows that there are lots of outliers from the sensor gyroscope (body gyro 0-2). It implys that the transformation will be need to reduce the noise. The Runs test The runs test is a randomness test based on binomial distribution, which can be used to decide if a data set is from a random process [48]. For each variable, we run this runs test. The test results show that all the p-values from the runs test are less than 0.0001, meaning we reject the null hypothesis, which is these data are generated in a random manner. Thus, this UCI data are not produced in the random manner. the reason for this might be because these data were collected from 30 volunteers and each of them performing for a sequence of activities, and this special way of data collection would generate lots of duplicate data points. In other words, it could be also indicate that personalized analysis might be needed.

From the data, we also have the sequences of the coordinate's streams from initial sensors, as shown in Figure 3 from one of the subjects. The bottom row in Figure 3 shows the activities, 1,2, and 3 represent the

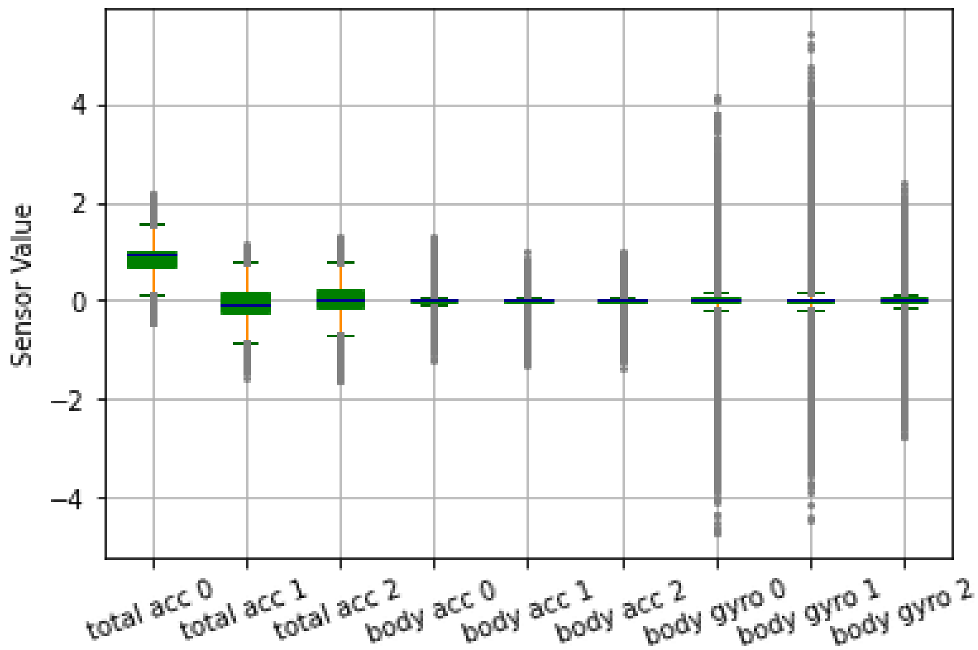


Figure 2. Box Plot for Raw Sensor Data

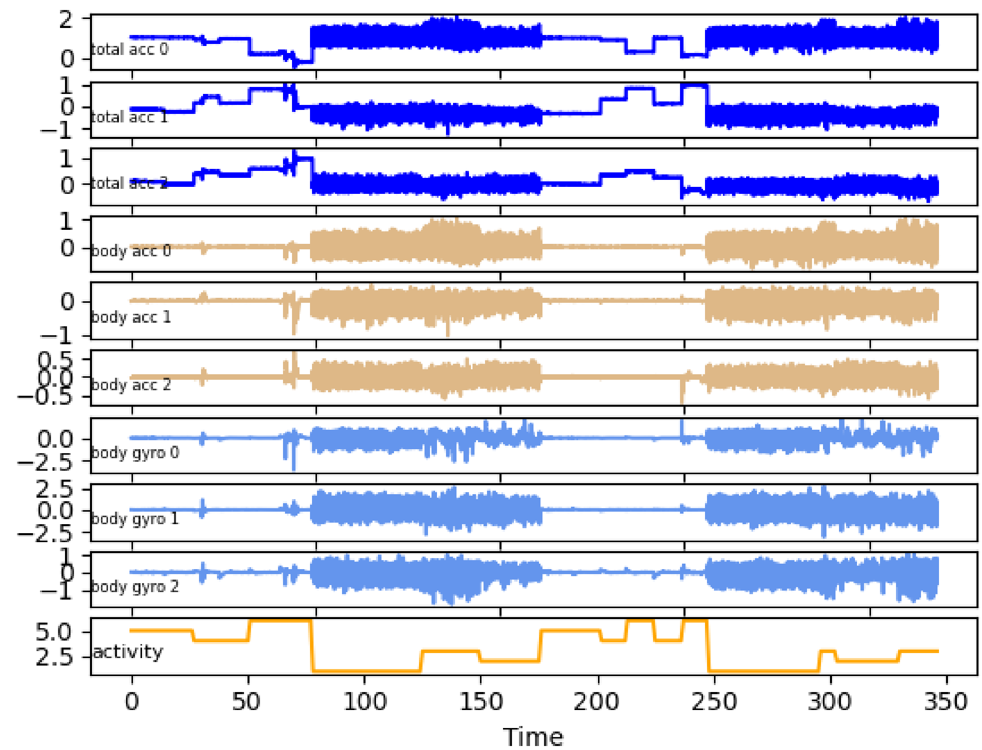


Figure 3. Coordinates Values of Activities

dynamic activities walking, upstairs, and downstairs, and 4,5, and 6 represent the static ones, sitting, standing, and lying. It is easy to notice that there are big differences between dynamic and static activities. Sensor coordinates values vary with dynamic activities while static ones have more flat lines. By visualization, it is hard to identify the activities from the initial sensor data.

The data was modified by applying several filters and same specific hertz rates to remove the noise. After this, a Fast Fourier Transform (FFT) was applied to these signal data. All of these transformations are based on a fixed window size with 50% of overlap. Finally, there are totally a vector of 561 features in each record, including 17 groups, each group has 17 statistical measurements, and 8 out of 17 groups are transferred for each axis, thus, there are $8 * 3 * 17 + 9 * 17 = 561$ features. The details of the transferred features are listed in Table 5 and Table 6. However, Figure 4 shows that there are still lots of outliers for each transformed variable, especially for te right part with negative values.

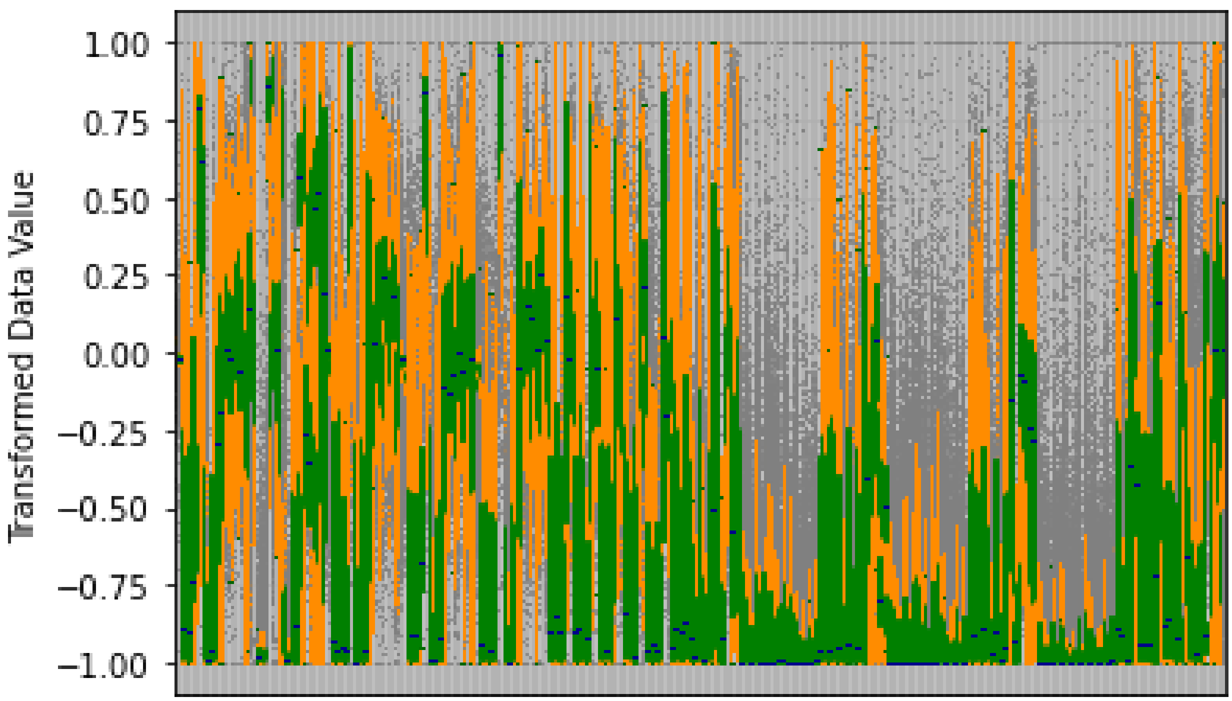


Figure 4. Box-plot for UCI Transformed Data Variables

The subjects were randomly selected into two groups, 70% of them in the training group and the rest in the testing group. In this case, we have 7352 records for training data and 2947 for testing data. The correlation coefficient matrix from the training data Figure 5 shows that 50% of the features have a coefficient larger than 0.25 with others. Almost 25% of the features have negative correlations. Meanwhile, there are highly positive correlations among groups.

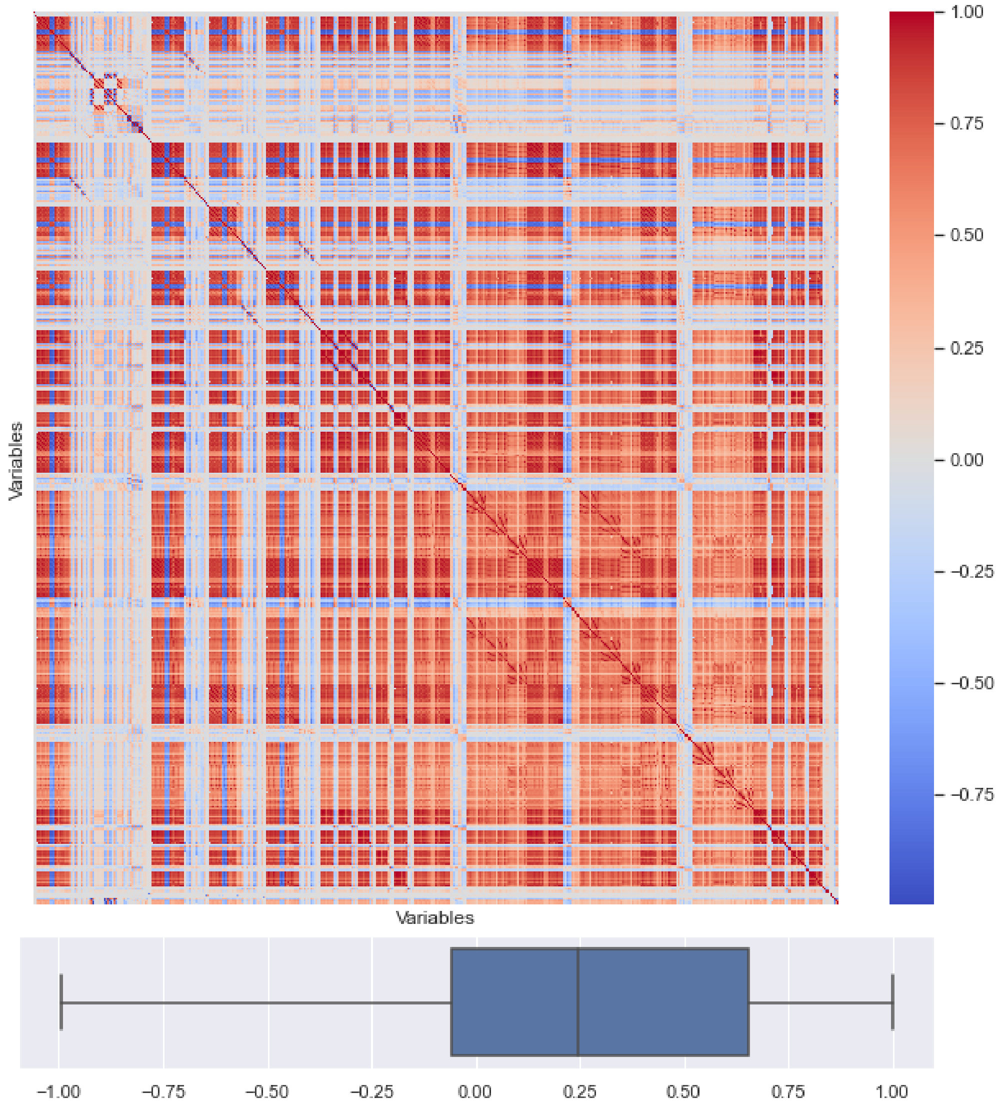


Figure 5. Correlation Coefficient Matrix for UCI Transformed Data

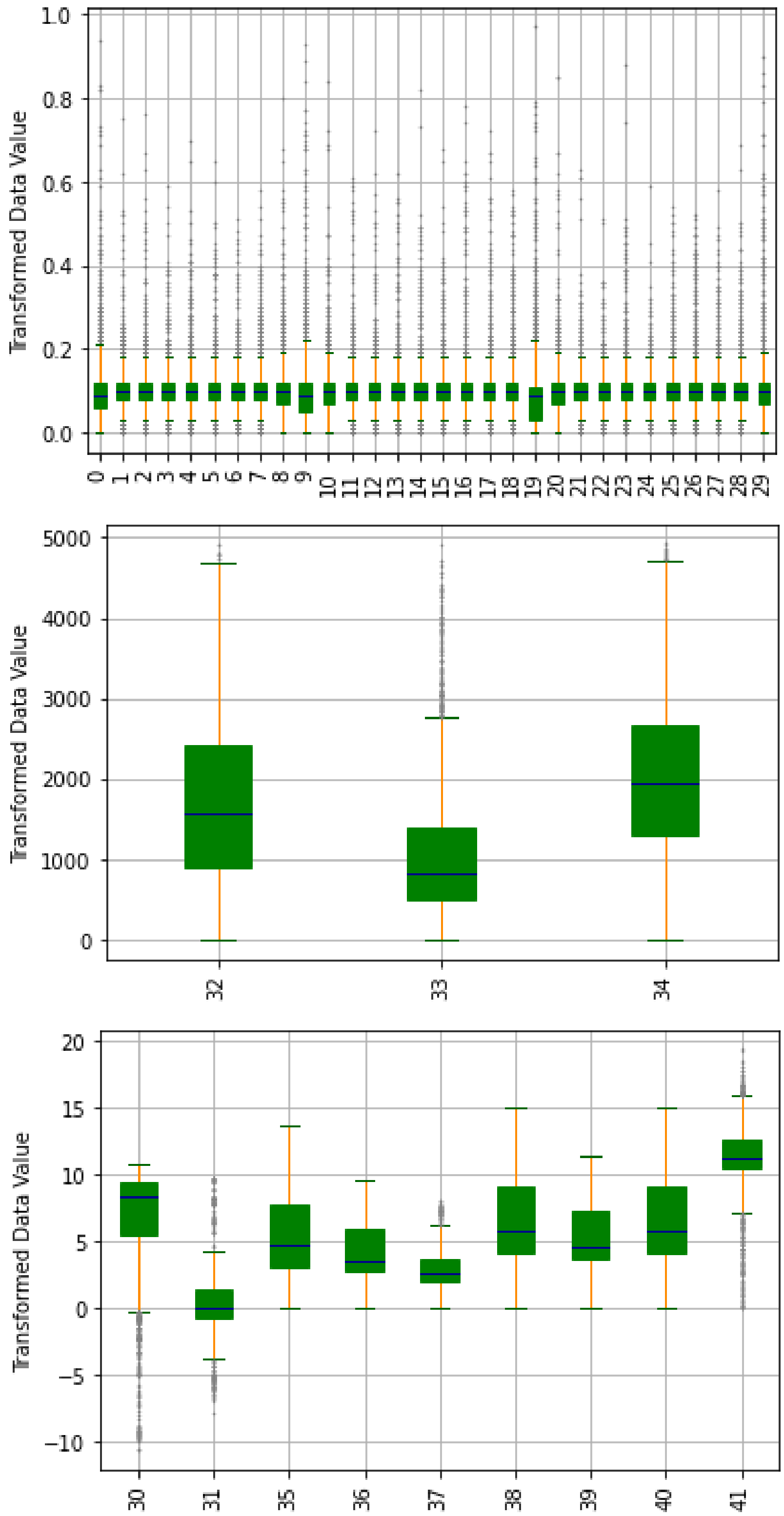
The WISDM dataset [56] had 36 volunteers with Android-based smartphones in their front pants pockets and they were asked to perform 6 activities for specific periods of time under monitoring, including walking, jogging, walking upstairs, walking downstairs, sitting, and standing. The data was recorded with 20Hz, lower than the UCI data (50Hz). The researcher used arffmagic program to transfer the raw data with a fixed 10 second window size. The new data has 43 features. These features include the difference between maximum and minimum, average sensor value, time between peaks, standard deviation, variance, and average resultant acceleration. Full description for the features is shown in Table 7. Since the average sensor value of X-axis is 0 for all the observations, we remove this feature in the analysis. The correlation coefficient matrix Figure 7 shows that most of the features have weak correlation with others, except the last

Table 5 UCI Features Groups

Groups	Name
Groups with signal axis ’-XYZ’ is used to denote 3-axial signals in the X, Y and Z directions	tBodyAcc-XYZ tGravityAcc-XYZ tBodyAccJerk-XYZ tBodyGyro-XYZ tBodyGyroJerk-XYZ fBodyAcc-XYZ fBodyAccJerk-XYZ fBodyGyro-XYZ
Groups with each pattern	tBodyAccMag tGravityAccMag tBodyAccJerkMag tBodyGyroMag tBodyGyroJerkMag fBodyAccMag fBodyAccJerkMag fBodyGyroMag fBodyGyroJerkMag

few variables (absoldev group, standded group, and the resultant). This is also implied in the box plot below, that most coefficients are in the range of $(-0.2, 0.35)$. With the boxplot for each transformed variable, we still see many outliers, especially for the first 30 variables. Again, the Runs test for the randomness gives all the p-values less than 0.0001, which is statistically significant for rejecting the null hypothesis for the randomness manner. As we noticed the same results from the UCI data, we will still use the data without any data process for our applications.

In brief, the UCI data includes $10,411$ total number of observations and WISDM includes 5,424. The percentage of each activity is shown in Table 8. As we did for UCI data, we split training and test data by 70% and 30% of the subjects, respectively. Since the researchers did not ask all the volunteers to perform all 6 activities, some subjects might have less activities.



a

b

c

Figure 6. Box-plot for WISDM Transformed Data

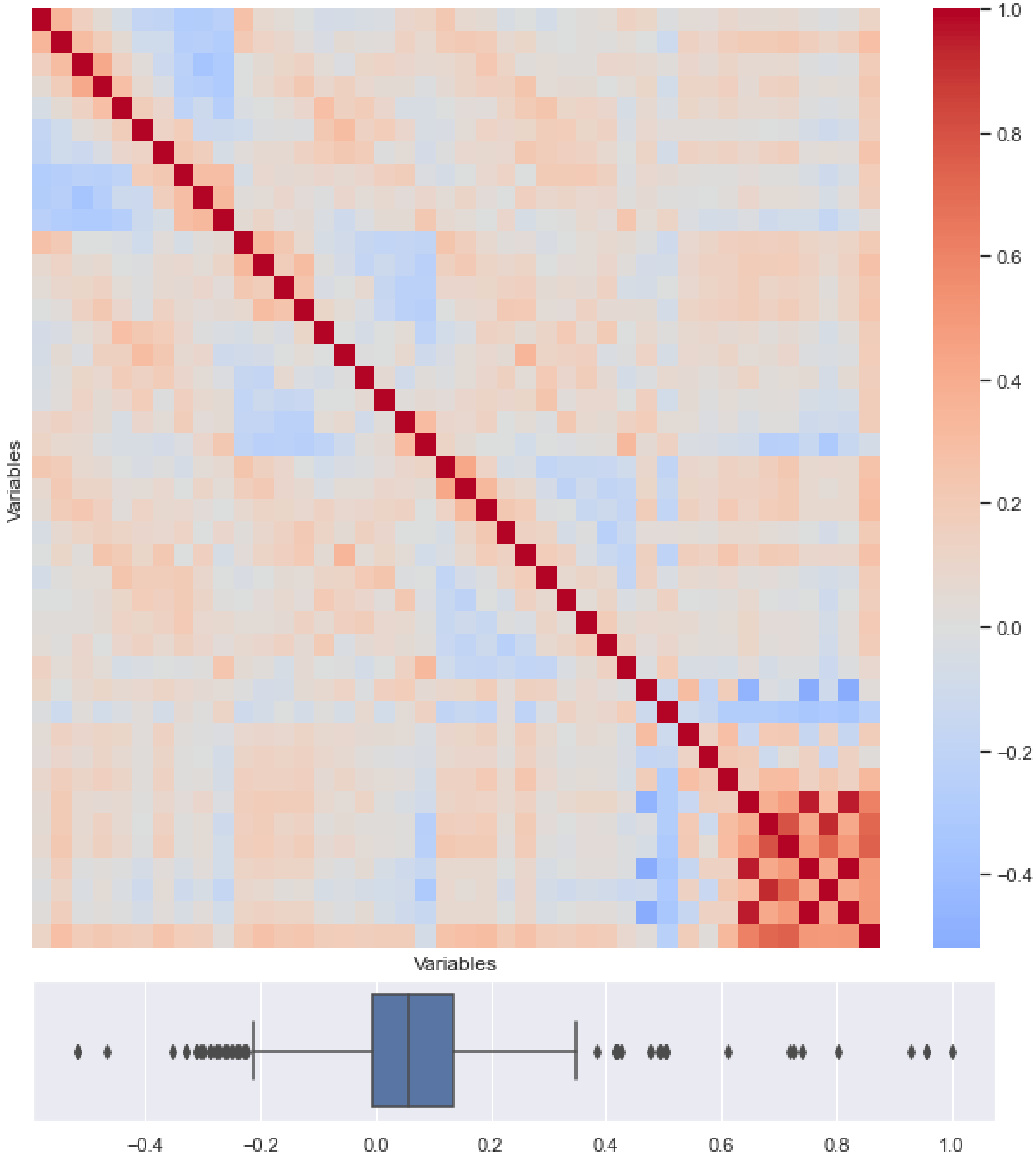


Figure 7. Correlation Coefficient Matrix for WISDM Transformed Data

Table 6 UCI Statistical Measurements

Variables	Details
mean()	Mean value
std()	Standard deviation
mad()	Median absolute deviation
max()	Largest value in array
min()	Smallest value in array
sma()	Signal magnitude area
energy()	Energy measure. Sum of the squares divided by the number of values
iqr()	Interquartile range
entropy()	Signal entropy
arCoeff()	Autorregresion coefficients with Burg order equal to 4
correlation()	correlation coefficient between two signals
maxInds()	index of the frequency component with largest magnitude
meanFreq()	Weighted average of the frequency components to obtain a mean frequency
skewness()	skewness of the frequency domain signal
kurtosis()	kurtosis of the frequency domain signal
bandsEnergy()	Energy of a frequency interval within the 64 bins of the FFT of each window.
angle()	Angle between to vectors

3.3 Related Works

The work of human activity recognition based on the sensors can be traced back to 1990s [32]. Sharma, Lee, and Chuang [80] applied neural networks (ANN) for a chest worn wireless sensor dataset and achieved 83.95% accuracy. Wu [93] 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 [10] used 561 transformed features to classify six different activities using a one vs. all support vector machine (SVM) and obtained as high as 89% accuracy. Kwapisz, Weiss, and Moore [56] from the WISDM Lab used Multilayer Perceptron and they got a best accuracy of 91.7%. Zhang, Wu, and Luo [100] 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, Liu, and Chen [42] performed a two layer and multi-strategy framework for sensor

Table 7 WISDM Statistical Measurements

Variables	Details
{X,Y,Z} {0-9}	These 30 features collectively show the distribution of values over the x, y, and z axes. We call this a binned distribution. For each axis we determine the range of values in the 10s window (max – min value), divide this range into 10 equal-sized bins, and then record the fraction of values in each bin.
{X,Y,Z}AVG	Average sensor value over the window (per axis).
{X,Y,Z}PEAK	Time in milliseconds between the peaks in the wave associated with most activities. Heuristically determined (per axis).
{X,Y,Z}ABSOLDEV	Average absolute difference between the each of the 200 readings and the mean of those values (per axis)
{X,Y,Z}STANDDEV	Standard deviation of the 200 values (per axis)
RESULTANT	Average resultant value, computed by squaring each matching x, y, and z value, summing them, taking the square root, and then averaging these values over the 200 readings.

Table 8 Total Data Structure Summary

	Walk	Up	Down	Sit	Stand	Lay	Jog
UCI (%)	16.5	14.8	13.5	17.3	19.0	18.8	
WISDM (%)	38.4	11.7	9.8	5.7	4.6		30.0

smartphone data and the result shows a 95.71% average accuracy. Besides, Ronao ad Cho [77] applied deep learning neural networks (DNN) to both raw sensor data and Fast Fourier Transformed smartphone data. Their work shows that the data with the transformed information provides average accuracy rate of 95.75%, which is 1% higher than the results from the raw data. Nakano and Chakraborty [69] 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. Ignatov [51] used CNN for the accelerometer data from smartphone. They obtain a 97.63% average accuracy with the statistical features. Besides, there are also some streaming works, which read the data immediately when it was generated. Na and Ramachandran [68] used the Online Bayesian Kernel Segmentation method for classifying 6 activities. The result shows a 92% average accuracy rate with the new segmentation data instead of fixed window data. In the paper, she first did the segmentation for new windows and then applied filters with these windows. Zhang and Ramachandran [102] used Very Fast Decision Tree method for online classification with the original fixed window transformed data. The result showed that the overall accuracy 85.9%. The advantage of this streaming method is that after inputting a user's own data to the model from lab data, the new model will be more personalized. Since not all the data and features are the same, it is hard to compare which method is better than others. But the challenge for most of the methods is the difficulty in discriminating between similar activities, especially for sitting and standing, walking upstairs and walking downstairs.

3.4 Methods

In this chapter we compare performance of some of the most popular machine learning methods for the smartphone based data with Python packages on a windows 10 laptop. including Support Vector Machine with three different kernels, K-Nearest Neighbor with different number of neighbors, Artificial Neural Network with 1 and 2 layers with different numbers of neurons, Linear Discriminant Analysis. Also, we compare these methods with the dimension reduction through Principal Components Analysis. For Support Vector Machine (SVM), Decision Tree and Random Forest methods, we used the grid search and 5-folds cross-validation to find the best hyper-parameters, since these parameters in the model cannot be estimated from the training data.

Linear Discriminant Analysis (LDA) is most commonly used as data classifier and dimensionality reduction technique in the pre-processing step for pattern-classification and machine learning applications. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability [96]. We use this method because of the high accuracy rate of the linear-SVM. The assumptions, such as the Gaussian distribution and the identical covariances with classes, are not verified. Thus, the results are subjected to the data.

SVM is an algorithm that finds classification boundaries so that categories are divided by a clear gap that is as wide as possible [25]. It can be defined as linear classifiers under the following the assumption that the margin should be as large as possible [5]. With the labels, the algorithm will output an optimal hyperplane which categorizes the data into different groups and the support vectors are the most useful data points because they are the ones most likely to be incorrectly classified. The basic form of SVM is a two-class classifier so we use the leave-one-out method, which considers the label y as 1 for the class or -1 as others. There are several commonly used kernels, including linear kernel, Gaussian radial basis function (rbf) kernel, Polynomial kernel, Gaussian kernel, etc., which could be applied to the data without prior knowledge. In this chapter, we applied the first 3 kernel functions. The reason of selecting these three kernels is that these three kernel functions show three different class boundaries, linear, non-linear, and curves. By the results from these kernels, we could then decide the boundaries. Before applying the SVM, we usually standardize the data. The common parameters we need to define are regularization parameter (C), Degree of the polynomial kernel function (degree), Kernel coefficient for rbf and polynomial kernel (gamma). In the experiments, we set C from 0.001 to 10 by 10, degree from 2 to 5 by 1, and gamma from 0.001 to 1 by 10.

K-Nearest Neighbor (KNN) is a non-parametric method that assign the class to a point by taking the majority of votes of its K neighbors [8]. KNN is based on the feature similarity, i.e., the more closely out-of-sample features resemble the training set, the more likely they are to be classified to a certain group. A characteristic of KNN is that it is sensitive to the local structure of the data. Euclidean Distance is used to measure the distance between two sample points.

$$d(x_1, x_2) = \sqrt{\sum_{i}^{n}(x_{1i} - x_{2i})^2}$$

where $d(x_1, x_2)$ is the distance between sample x_1 and x_2, n is the number of variables, and i is the i^{th} variable for the sample. We apply KNN here to see if there are any large distances between any two classes. The number of neighbors starts from 5 to 20.

Decision Tree (DT) and Random Forest (RF) are other two non-linear methods. We try Decision Tree here to see that if the greedy method can find a good cut-point for these continuous variables and select the better variables to do the split. Random Forest (RF) is an ensemble learning method that operates by constructing a large number of decision trees at training time and outputting the class that is the mode of the classes of the individual trees [49]. Since Random Forest do the voting and regardless of the collinearity, we also try this method. The most important parameter for these two are the branch split rules and the stop criteria. We use the Gini index for the measure of split quality because of less computation, the minimum number of samples for split is from 2 to 10 by 1, the maximum depth of the tree is from 3 to 10 by 1, and the number of trees in the forest is from 20 to 100.

Artificial Neural Network (ANN) extracts linear combinations of the inputs as derived features, and then model the target as a nonlinear function of these features and evolves to encompass a large class of models and learning methods [33]. Each neuron has an associated weight vector, which is assigned on the basis of its relative importance to the inputs. With the activation function, the neurons output the non-linear results. The advantage of using this method is that ANN has a flexibility to capture the non-linearities in the data. Moreover, it could extract different linear relations.

Principal Component Analysis (PCA) is a dimensionality-reduction technique that is often used to transform a high-dimensional dataset into a lower dimensional subspace. PCA finds the principal components of the dataset by transforming the data into a new coordinate system. In the new subspace, the first axis corresponds to the first principal component, which is the component that explains the greatest amount of the variance in the data. Considering the UCI data has 561 variables, it is very unlikely that all the variables are independent. To overcome this problem, in this paper we also implemented PCA and analyzed the resulting data. Considering the small number of the features (42) in the WISDM data, we did not apply PCA to this data.

3.5 Experiments and Results

We apply these machine learning methods to 70% of the subjects to train our model with grid searching for best parameters. In this case the validation data for parameter optimization is pulled from the training data. The number of observations for each activity in training data is shown in Table 9. The UCI data has balanced groups, while WISDM data has different training observations for each activity. This is an unbalanced dataset, but here we will conduct the experiment same as if it was a balanced data. As we see from the table, Upstairs, Downstairs, sitting and Standing have less observations in WISDM data than in UCI data.

Table 9 Number of Observations in Training Data

Activities	Walk	Up	Down	Sit	Stand	Lay	Jog
UCI	1226	1073	987	1293	1423	1413	
WISDM	1529	447	375	190	170		1200

Since we used grid search for some of the parameters, Table 10 shows all the optimized parameters and the time to converge for each method. Grid search with 5 folds are time consuming, for example, SVM with polynomial kernel (Poly-SVM) and RBF (RBF-SVM) costed more than 6000 seconds, even for KNN run for 1210 seconds. And we show the accuracy rate of the test data in Table 11 to compare the performances of the models and make a rough conclusion of the relations between different activities. It shows that some of the models with WISDM data perform better than those with UCI data, such as the SVM with RBF kernal and the three ANNs, while others are opposite. From this table, the best classifier for these two datasets are different. It shows that SVM with linear kernel and LDA have the best performance with dataset UCI (96.2%). This means that the different activities exist linear classifiers based on the feature combination.

Table 10 Optimized Parameters and Time to Convergence for UCI

Methods	Parameter Settings	Time (seconds)
Linear-SVM	C=1	987
RBF-SVM	C=10 , gamma=0.01 , degree =2	8562
Poly-SVM	C=0.1, gamma= 0.01, degree =3	6528
KNN	n=10	1210
LDA	None	≤ 10
DT	max_depth= 8, min_split= 2	572
RF	# of trees = 100, max_depth= 5, min_split= 2	$\geq$ one day
One-layer-NN	layers=1, unites =30, activation='relu'	≤ 10
One-layer-NN	layers=1, unites =50, activation='relu'	≤ 10
Two-layer-NN	layers=2, unites =(30,6), activation='relu'	≤ 10

This is much better than the result of 89.3% in [10], which also used SVM methods. The reason might be because of the setting of hyper-parameters, which we got from grid search and cross validation. And The ANNs have the best performance for dataset WISDM (upto 98.3%). This is also better than the result of 91.7% from [56], which used Multilayer Perceptron. This is because we checked the data structure and features and found the feature named average sensor value of X-axis is a constant 0. Thus, we removed it before the analysis. The result shows that better data pre-process might improve the model performance.

The KNN, LDA, DT and RF perform much better with the UCI data than with WISDM data. This might have many reasons. Comparing these two datasets, there are two main differences. Firstly, the different ways of collecting and transforming of raw sensor data. Secondly, the different body locations for the sensors placed. The KNN and DT do not perform as good as others with both datasets. In other words, the Euclidean distance used in KNN might not be suitable for this type of data, which means the variance within each activity might vary. And finding the important features and splitting these might have difficulties for DT. Moreover, this comparison also shows that the neural network does not always "win". It depends on the data type or might be the proportion from the number of variables and the total sample size. Besides, increasing the layers and the neurons does not improve the accuracy much in the neural network, which also means that the dataset might have linear relationships among activities. It also implies that non-linear methods could do better identification for WISDM data with a small amount of observations.

Table 12 and Table 13 show the prediction details from each method. These two tables show that most of the methods can successfully identify Walking, Lying and Jogging with more than 95% in accuracy rates.

Table 11 Methods Comparison with Average Accuracy (%)

Methods	UCI	WISDM
Linear-SVM	**96.4**	89.9
RBF-SVM	95.3	**97.9**
Poly-SVM (2)/(3)	93.7	90.8
KNN (10)	88.5	78.8
LDA	**96.2**	76.0
DT	86.0	73.1
RF	92.5	80.4
One-layer-NN (30)	94.6	97.3
One-layer-NN (50)	94.4	97.3
Two-layer-NN (30, 6)	95.2	**98.3**

They are also able to identify the Upstairs and Downstairs with a decent accuracy rate, which is around 90% for UCI data. But it is a big challenge for WISDM, some of the accuracy rates are even below 50%. The biggest challenge is to identify the sitting and standing for UCI, but walking upstairs and walking downstairs for WISDM. These two activities have the similar pattern and very small differences with the raw sensor data, such as the frequency. Thus, they are difficult to be classified correctly from each other. For example, the result from Linear-SVM for the UCI data shows that, 59 out of 66 misclassified sitting are identified as standing, and all of the misclassified standing are identified as sitting. The same thing to the LDA, 62 out of 63 misclassified sitting cases are identified as standing and all of the 23 misclassified standing are identified as sitting.

Then, we use PCA to reduce the data dimensions for the UCI data. Since the WISDM only has 42 features, PCA is not necessary. Figure 8 shows that the proportion of variance explained by each component is less than 0.1% after the first 40 principal components. And Figure 9 shows that the cumulative proportion of variance explained by the components are close to 1 (99.99978%) after the first 200 principal components. Then we can have a conclusion that the first 200 principal components are sufficiently explained the data information and we can reduce the data dimension from 561 to 200.

Table 12 UCI Test Average Accuracy for Activities

Methods	Walk	Up	Down	Sit	Stand	Lay
Linear-SVM	0.99	0.97	0.97	0.87	0.97	1.0
RBF-SVM	0.97	0.97	0.92	0.90	0.96	1.0
Poly-SVM (2)	0.99	0.94	0.84	0.89	0.95	1.0
KNN (10)	0.98	0.88	0.73	0.81	0.93	0.94
LDA	0.99	0.98	0.96	0.88	0.96	1.0
DT	0.88	0.78	0.83	0.77	0.88	1.0
RF	0.97	0.90	0.84	0.88	0.94	1.0
One-layer-NN (30)	0.99	0.94	0.93	0.89	0.95	0.97
One-layer-NN (50)	0.99	0.95	0.92	0.87	0.96	0.96
Two-layer-NN (30, 6)	0.99	0.94	0.93	0.88	0.97	1.0

Table 13 WISDM Test Average Accuracy for Activities

Methods	Walk	Up	Down	Sit	Stand	Jog
Linear-SVM	1.0	0.54	0.69	0.84	0.97	1.0
RBF-SVM	0.99	0.99	0.99	0.91	0.79	1.0
Poly-SVM (3)	0.96	0.70	0.89	0.68	0.91	1.0
KNN (10)	0.94	0.36	0.14	0.87	0.82	0.98
LDA	0.66	0.33	0.3	1.0	0.8	0.97
DT	0.71	0.47	0.3	0.85	0.95	0.96
RF	0.76	0.55	0.0.47	0.98	0.84	0.97
One-layer-NN (30)	1.0	0.99	0.99	0.73	0.91	1.0
One-layer-NN (50)	1.0	1.0	0.99	0.75	0.93	1.0
Two-layer-NN (30, 6)	1.0	0.99	0.99	0.84	0.93	1.0

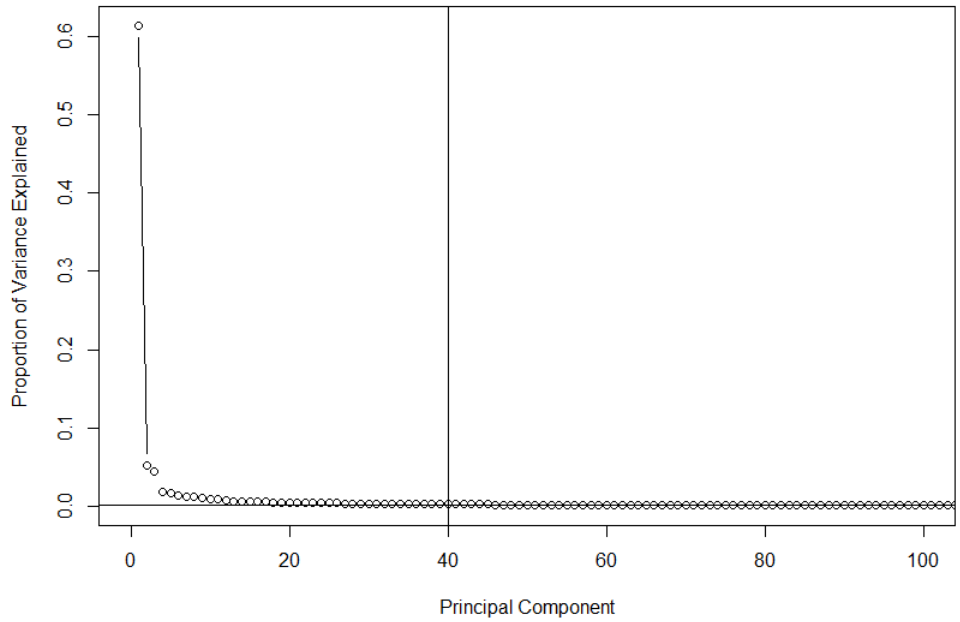


Figure 8. Proportion of Variance

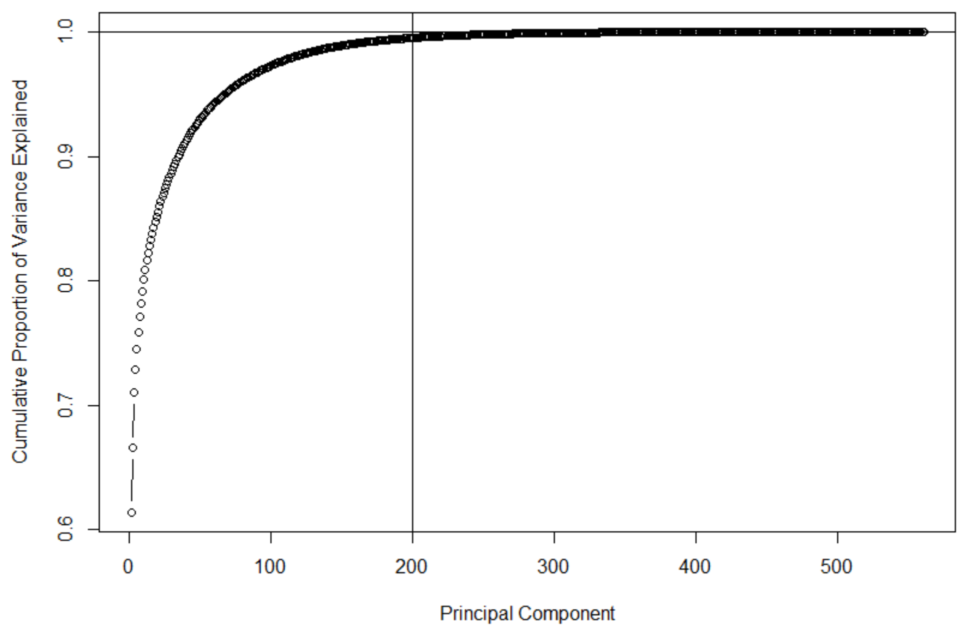


Figure 9. Cumulative Proportion of Variance

Table 14 Comparison With PCA Methods

Methods	# of Components	Avg. accuracy (%)	Methods	# of Components	Avg.accuracy (%)
Linear-SVM	50	91.49	LDA	100	93.69
	100	94.43		150	95.16
	150	96.06		200	95.96
	200	96.46		250	96.36
	250	96.06		300	96.53
RBF-SVM	100	96.43	KNN(10)	100	89.89
	150	94.73		150	90.02
	200	93.52		200	90.39
Poly-SVM(2)	100	92.59	KNN(20)	100	90.15
	150	92.39		150	90.62
	200	91.32		200	90.45
One-Layer-	100	94.55	Two-Layer-	100	95.45
NN(30)	150	95.02	NN(30,10)	150	95.10
	200	95.78		200	96.14

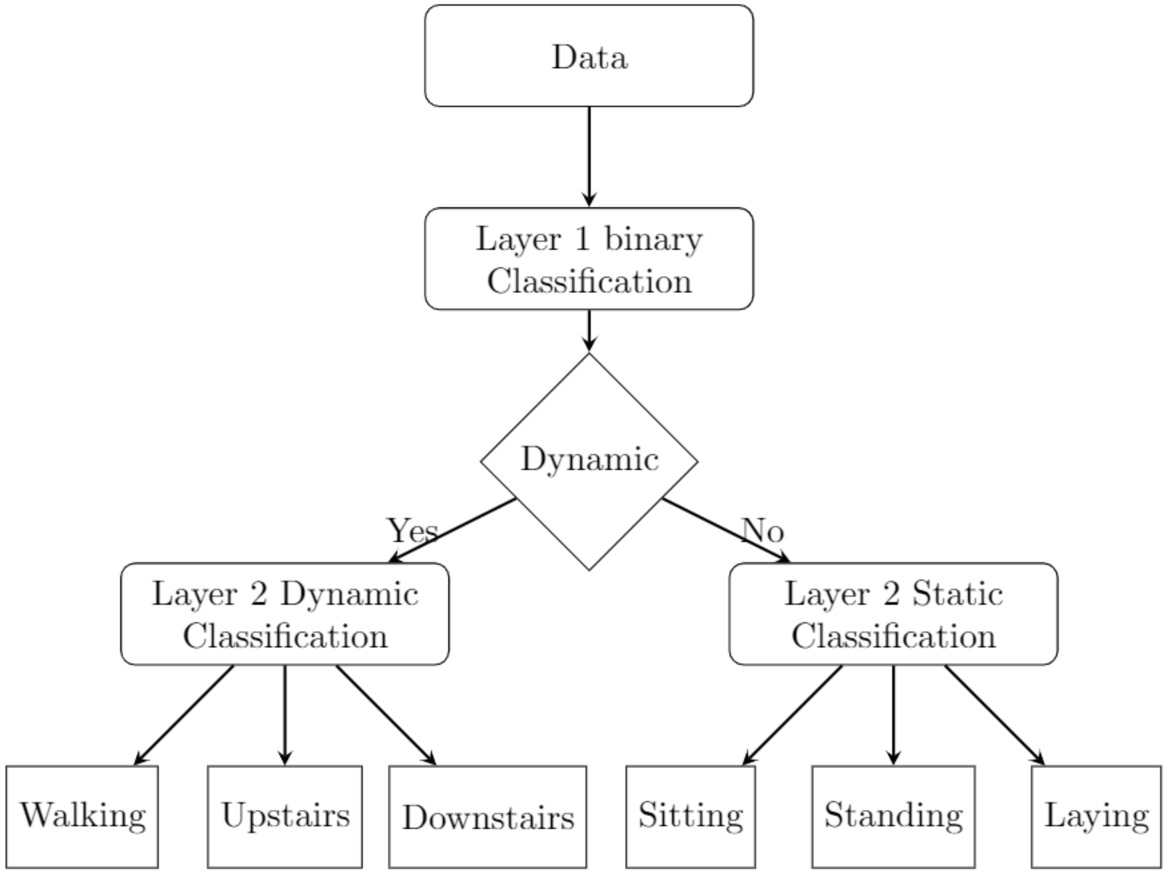


Figure 10. Two Layers Method Flowchart

As we can see from Table 14, the PCA method successfully reduces the data dimension to less than half of the original data dimensions without losing the features variance. Again, the linear classifiers have better performance, since the highest accuracy rates are from Linear-SVM with 200 principal components and the LDA with 250 and 300 principal components, with accuracy 96.06%, 96.36%, and 96.53%, respectively. Additionally, Poly-SVM and KNN do not perform as well as others. From this, it is reasonable to think that PCA reduces the data dimensions but does not change the data structure and the relationships among categories. However, there is little improvement for the average accuracy, which still around 96%. In this point of view, PCA reduces the dimensions effectively without loss in performance.

Table 12 and Table 13 also imply that most of the algorithms are good at identify dynamic activities but have worse performance with static activities. To create an algorithm which has more flexibility to adapt two different methods according to the type of activity, we propose a two layers method for UCI data, as shown in Figure 10. By experiments, we selected the LDA as the first layer binary classifier. There are two reasons. First, LDA gives a good result for identifying dynamic and static activities. It has a high degree of average test accuracy rate, close to 100%, for 1 out of 1609 static cases misclassified as dynamic. Secondly, LDA has stable result. All the parameters are from training data, while linear-SVM have hyper-parameters. We used Linear-SVM and Poly-SVM for dynamic and static classification, respectively. These selections were based on the performance of the algorithm with the single group data. Table 15 shows that the average test accuracy is improved a little by using different classifiers. The biggest challenge is Sitting and Standing identification. The problem might imply that the transition of these two activities sensor data might not be appropriate. It loses the power to extra characteristic for these two.

3.6 Conclusion

Comparing results from two different datasets, the best method for each datasets are different. It seems that the linear methods prefer better than others for UCI, while the non-linear methods prefer better for WISDM. It is also obvious that the methods perform better with the UCI data than these with WISDM data. The reasons are not clear yet. But we might guess that the way of collecting and transforming the raw data. UCI data had higher frequency 50Hz and first applied median and butterworth filters to reduce noise and used FFT to transform the signal data, while WISDM lab collected in 20Hz and used the original signal data with arffmagic program. More detail information is collected for UCI. This might affect the window fixed transform data. Based on this result, we might have a conclusion that for human activities recognition with smartphone sensors, the FFT with statistical transformation is better than arffmagic program and might be in the lab experiment, noise reduction might be necessary. Besides, without resorting to complicated methods, simple linear classification is sufficient for data analysis.

Table 15 UCI Test Average Accuracy for Activities Two Layers Method

Methods	Two Layers Method							Linear-SVM
Activity	Walk	Up	Down	Sit	Stand	Lay	Recall %	Recall %
Walk	**492**	2	2	0	0	0	99	99
Up	20	**451**	0	0	0	0	96	97
Down	4	9	**407**	0	0	0	97	97
Sit	0	1	0	**464**	43	0	91	87
Stand	0	0	0	19	**537**	0	97	97
Lay	0	0	0	2	0	**543**	100	100
Precision%	95	97	100	96	93	100	**96.6**	96.4

[1]The number in the circle is the misclassified case in the first layer.

Chapter 4

Adjusted Method for Human Activity Recognition

Note to Reader:

Portions of this chapter have been previously published in IEEE, 2019, pp 1-6, and have been reproduced with permission from IEEE Publishing. © [2019] IEEE. Reprinted, with permission, from [Yanjia Zhang and Kandethody M. Ramachandran, Human Activity Recognition with Streaming Smartphone Data, IEEE GCAT, and Oct.2019]

4.1 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 health care, fitness monitoring, and elder monitoring, are based on this kind of dynamic data. Although there are many online methods that have made a great progress in analyzing these kinds of data, it is still a big challenge to get good results from a streaming data perspective. Currently, streaming data analysis methods are in its infancy. In this chapter, we use an online method called Very Fast Decision Tree (VFDT) to mimic the real scenario, since we do not have data collected in a streaming basis. 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 close to 100%.

4.2 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 [44]. Meanwhile, there are several challenges and bottlenecks for these data from wearable devices to be more useful and reliable in medical purposes [44]. 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 [1] as the case study. We also use WISDM lab data [56] to explore the adaptive power of this model.

4.2.1 Previous Works

The work of human activity recognition based on the sensors can be traced back to 1990s [32]. Sharma et al.[80] applied neural networks (ANN) for a chest worn wireless sensor dataset and achieved 83.95% accuracy. Kwapisz [55] 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 [47] combined Principal Components Analysis (PCA) and Support Vector Machine (SVM) to classify four activities and got 97.5% average accuracy. Sohn and Khan [81] 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. [93] 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.[10] used 561 transformed features to classify six different activities using a one vs.all SVM and obtained as high as 96% accuracy. Fergani [31] used PCA based multi-classfier to get 96.9% average accuracy for daily activities. Zhang, Wu and Luo [100] 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. [42] 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 [77] 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 [69] 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 [51] used CNN for the 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 [50]. 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 [43] proposed an on-line HAR mobile system. Wang, Liang, et al. [90] 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. [71] 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. [4] used a sequential extreme learning machine method (OSELM) and achieved an average accuracy of 82.05%. Shuang Na, et al [68] 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 16. 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 are more suitable for recording daily activities. [4] and [68] are two examples of this. They both use the same data from UCI. Unfortunately, [4] needs a large size window segmentation to train the hidden layers. And [68] only uses the last data window to create a new classifier but forgets all the previous information. Both [4] and [68] lack the ability to adapt the incoming labeled data from single users and might violate the stationary assumption at the very beginning.

Table 16 Online Methods Summary

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

To address the above challenges and try to improve the existing methods, we propose an online tree based method with pre-processed 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 [29]. The purpose of this algorithm is to deal with continuous data streams by building decision trees using constant memory and time per example [29]. This method is used in many streaming fields, including fraud detection [65, 66], and sensor networks [24, 35, 36]. It can also be applied for handling missing values [82] and implementing in distributed environment [27]. 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.

4.2.2 Structure

In this chapter, 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.

4.3 Data Processing

In this paper, we used the smartphone data from UCI [1] and WISDM Lab [56]. 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 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 activity in Table 17, 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.

Table 17 Size of Activities in UCI Training data

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

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 18. Since there are two volunteers who did not perform Jogging, we ignored these data in out testing. Thus, there are 25 cases.

Table 18 Size of Activities in WISDM Training data

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

4.4 Proposed Streaming Methods

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 Principal Components Analysis (PCA), which is most used in the research, such as [47, 81] and [31].

4.4.1 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 [57]. 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 [6]. 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 Figure 11.

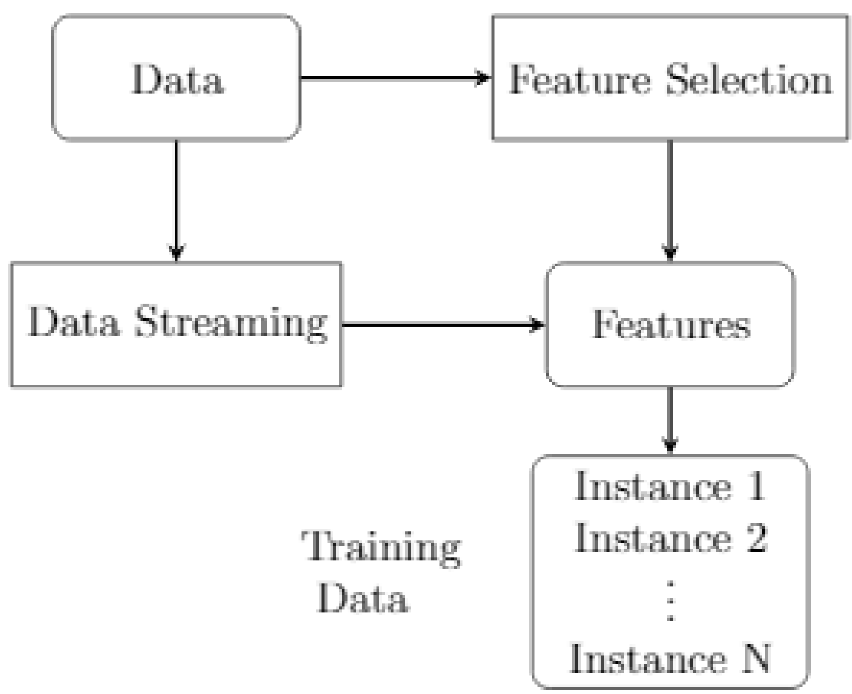


Figure 11. Feature Selection and Streaming Data Creation

4.4.2 Very Fast Decision Tree (VFDT)

Geoff,etc.[50] 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 bound, which states that with probability *(*$1 - \delta$*)*, 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:

$$\epsilon = \sqrt{\frac{R^2 ln(1/\delta)}{2n}}. \tag{4.1}$$

Equation 4.1 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 [74], and Gini Index (GI) which estimates the probability of misclassification under the split [62]. 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 Figure 12.

The VFDT Algorithm

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* VFDT

VFDT {*S, X, IG,* δ, τ, 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* each (x, y) in S *Sort (x, y) into leaf* l *using DT* each x_{ij} in x such that $X_i \in X_1$ *Increment* $n_{ijy}(l)$ *Label* l *with the majority class among* $n_{ijk}(l)$

(examples at l are not in the same class and $n_{ijy}(l)$ mod $n_{min} = 0$) *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* $\epsilon = \sqrt{\frac{R^2 \ln(1/\delta)}{2n_{ijk}(l)}}$ *(*$\triangle IG > \epsilon$ *or* $\triangle IG \leq \epsilon < \tau$ *and* $X_{i1} \neq X_\emptyset$*) Add new split to* l *with* X_{i1} *and have a new leaf* l_m *Let* $n_{ijk}(l_m) = 0$

Return VFDT

4.5 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. The number of features is reduced to be 36. Then we used these 36 features to create an online tree model. 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 19.

Table 19 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[1]	0.92	0.87	0.77	0.65	0.91	1.0	0.859

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

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 Figure 15, 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 Figure 14. The rest of 3 activities are more complex as the changes between them are tiny, such as in Figure 14 and Figure 16.

To show the ability of updating our model to new activities, we use another data set from WISDM Lab [56]. 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 close to 100% and 16 out of 25 accuracy rates are higher than 90%. More details can be found in Table 20. 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 20 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[2]	0.84

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

4.6 Conclusion

To provide a human activity recognition system with automatic updating and adjusting, an online streaming system is required. Most of the methods in the literature are online, 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 close to 100% of accuracy. For a new activity, VFDT gives an average of 84% accuracy rate and above 90% accuracy for 64% of the testing people.

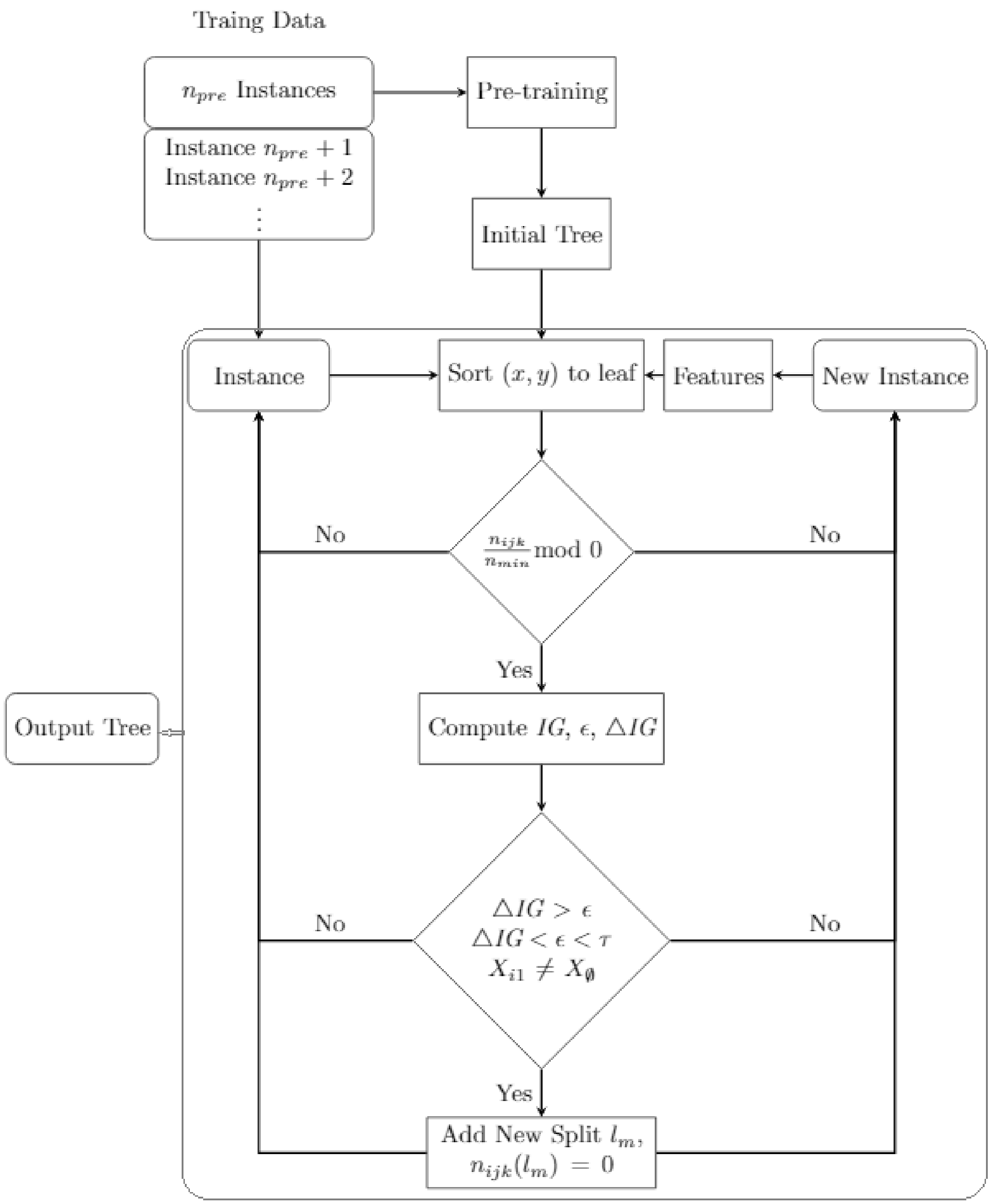


Figure 12. VFDT with DT Pre-training Diagram.

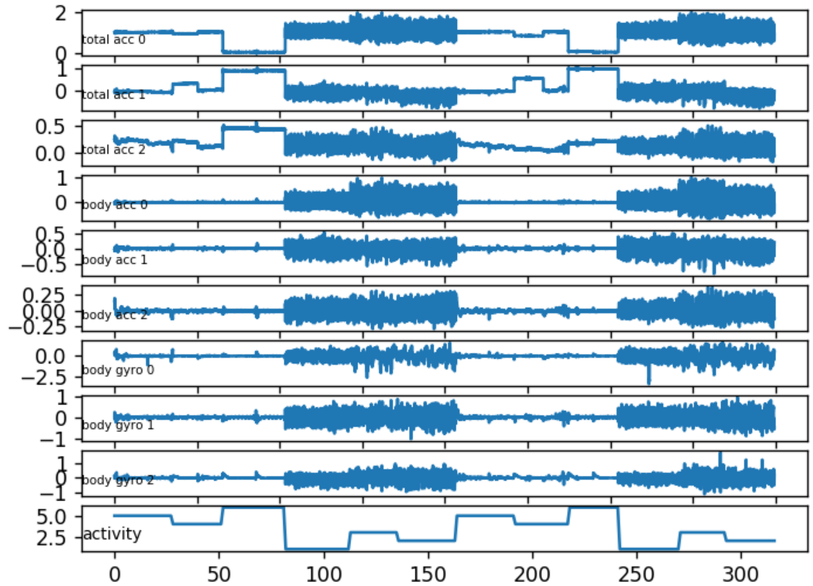


Figure 13. Example Sequence for Subject 2

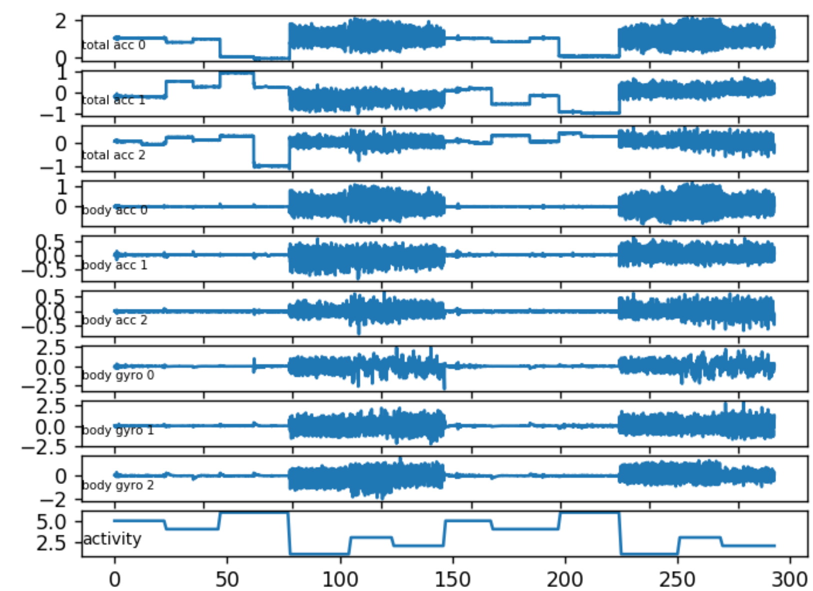


Figure 14. Example Sequence for Subject 4

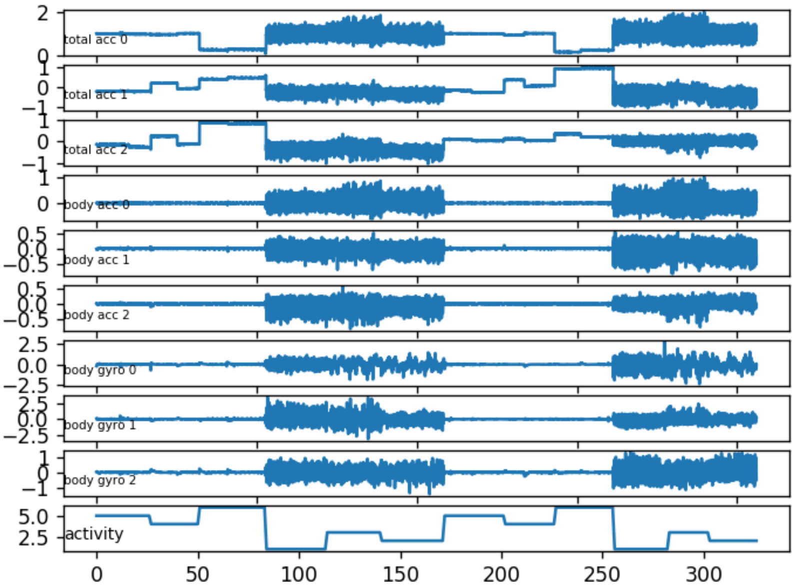


Figure 15. Example Sequence for Subject 6

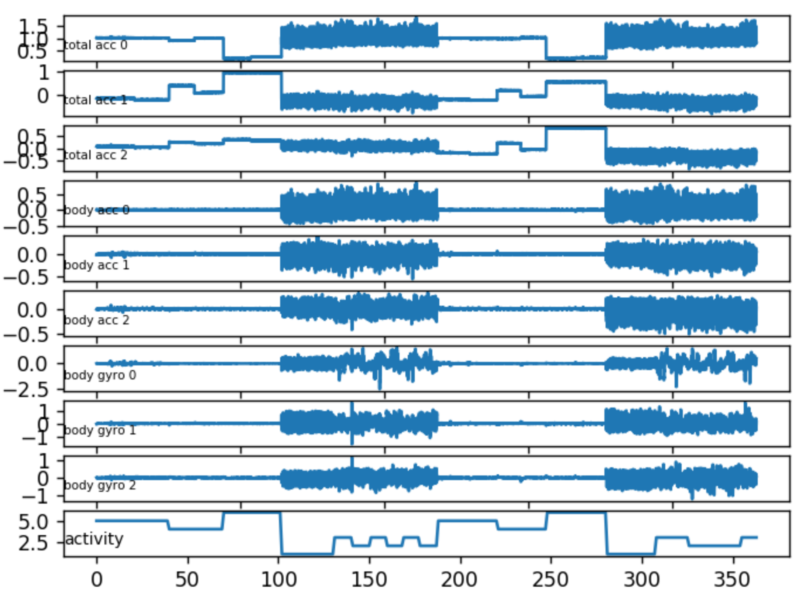


Figure 16. Example Sequence for Subject 7

Chapter 5

Personalized Adjusted Method for Human Activity Recognition

As mentioned in Chapter 2, most research focus on improving the results accuracy rate, including creating classification models and transforming sensor data, reducing the process time, etc. Although, the final overall accuracy rates are improved by the complex models and some of the activities could be identified for near to 100%, there are still some activities are confused to other activities, such as walking upstairs and downstairs, sitting and standing. Meanwhile, from our previous study, the data is different from person to person. Thus few research start to pay attention to the personal adjustable HAR system, for example [14].

A personal HAR system means that this system could be adjusted as customer's requirement, such as transfer the classifier model from the general data to customer's data model and add new frequently occurred activities to the system. With this aim, we propose an HAR system which could be considered as personalized adjusted HAR system. In this system, we show the performance of a combination method with unsupervised and supervised learning. For the unsupervised method, we choose the k-means cluster as it could gather the same activity as a cluster. For supervised method, we choose discriminant analysis. According to [103], LDA gives well performance for identifying the activities for both UCI and WISDM data, we first consider the discriminant analysis method as the classifier. This model is flexible to adjust personal based activity patterns and add new activities. While the human activity patterns is different from person to person, it is better to use the updating data to modify the parameters to improve the result and adjust the system with personal performance, for example adding some additional activities, instead of using the built-in model from the lab training parameters. Besides, it is also beneficial to discover if the lab data could be generalized to personal activity pattern. With this purpose, we create 4 different datasets from the WISDM data.

5.1 Data

These two data sets are from WISDM lab. It has totally 36 subjects with smartphone sensors and each subject performs 6 activities. Each sensor records the raw data in 20Hz with x, y, and z dimensions. The data we

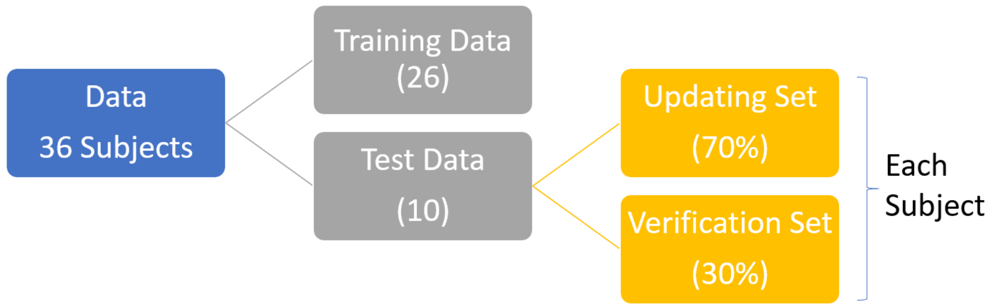


Figure 17. Data Sets

used here is transformed data based on a fixed 10 seconds window. In other words, each observation from the transformed data is based on 200 raw records (10s x 20 records/s). The features contain 43 variables, include Average (3), Standard Deviation (3), Average Absolute Difference (3), Average Resultant Acceleration (1), Time Between Peaks (3), and Binned Distribution (30). The details of the variables are shown in the table Table 7.

We randomly divide the data into training set and test set with 26 and 10 subjects from the WISDM1, respectively. To create a personal activity recognition model, we then randomly divide each subject in the test data into two groups: 70% for updating set and 30% for verification set for each activity. The data is divided as Figure 17.

To compare the results from adaptive discriminant model with the results from the whole training data with 70% of the personal data and the personal data only, three different datasets, Data 1, Data 2, and Data 3, are used. Meanwhile, Data 4 is a new training set which excluding the Jogging activity from the old training set and keeping the same for the updating and verification set. The details for these datasets are shown in Table 21. For example, subject s is from test data, and we then divide this series data into 70% for updating the personal model and 30% for test the model's performance for each activities, respectively. Specially, Data 4 has a new activity Jogging in the test set, which not included in the training set.

The purpose of setting such different kinds of dataset in the training, we try to discover the potential relation between the model and the single person's activity pattern. The training set in Data 1 includes the updating set, means it includes customer's activities information in the training, and the data in test set could find the same person's information. The training set in Data 2 only includes the updating set, means it only has the personal data, and the test data is from the same person in training. Then, training in Data 3 includes the training set, and uses the updating set to adjust the training model, then test the adjusted model. Data 4 is used for adding new activities. Jogging is not in the training set, but in the updating set, and then test

Table 21 Training and Test Data Sets

Name	Training Set	Updating Set	Test Set
Data 1	Training + Updating Set	None	Verification Set
Data 2	Updating Set	None	Verification Set
Data 3	Training	Updating Set	Verification Set
Data 4	Training without Jogging	Updating Set for Jogging	Verification Set

the model with all 6 activities. By comparing the results, we might know if the activity pattern could be generalized from sample lab data.

5.2 Method

The personal activity recognition model includes two stages: one is the unsupervised machine learning stage, which is used to group a set of objects with similar characteristic, which means they could possibly be interpreted as the same activity. With this purpose, the best choice will be the cluster methods. Considering the computational time and the useful of centroids, the k-means is a suitable method for the data. Then, we could use the centroid set of each cluster comparing with the mean set of each activity to detect if new activities are performed. Another stage is the supervised learning, Quadratic Discriminant Analysis (QDA) is used for classification. Instead of assuming the data has equal covariances for classes to use linear classifier, QDA is more adaptive to the different covariances with the quadratic discriminant functions. And from boxplot of the correlation distribution for each activity in fig. 7, we could see that most likely, the covariance matrix for each activity is different, especially for the walking upstairs and downstairs. To proof this, we use two sample Kologorov-Smirnov (KS) test, which is sensitive to differences in both location and shape of the empirical cumulative distribution functions, to test if the covariance matrix are significant by pair. The most important reason for us to choose the KS test is that it is a non-parametric statistic, which means assumptions are not required. According to Figure 18, the distributions from Walking Upstairs and Downstairs are different than others, Walking, Sitting, and Jogging might have a similar distribution. We apply the KS test on the Walking vs. Walking Upstairs, Walking vs. Walking Downstairs, Walking vs. Jogging, and Sitting vs. Standing. The Hypotheses are these two covariance matrices for two activities are the same. The test results are shown in Table 22. As we expected that Walking Upstairs and Walking Downstairs are significantly different from the Walking, while Jogging is statistically not significantly different from Walking based on 5% significance level. However, the p-value from Sitting vs. Standing is also less than 0.0001.

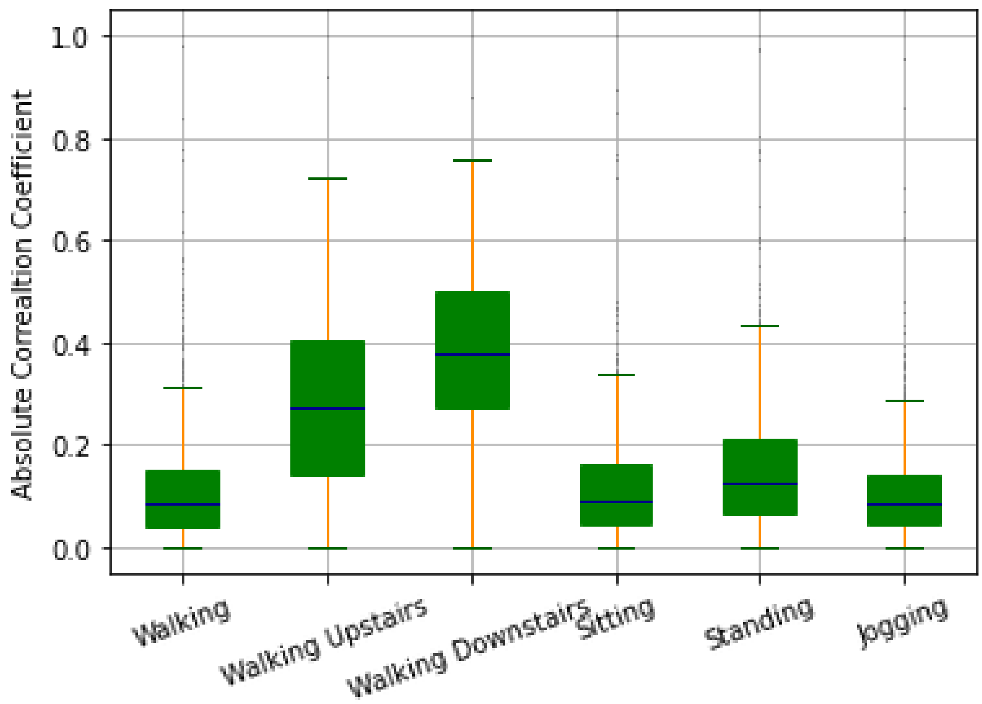


Figure 18. Correlation Coefficient Distribution for Classes from

Table 22 Kologorov-Smirnov Test Results

Activities	Results (p-value)
Walking vs. Walking Upstairs	p-value < 0.0001
Walking vs. Walking Downstairs	p-value < 0.0001
Walking vs. Jogging	p-value $= 0.513$
Walking vs. Sitting	p-value $= 0.046$
Sitting vs. Standing	p-value < 0.0001

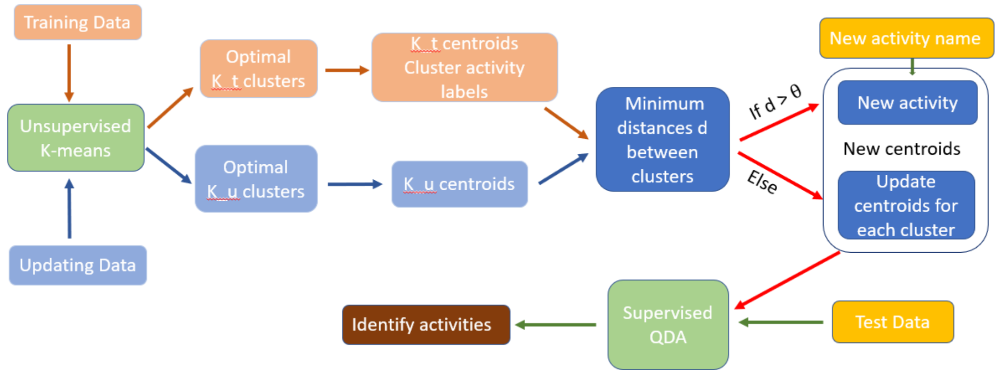


Figure 19. Personalized Adjusted HAR Process

After we decide these methods, the two step process will be shown in following figure Figure 19.

Suppose that we have a set of $N \times D$ dimensional samples, D is for number of features, $x_1, x_2, \cdots, x_N$ belonging to K different classes and n_k represents the number of instances of training data in class k. By minimizing the within cluster sum of squares, K-means method shows C clusters, C as the number of clusters from the training result. Let n_j represents the number of instances in cluster C_j. The distortions objective function will be:

$$J = \frac{\sum_j^C \sum_{i_j}^{n_j} ||x_{i_j} - c_j||^2}{C}$$

The elbow point will be selected as the optimal C and the euclidean distances between centroids set and the mean set μ from true classes are calculated and the criteria is made, e.g. the average of the distances will be the θ. If the distances are larger than the threshold θ, then new activities will be considered. Suppose data U is the set from clusters that have a larger θ from any μ.

$$\delta_k(x) = -\frac{1}{2} log|\Sigma_k| - \frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k) + log\pi_k,$$

here

$$\Sigma_k = \sum_{k=1}^{K} \sum_{i=1}^{n_k} (x_i - \mu_k)(x_i - \mu_k)^T / (N - K),$$

$$\pi_k = N_k / N,$$

and μ_k is the mean vector of the training data belonging to class k.

The decision boundary between class k_1 and k_2 can be described by

$$\{x : \delta_{k1}(x) = \delta_{k2}(x)\}$$

Since discriminant function is based on the class means, updating classifiers can be modified by the training class means (μ_k) and the added class means ($\mu_k^{(2)}$). Let new observations of the updating data $x'_1, x'_2, \cdots, x'_L$ set L is incremented. l_k is the number of observations of added data belonging to class k. The new class is possible to be incremented by the updating data, and the number of classes becomes K'.

Thus the new mean vector for data belongs to K is

$$\mu'_k = (1-p)\mu_k + p\mu_k^{(2)},$$

here p is the weight for the added class mean vector. For new activity as $(K+1)^{th}$ class,

$$\mu'_{(K+1)} = \mu_{(K+1)}^{(2)},$$

and the new co-variance matrix can be written as

$$\Sigma' = \sum_{k=1}^{K'} \sum_{i=1}^{l_k} (x'_i - \mu'_k)(x'_i - \mu'_k)^T / (L - K')$$

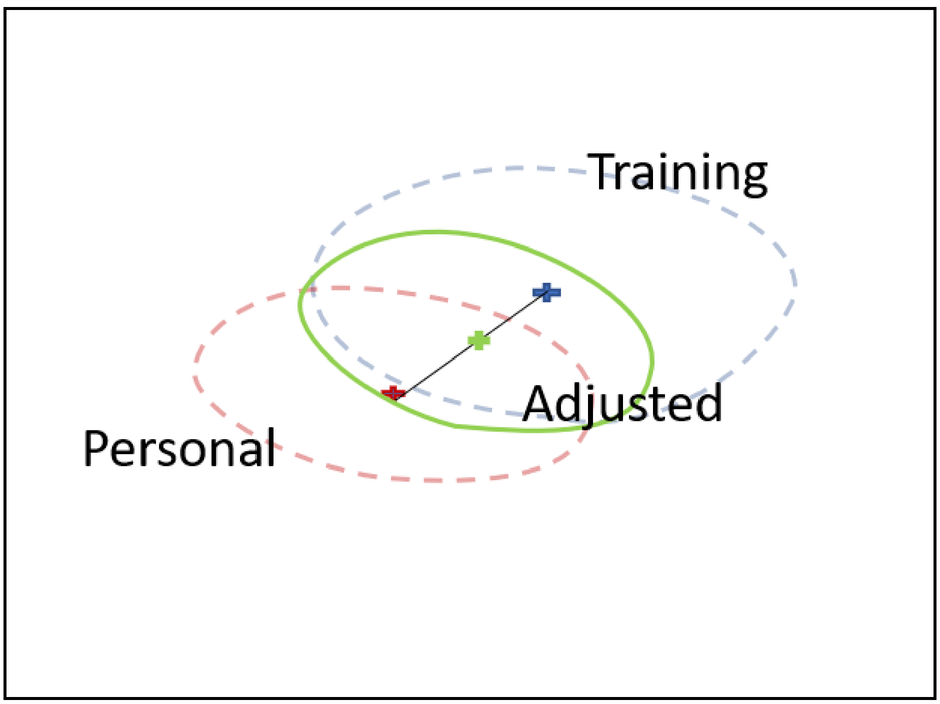


Figure 20. Adjusting Illustration

Figure 20 illustrates the idea for personal adjusting. If the personal data centroid is away from the training data centroid, the model will adjust the centroid from both and make the prediction more reliable. Moreover,

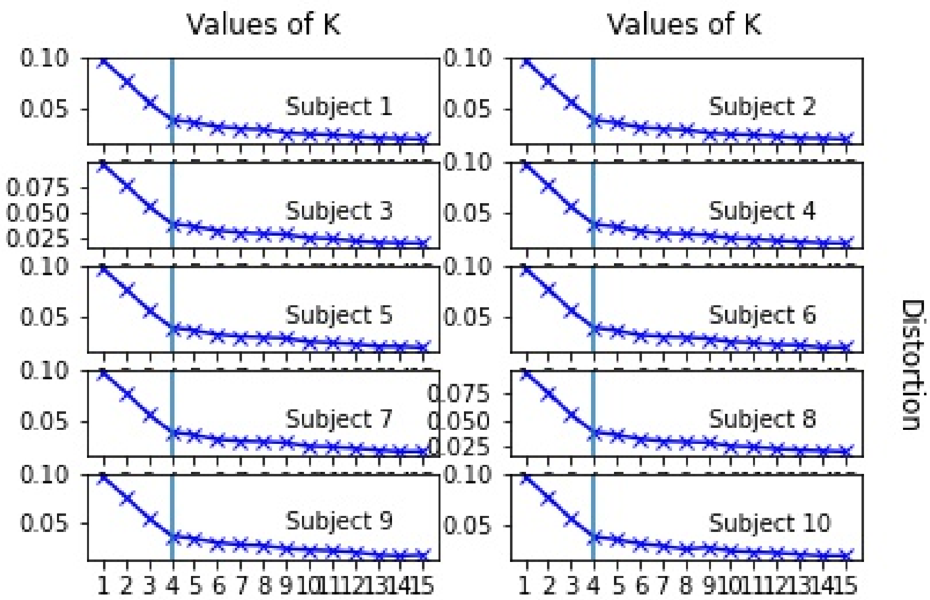


Figure 21. The Elbow Method Using Distortions for Data 1

the model can be updated with new activities with the updating dataset by creating a new class. In this case, we only need to save the center information in the system regardless the details from the training set.

5.3 Results

5.3.1 K-means Results

The experiments for Data 1 and Data 2 are subject-dependent, while experiments for Data 3 and Data 4 are subject-independent. The subject-dependent experiment includes the test subjects' information in the training, such as the Updating Set is included in Data 1 and Data 2. While subject-independent experiment is ones without any information from the test subjects, such as Data 3 and Data 4. The optimal number of clusters are different for each test subject. Figure 22 shows that the optimal number of clusters for each subject. However, the clusters are not accurate comparing to the true activity classes. For example, subject 2 only performed 2 activities, Walking and Jogging, the K-means gives 5 clusters. In this case, we could get the conclusion that even for the same person, the same activity data might not center well. And from Figure 21 and Figure 23 we see that the updating set has a small impact on the whole training data. And without Jogging, the k-means cluster method gives 5 optimal clusters which is exactly true number of activity classes.

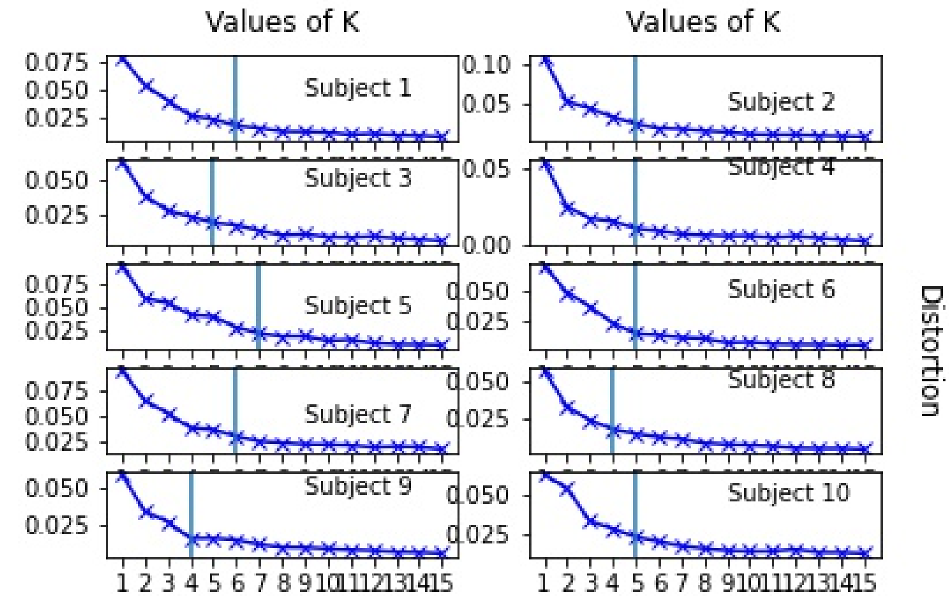


Figure 22. The Elbow Method Using Distortions for Data 2

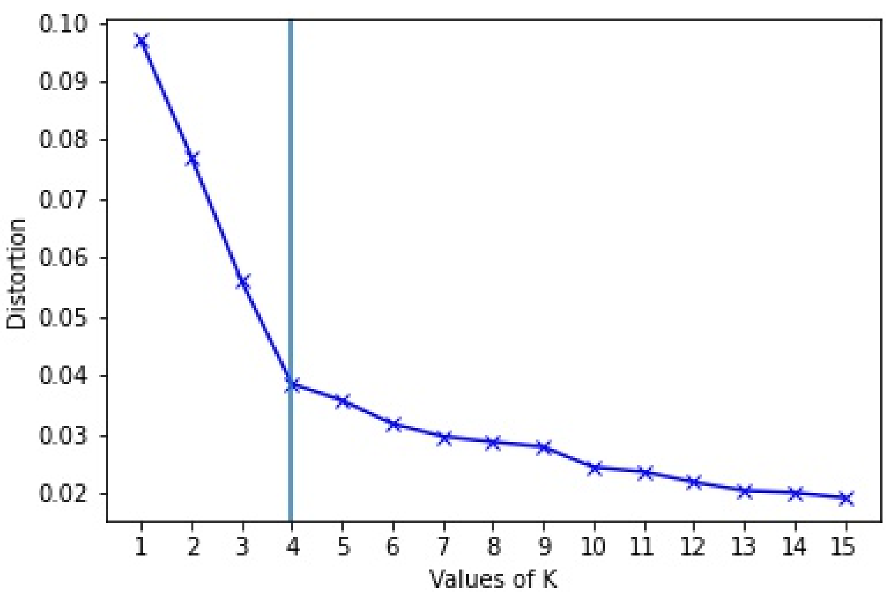


Figure 23. The Elbow Method Using Distortions for Data 3

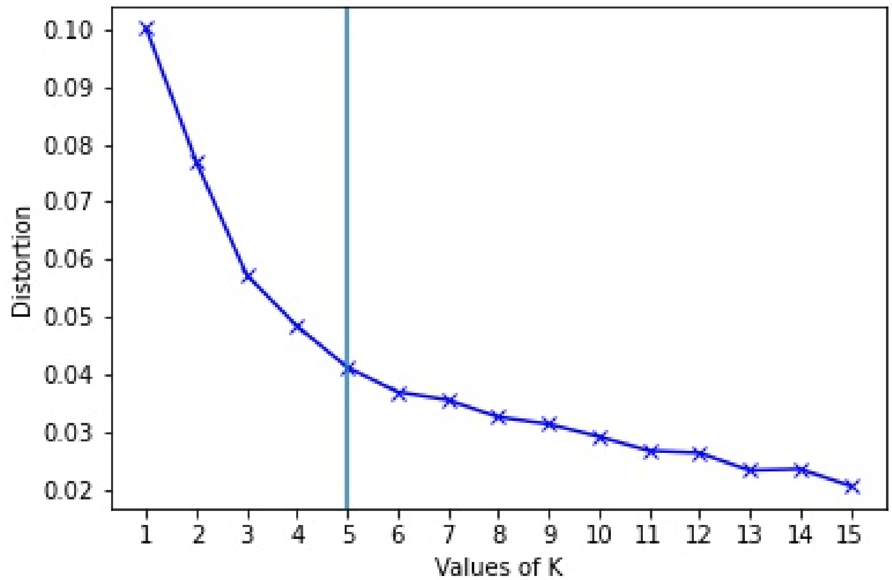


Figure 24. The Elbow Method Using Distortions for Data 4

5.3.2 Discriminant Analysis Results

In the previous step, K-means method gives the optimal number of classes without a specific activity name. Thus, in the step, we need to give the activity name to each cluster with major vote. The new class is determined by ”maximum value” in discriminant functions from the training parameters. The accuracy rates achieved by different datasets are listed in Table 23 and the error rates is shown in Figure 25. It is obvious that without adaptive step, Data 1, which includes the whole training data and 70% of personal data from the test subject, has lowest accuracy rate. This implies that the activity patterns are different from person to person, and even with some personal data, the model could not identify the activities accurately. This personal differences also could be found from Data 2, which only includes 70% of the same person’s data from the test but has the lowest error rates. For example, the accuracy rates for Subject 2 is improved from 61% to 97% with the personal data. The density centroids of activities from subject is different to the general training data. Thus adjustment step is necessary. The accuracy rates from Data 3 are after adjustment. The performances for Subject 2, 5, and 10 are as good as those from Data 2 or even better. Most of the results lose 2% in accuracy rate. By adding a new activity to the updating data, the performance of adjustment model for Data 4 is as good as for Data 3, and even has improvement for Subject 8.

For single activity precision rates, the big improvement from Data 1 to Data 2,3, and 4 is the performance on Upstairs and Downstairs. It is a big challenge to identify these two activities according to Table 24. In

Table 23 Average Accuracy Rates from Datasets

	Activities[1]	Data 1	Data 2	Data 3	Data 4
Subject 1	1,2,3,6	0.78	**0.95**	0.93	0.93
Subject 2	1, 6	0.61	0.97	**1.0**	**1.0**
Subject 3	1,2,3,4,5,6	0.72	**0.93**	0.91	0.91
Subject 4	1,2,3,6	0.83	**0.95**	0.93	0.93
Subject 5	1,2,3,4,5,6	0.57	**0.95**	**0.95**	**0.95**
Subject 6	1,2,3,6	0.56	**0.94**	0.91	0.91
Subject 7	1,2,3,4,5,6	0.75	**0.83**	0.80	0.80
Subject 8	1,2,3,6	0.76	**0.96**	0.92	**0.96**
Subject 9	1,2,3,5	0.67	**0.88**	0.85	0.85
Subject 10	1,2,3,4,5,6	0.82	**0.96**	0.98	0.98

[1] Activities column shows the activities performed by the subject, 1: walking, 2: Upstairs, 3: Downstairs, 4: Sitting, 5: Standing, 6: Jogging.

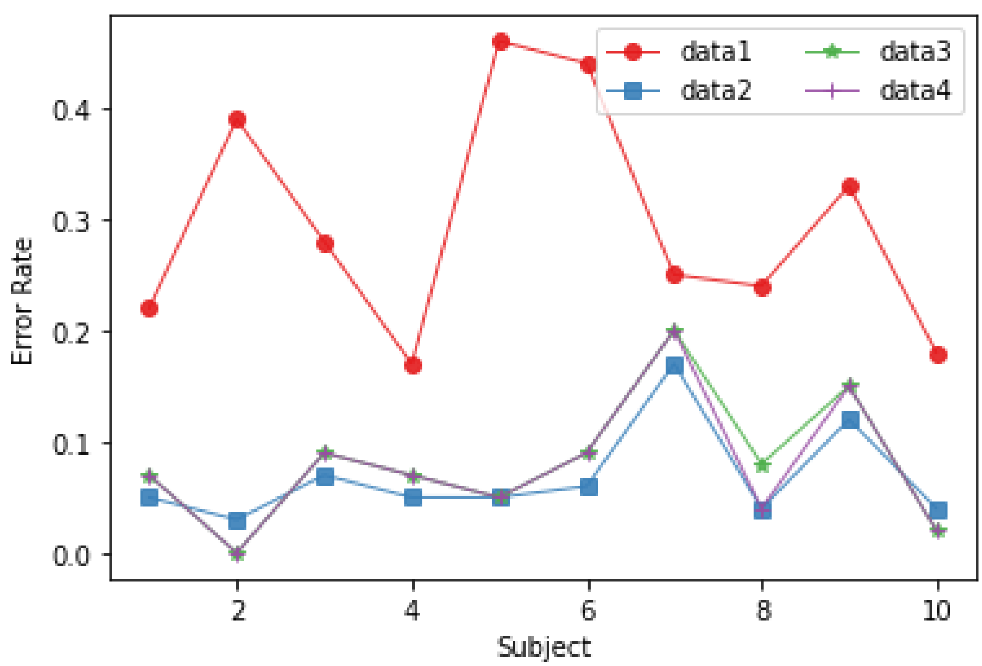


Figure 25. Error Rate for Datasets

Table 24 Precision Rates for Each Activity from Data 1

	Walking	Upstairs	Downstairs	Sitting	Standing	Jogging
Subject 1	0.70	0	0			0.89
Subject 2	0.57					1.0
Subject 3	0.65	0.50	0.5	1.0	0.5	1.0
Subject 4	0.76	0	0.5			1.0
Subject 5	0.45	0	0	1.0	1.0	0.86
Subject 6	0.43	0	0			1.0
Subject 7	0.53	0	0	1.0	1.0	0.90
Subject 8	0.65	0	0			1.0
Subject 9	0.77	0.25	0		1.0	
Subject 10	0.69	0	1.0	1.0	1.0	1.0

most subject cases, the precision rates are 0s. While, in other three Datasets,from Table 25, Table 26, and Table 27, the precision rates are much better, some of them have a high degree of accuracy, which is close to 100%.

Although the model with 70% personal data shows good performance for classification, the system with this model is fixed, means it could not be adaptive to new activities or to new activity patterns. The adaptive model solves this concern. The training set in Data 4 excludes the Jogging and the updating and verification sets in Data 5 is from different experiment and adding Kicking. The precision rates in Table 27 and **??** show that the model adjusts to the new activities very well. In most of the case, the model identifies the new activity with almost 100% accuracy.

5.4 Discussion

We notice that some of the precision rates has a low value, for example Subject 7 in Data 2 and Data 4, the precision rate for activities Sitting and Standing are 50% and 0, while in Data 1, they are close to 100%. The most possible reason might be because of the sample size limitation. Table 28 shows the number in the verification set, there is only 1 case for both activities (Sitting and Standing) in Subject 7, thus, they either 100% correct or wrong. It also implies that there might be only 2 or 3 cases/windows are used for updating, which means these activities might not being updated fully. Thus, the improvement for the model will need to consider the way to use small sample size to update the classifier. Meanwhile, the model based

Table 25 Precision Rates for Each Activity from Data 2

	Walking	Upstairs	Downstairs	Sitting	Standing	Jogging
Subject 1	1.0	1.0	0.71			1.0
Subject 2	1.0					0.95
Subject 3	1.0	0.75	1.0	1.0	0.5	1.0
Subject 4	1.0	1.0	0.67			1.0
Subject 5	1.0	0.88	0.88	1.0	1.0	1.0
Subject 6	1.0	0.89	0.83			1.0
Subject 7	0.75	1.0	0.75	0.50	0	0.90
Subject 8	0.91	1.0	1.0			1.0
Subject 9	1.0	0.75	0.57		1.0	
Subject 10	1.0	0.75	1.0	1.0	1.0	1.0

Table 26 Precision Rates for Each Activity from Data 3

	Walking	Upstairs	Downstairs	Sitting	Standing	Jogging
Subject 1	1.0	1.0	0.62			1.0
Subject 2	1.0					1.0
Subject 3	1.0	0.67	0.83	1.0	1.0	1.0
Subject 4	1.0	1.0	0.57			1.0
Subject 5	1.0	0.88	0.88	1.0	1.0	1.0
Subject 6	1.0	0.80	0.80			1.0
Subject 7	0.69	0	0.6	1.0	0	0.95
Subject 8	0.87	0.80	1.0			1.0
Subject 9	1.0	0.57	0.5		1.0	
Subject 10	1.0	0.86	1.0	1.0	1.0	1.0

Table 27 Precision Rates for Each Activity from Data 4

	Walking	Upstairs	Downstairs	Sitting	Standing	Jogging
Subject 1	1.0	1.0	0.62			1.0
Subject 2	1.0					1.0
Subject 3	1.0	0.75	0.83	1.0	0.5	1.0
Subject 4	1.0	1.0	0.57			1.0
Subject 5	1.0	0.88	0.88	1.0	1.0	1.0
Subject 6	1.0	0.80	0.80			1.0
Subject 7	0.75	0.50	0.67	0.5	0	0.9
Subject 8	0.91	1.0	1.0			1.0
Subject 9	1.0	0.57	0.5		1.0	
Subject 10	1.0	0.86	1.0	1.0	1.0	1.0

on the personal data from beginning give a better result for identifying the activities. This could imply that if the HAR system could collect customer's information at the beginning and create the classifier based on these collected data, the performance might be improved dramatically. For this reason, we think that add an interactive screen might be a good idea.

Table 28 Verification Observations in Each Activity

	Walking	Upstairs	Downstairs	Sitting	Standing	Jogging	Total
Subject 1	19	5	5			17	46
Subject 2	18					18	36
Subject 3	18	6	5	3	2	10	44
Subject 4	17	4	4			16	41
Subject 5	10	8	7	2	4	6	37
Subject 6	11	8	6			9	34
Subject 7	9	5	5	1	1	19	40
Subject 8	20	6	6			18	50
Subject 9	21	5	5		2		33
Subject 10	20	6	5	3	2	19	55

Chapter 6

Conclusion and Future Work

In previous chapters, We present several machine learning classifications and two streaming based adjusted activity recognition algorithms. All of these methods are used to apply on smartphone platform for human activity identification. We introduce some algorithms in Chapter 3, including SVM with different kernel functions, LDA, KNN, ANN, and a two layers LDA-SVM classifier. The basic machine learning algorithms give the best options for us to create the two layers classifier. For example, we select the LDA as the first layer classifier and select the Linear-SVM and Poly-SVM as the second layer classifier for dynamic and static, respectively. The reason is that the LDA gives an overall accuracy rate of 96.2% for UCI data, which is stable and one of the best output, the Linear-SVM gives 99%, 97%, and 97% for dynamic activities (walking, walking upstairs, and walking downstairs), respectively, and the Poly-SVM gives a better results for static activities (sitting, standing, and lying down). The two layers algorithm gives the highest overall accuracy rate 96.6%, with a high degree of accuracy rate for the first layer.

Considering the nature of the HAR system on wearable devices, we then introduce VFDT to mimic the streaming data incoming and adjust the model with these data. With the personal data updating, the VFDT has personal accuracy rates from 60.5% to 99.6%. The reason might because of the limit new cases. The results from Chapter 4 imply that the HAR system should be considered from person to person and the VFDT could classify the specific customer's activities for almost 100% certain. Because of these conclusion, we then proposed a combination algorithm with unsupervised and supervised methods. The largest advantage of this propose method is that it is possible to add new activities into the system, every customer could have his/her own specific HAR system. When apply to the WISDM data, the personal accuracy rates are from 80% to high degree, even with a new activity, which was not included in the first training process. Meanwhile, it also shows that the updating process only requires small amount of the data, which is a beneficial for us to apply this model to the real world. Also, these personalized model could adjust the classifiers by time goes. It is very useful in practice. For example, after several years, the customer has health issues, which might change his/her activity patterns, based on the built-in functions, the HAR system could possibly identify these activities with high error. But with our adjusted system, these classifiers could gradually update to the new recognition setting, which was adjusted by the customer's current activity pattern. This could have a better performance.

The future work will mainly focus on the feature transformation, a personalized adjusted HAR system, and the multiple sensor combination. In the previous chapters, we do not include any sensor data transformation method. We apply all the methods to the transformed datasets, UCI and WISDM, which only include the statistical features, such as means, standard deviations, correlations, etc. There are many other method, such as time and energy domain. Besides, the window drift and flexible window size methods are also interesting directions. The personalized HAR system is the future trend from our results, thus, have a simple algorithm with low power and small space requirement will be the next step. Also, with the development of the technology, it is also possible to build more sensors in smartphone and smartwatch, such as heart rate recorder and environment detector. The multiple sensor combination will be the big issue for creating model.

References

[1] Uci machine learning web.

[2] Abidine, B. M., Fergani, L., Fergani, B., and Oussalah, M. The joint use of sequence features combination and modified weighted svm for improving daily activity recognition. *Pattern Analysis and Applications 21*, 1 (2018), 119–138.

[3] Ahad, M. A. R., Tan, J., Kim, H., and Ishikawa, S. Human activity recognition: Various paradigms. In *2008 international conference on control, automation and systems* (2008), IEEE, pp. 1896–1901.

[4] Al Jeroudi, Y., Ali, M., Latief, M., and Akmeliawati, R. Online sequential extreme learning machine algorithm based human activity recognition using inertial data. In *2015 10th Asian Control Conference (ASCC)* (2015), IEEE, pp. 1–6.

[5] Alpaydin, E. *Introduction to machine learning*. MIT press, 2020.

[6] Alpaydin, E. *Introduction to machine learning*. MIT press, 2020.

[7] Alsheikh, M. A., Selim, A., Niyato, D., Doyle, L., Lin, S., and Tan, H.-P. Deep activity recognition models with triaxial accelerometers. In *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence* (2016).

[8] Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician 46*, 3 (1992), 175–185.

[9] Altun, K., Barshan, B., and Tunçel, O. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognition 43*, 10 (2010), 3605–3620.

[10] Anguita, D., Ghio, A., Oneto, L., Parra, X., and Reyes-Ortiz, J. L. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *International workshop on ambient assisted living* (2012), Springer, pp. 216–223.

[11] ATTAL, F., MOHAMMED, S., DEDABRISHVILI, M., CHAMROUKHI, F., OUKHELLOU, L., AND AMIRAT, Y. Physical human activity recognition using wearable sensors. *Sensors 15*, 12 (2015), 31314–31338.

[12] BALABKA, D. Semi-supervised learning for human activity recognition using adversarial autoencoders. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (2019), pp. 685–688.

[13] BERCHTOLD, M., BUDDE, M., GORDON, D., SCHMIDTKE, H. R., AND BEIGL, M. Actiserv: Activity recognition service for mobile phones. In *International symposium on wearable computers (ISWC) 2010* (2010), IEEE, pp. 1–8.

[14] BETTINI, C., CIVITARESE, G., AND PRESOTTO, R. Personalized semi-supervised federated learning for human activity recognition. *arXiv preprint arXiv:2104.08094* (2021).

[15] BLASCO, R., MARCO, Á., CASAS, R., CIRUJANO, D., AND PICKING, R. A smart kitchen for ambient assisted living. *Sensors 14*, 1 (2014), 1629–1653.

[16] BOULOS, M. N. K., ROCHA, A., MARTINS, A., VICENTE, M. E., BOLZ, A., FELD, R., TCHOUDOVSKI, I., BRAECKLEIN, M., NELSON, J., LAIGHIN, G. Ó., ET AL. Caalyx: a new generation of location-based services in healthcare. *International journal of health geographics 6*, 1 (2007), 1–6.

[17] BRÉMOND, F., THONNAT, M., AND ZÚNIGA, M. Video-understanding framework for automatic behavior recognition. *Behavior Research Methods 38*, 3 (2006), 416–426.

[18] BULLING, A., BLANKE, U., AND SCHIELE, B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR) 46*, 3 (2014), 1–33.

[19] CARDOSO, H. L., AND MENDES-MOREIRA, J. Improving human activity classification through online semi-supervised learning. In *STREAMEVOLV@ ECML-PKDD* (2016).

[20] CHANG, M.-C., KRAHNSTOEVER, N., LIM, S., AND YU, T. Group level activity recognition in crowded environments across multiple cameras. In *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance* (2010), IEEE, pp. 56–63.

[21] Chen, Y., and Xue, Y. A deep learning approach to human activity recognition based on single accelerometer. In *2015 ieee international conference on systems, man, and cybernetics* (2015), IEEE, pp. 1488–1492.

[22] Cheng, L., Guan, Y., Zhu, K., and Li, Y. Recognition of human activities using machine learning methods with wearable sensors. In *2017 IEEE 7th annual computing and communication workshop and conference (CCWC)* (2017), IEEE, pp. 1–7.

[23] Chernbumroong, S., Cang, S., Atkins, A., and Yu, H. Elderly activities recognition and classification for applications in assisted living. *Expert Systems with Applications 40*, 5 (2013), 1662–1674.

[24] Cohen, L., Avrahami-Bakish, G., Last, M., Kandel, A., and Kipersztok, O. Real-time data mining of non-stationary data streams from sensor networks. *Information Fusion 9*, 3 (2008), 344–353.

[25] Cortes, C., and Vapnik, V. Support-vector networks. *Machine learning 20*, 3 (1995), 273–297.

[26] Demiris, G., Hensel, B. K., Skubic, M., and Rantz, M. Senior residents' perceived need of and preferences for "smart home" sensor technologies. *International journal of technology assessment in health care 24*, 1 (2008), 120–124.

[27] Desai, S., Roy, S., Patel, B., Purandare, S., and Kucheria, M. Very fast decision tree (vfdt) algorithm on hadoop. In *2016 International Conference on Computing Communication Control and automation (ICCUBEA)* (2016), IEEE, pp. 1–7.

[28] Dohr, A., Modre-Opsrian, R., Drobics, M., Hayn, D., and Schreier, G. The internet of things for ambient assisted living. In *2010 seventh international conference on information technology: new generations* (2010), Ieee, pp. 804–809.

[29] Domingos, P., and Hulten, G. Mining high-speed data streams. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining* (2000), pp. 71–80.

[30] Erfani, S. M., Rajasegarar, S., Karunasekera, S., and Leckie, C. High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning. *Pattern Recognition 58* (2016), 121–134.

[31] FERGANI, B., ET AL. Evaluating a new classification method using pca to human activity recognition. In *2013 International Conference on Computer Medical Applications (ICCMA)* (2013), IEEE, pp. 1–4.

[32] FOERSTER, F., SMEJA, M., AND FAHRENBERG, J. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. *Computers in human behavior 15*, 5 (1999), 571–583.

[33] FRIEDMAN, J., HASTIE, T., TIBSHIRANI, R., ET AL. *The elements of statistical learning*, vol. 1. Springer series in statistics New York, 2001.

[34] FUSIER, F., VALENTIN, V., BRÉMOND, F., THONNAT, M., BORG, M., THIRDE, D., AND FERRYMAN, J. Video understanding for complex activity recognition. *Machine Vision and Applications 18*, 3 (2007), 167–188.

[35] GABER, M. M., KRISHNASWAMY, S., AND ZASLAVSKY, A. On-board mining of data streams in sensor networks. In *Advanced methods for knowledge discovery from complex data*. Springer, 2005, pp. 307–335.

[36] GAMA, J., AND PEDERSEN, R. U. Predictive learning in sensor networks. In *Learning from Data Streams*. Springer, 2007, pp. 143–164.

[37] GANNAPATHY, V. R., IBRAHIM, A., ZAKARIA, Z. B., OTHMAN, A. R. B., AND LATIFF, A. A. Zigbee-based smart fall detection and notification system with wearable sensor (e-safe). *Int J Res Eng Technol 2*, 8 (2013), 337–344.

[38] GARCIA, K. D., CARVALHO, T., MENDES-MOREIRA, J., CARDOSO, J. M., AND DE CARVALHO, A. C. A preliminary study on hyperparameter configuration for human activity recognition. *arXiv preprint arXiv:1810.10956* (2018).

[39] GARCIA, K. D., DE SÁ, C. R., POEL, M., CARVALHO, T., MENDES-MOREIRA, J., CARDOSO, J. M., DE CARVALHO, A. C., AND KOK, J. N. An ensemble of autonomous auto-encoders for human activity recognition. *Neurocomputing 439* (2021), 271–280.

[40] GOLDSTONE, J. A. The new population bomb: the four megatrends that will change the world. *Foreign Aff. 89* (2010), 31.

[41] GUAN, D., YUAN, W., LEE, Y.-K., GAVRILOV, A., AND LEE, S. Activity recognition based on semi-supervised learning. In *13th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA 2007)* (2007), IEEE, pp. 469–475.

[42] GUO, Q., LIU, B., AND CHEN, C. W. A two-layer and multi-strategy framework for human activity recognition using smartphone. In *2016 Ieee International Conference on Communications (Icc)* (2016), IEEE, pp. 1–6.

[43] GYŐRBÍRÓ, N., FÁBIÁN, Á., AND HOMÁNYI, G. An activity recognition system for mobile phones. *Mobile Networks and Applications 14*, 1 (2009), 82–91.

[44] HAGHI, M., THUROW, K., AND STOLL, R. Wearable devices in medical internet of things: scientific research and commercially available devices. *Healthcare informatics research 23*, 1 (2017), 4–15.

[45] HASSAN, M. M., UDDIN, M. Z., MOHAMED, A., AND ALMOGREN, A. A robust human activity recognition system using smartphone sensors and deep learning. *Future Generation Computer Systems 81* (2018), 307–313.

[46] HE, W., GUO, Y., GAO, C., AND LI, X. Recognition of human activities with wearable sensors. *EURASIP Journal on Advances in Signal Processing 2012*, 1 (2012), 1–13.

[47] HE, Z., AND JIN, L. Activity recognition from acceleration data based on discrete consine transform and svm. In *2009 IEEE International Conference on Systems, Man and Cybernetics* (2009), IEEE, pp. 5041–5044.

[48] HECKERT, N. A., FILLIBEN, J. J., CROARKIN, C. M., HEMBREE, B., GUTHRIE, W. F., TOBIAS, P., PRINZ, J., ET AL. Handbook 151: Nist/sematech e-handbook of statistical methods.

[49] HO, T. K. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (1995), vol. 1, IEEE, pp. 278–282.

[50] HULTEN, G., SPENCER, L., AND DOMINGOS, P. Mining time-changing data streams. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining* (2001), pp. 97–106.

[51] IGNATOV, A. Real-time human activity recognition from accelerometer data using convolutional neural networks. *Applied Soft Computing 62* (2018), 915–922.

[52] INCEL, O. D. Analysis of movement, orientation and rotation-based sensing for phone placement recognition. *Sensors 15*, 10 (2015), 25474–25506.

[53] JIANG, W., AND YIN, Z. Human activity recognition using wearable sensors by deep convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia* (2015), pp. 1307–1310.

[54] KAUTZ, T., GROH, B. H., HANNINK, J., JENSEN, U., STRUBBERG, H., AND ESKOFIER, B. M. Activity recognition in beach volleyball using a deep convolutional neural network. *Data Mining and Knowledge Discovery 31*, 6 (2017), 1678–1705.

[55] KWAPISZ, J. R., WEISS, G. M., AND MOORE, S. A. Cell phone-based biometric identification. In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)* (2010), IEEE, pp. 1–7.

[56] KWAPISZ, J. R., WEISS, G. M., AND MOORE, S. A. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter 12*, 2 (2011), 74–82.

[57] LANSANGAN, J. R. G., AND BARRIOS, E. B. Principal components analysis of nonstationary time series data. *Statistics and Computing 19*, 2 (2009), 173.

[58] LARA, O. D., AND LABRADOR, M. A. A mobile platform for real-time human activity recognition. In *2012 IEEE consumer communications and networking conference (CCNC)* (2012), IEEE, pp. 667–671.

[59] LARA, O. D., AND LABRADOR, M. A. A survey on human activity recognition using wearable sensors. *IEEE communications surveys & tutorials 15*, 3 (2012), 1192–1209.

[60] LI, F., SHIRAHAMA, K., NISAR, M. A., KÖPING, L., AND GRZEGORZEK, M. Comparison of feature learning methods for human activity recognition using wearable sensors. *Sensors 18*, 2 (2018), 679.

[61] LIU, G., LIANG, J., LAN, G., HAO, Q., AND CHEN, M. Convolution neutral network enhanced binary sensor network for human activity recognition. In *2016 IEEE SENSORS* (2016), IEEE, pp. 1–3.

[62] LOH, W.-Y. Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery 1*, 1 (2011), 14–23.

[63] MA, Y., AND GHASEMZADEH, H. Labelforest: Non-parametric semi-supervised learning for activity recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence* (2019), vol. 33, pp. 4520–4527.

[64] MAURER, U., SMAILAGIC, A., SIEWIOREK, D. P., AND DEISHER, M. Activity recognition and monitoring using multiple sensors on different body positions. In *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)* (2006), IEEE, pp. 4–pp.

[65] MINEGISHI, T., AND NIIMI, A. Detection of fraud use of credit card by extended vfdt. In *2011 World Congress on Internet Security (WorldCIS-2011)* (2011), IEEE, pp. 152–159.

[66] MINEGISHI, T., AND NIIMI, A. 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), 229–235.

[67] MOESLUND, T. B., HILTON, A., AND KRÜGER, V. A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding 104*, 2-3 (2006), 90–126.

[68] NA, S., RAMACHANDRAN, K. M., JI, M., AND TU, Y. Real time activity recognition using smartphone accelerometer. *International Journal of Trend in Scientific Research and Development* (2016), 533–542.

[69] NAKANO, K., AND CHAKRABORTY, B. Effect of dynamic feature for human activity recognition using smartphone sensors. In *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)* (2017), IEEE, pp. 539–543.

[70] OH, S., ASHIQUZZAMAN, A., LEE, D., KIM, Y., AND KIM, J. Study on human activity recognition using semi-supervised active transfer learning. *Sensors 21*, 8 (2021), 2760.

[71] OKEYO, G., CHEN, L., WANG, H., AND STERRITT, R. Dynamic sensor data segmentation for real-time knowledge-driven activity recognition. *Pervasive and Mobile Computing 10* (2014), 155–172.

[72] ORDÓÑEZ, F. J., AND ROGGEN, D. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors 16*, 1 (2016), 115.

[73] PLÖTZ, T., HAMMERLA, N. Y., AND OLIVIER, P. L. Feature learning for activity recognition in ubiquitous computing. In *Twenty-second international joint conference on artificial intelligence* (2011).

[74] RAILEANU, L. E., AND STOFFEL, K. Theoretical comparison between the gini index and information gain criteria. In *Annals of Mathematics and Artificial Intelligence* (2004), pp. 77–93.

[75] RASHIDI, P., COOK, D. J., HOLDER, L. B., AND SCHMITTER-EDGECOMBE, M. Discovering activities to recognize and track in a smart environment. *IEEE transactions on knowledge and data engineering 23*, 4 (2010), 527–539.

[76] RAVI, D., WONG, C., LO, B., AND YANG, G.-Z. A deep learning approach to on-node sensor data analytics for mobile or wearable devices. *IEEE journal of biomedical and health informatics 21*, 1 (2016), 56–64.

[77] RONAO, C. A., AND CHO, S.-B. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert systems with applications 59* (2016), 235–244.

[78] RONAOO, C. A., AND CHO, S.-B. Evaluation of deep convolutional neural network architectures for human activity recognition with smartphone sensors. *KIISE Korea Computer Congress* (2015), 858–860.

[79] ROSER, M. Fertility rate. *Our World in Data* (2014). https://ourworldindata.org/fertility-rate.

[80] SHARMA, A., LEE, Y.-D., AND CHUNG, W.-Y. High accuracy human activity monitoring using neural network. In *2008 third international conference on convergence and hybrid information technology* (2008), vol. 1, IEEE, pp. 430–435.

[81] SOHN, W., AND KHAN, Z. Real time human activity recognition system based on radon transform. In *IJCA special issue on artificial intelligence techniques novel approaches practical applications," AIT* (2011), Citeseer.

[82] SUBRAHMANYAM, M., AND VENKATESWARA, R. Vfdt algorithm for decision tree generation. *International Journal for Development of Computer Science and Technology (IJDCST) 1*, 7 (2013).

[83] TAPIA, E. M., INTILLE, S. S., HASKELL, W., LARSON, K., WRIGHT, J., KING, A., AND FRIEDMAN, R. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *2007 11th IEEE international symposium on wearable computers* (2007), IEEE, pp. 37–40.

[84] TUNCA, C., ALEMDAR, H., ERTAN, H., INCEL, O. D., AND ERSOY, C. Multimodal wireless sensor network-based ambient assisted living in real homes with multiple residents. *Sensors 14*, 6 (2014), 9692–9719.

[85] VAN KASTEREN, T., ENGLEBIENNE, G., AND KRÖSE, B. J. An activity monitoring system for elderly care using generative and discriminative models. *Personal and ubiquitous computing 14*, 6 (2010), 489–498.

[86] VAN LAERHOVEN, K., AND GELLERSEN, H.-W. Spine versus porcupine: A study in distributed wearable activity recognition. In *Eighth International Symposium on Wearable Computers* (2004), vol. 1, IEEE, pp. 142–149.

[87] VOLLMER, C., GROSS, H.-M., AND EGGERT, J. P. Learning features for activity recognition with shift-invariant sparse coding. In *International conference on artificial neural networks* (2013), Springer, pp. 367–374.

[88] WANG, K., HE, J., AND ZHANG, L. Attention-based convolutional neural network for weakly labeled human activities' recognition with wearable sensors. *IEEE Sensors Journal 19*, 17 (2019), 7598–7604.

[89] WANG, L. Recognition of human activities using continuous autoencoders with wearable sensors. *Sensors 16*, 2 (2016), 189.

[90] WANG, L., GU, T., TAO, X., AND LU, J. A hierarchical approach to real-time activity recognition in body sensor networks. *Pervasive and Mobile Computing 8*, 1 (2012), 115–130.

[91] WANG, Z., WU, D., CHEN, J., GHONEIM, A., AND HOSSAIN, M. A. A triaxial accelerometer-based human activity recognition via eemd-based features and game-theory-based feature selection. *IEEE Sensors Journal 16*, 9 (2016), 3198–3207.

[92] WU, C., KHALILI, A. H., AND AGHAJAN, H. Multiview activity recognition in smart homes with spatio-temporal features. In *Proceedings of the fourth ACM/IEEE international conference on distributed smart cameras* (2010), pp. 142–149.

[93] WU, W., DASGUPTA, S., RAMIREZ, E. E., PETERSON, C., NORMAN, G. J., ET AL. Classification accuracies of physical activities using smartphone motion sensors. *Journal of medical Internet research 14*, 5 (2012), e2208.

[94] YAO, R., LIN, G., SHI, Q., AND RANASINGHE, D. C. Efficient dense labelling of human activity sequences from wearables using fully convolutional networks. *Pattern Recognition 78* (2018), 252–266.

[95] YAO, S., HU, S., ZHAO, Y., ZHANG, A., AND ABDELZAHER, T. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In *Proceedings of the 26th International Conference on World Wide Web* (2017), pp. 351–360.

[96] YE, J., JANARDAN, R., AND LI, Q. Two-dimensional linear discriminant analysis. *Advances in neural information processing systems 17* (2004), 1569–1576.

[97] ZDRAVEVSKI, E., LAMESKI, P., TRAJKOVIK, V., KULAKOV, A., CHORBEV, I., GOLEVA, R., POMBO, N., AND GARCIA, N. Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering. *Ieee Access 5* (2017), 5262–5280.

[98] ZENG, M., NGUYEN, L. T., YU, B., MENGSHOEL, O. J., ZHU, J., WU, P., AND ZHANG, J. Convolutional neural networks for human activity recognition using mobile sensors. In *6th International Conference on Mobile Computing, Applications and Services* (2014), IEEE, pp. 197–205.

[99] ZENG, M., YU, T., WANG, X., NGUYEN, L. T., MENGSHOEL, O. J., AND LANE, I. Semi-supervised convolutional neural networks for human activity recognition. In *2017 IEEE International Conference on Big Data (Big Data)* (2017), IEEE, pp. 522–529.

[100] ZHANG, L., WU, X., AND LUO, D. Human activity recognition with hmm-dnn model. In *2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)* (2015), IEEE, pp. 192–197.

[101] ZHANG, M., AND SAWCHUK, A. A. Human daily activity recognition with sparse representation using wearable sensors. *IEEE journal of Biomedical and Health Informatics 17*, 3 (2013), 553–560.

[102] ZHANG, Y., AND RAMACHANDRAN, K. M. Human activity recognition with streaming smartphone data. In *2019 Global Conference for Advancement in Technology (GCAT)* (2019), IEEE, pp. 1–6.

[103] ZHANG, Y., AND RAMACHANDRAN, K. M. Offline machine learning for human activity recognition with smartphone. IEEE.

[104] ZHU, C., AND SHENG, W. Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 41*, 3 (2011), 569–573.

[105] ZOUBA, N., BREMOND, F., THONNAT, M., AND VU, V. T. Multi-sensors analysis for everyday activity monitoring. *Proc. of SETIT* (2007), 25–29.

Appendix A

Copyright Permission

The following Figure A.1 is the permission proof for Chapter 4.

IEEE Requesting permission to reuse content from an IEEE publication

Human Activity Recognition with Streaming Smartphone Data

Conference Proceedings: 2019 Global Conference for Advancement in Technology (GCAT)

Author: Yanjia Zhang

Publisher: IEEE

Date: Oct. 2019

Copyright © 2019, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK CLOSE WINDOW

Figure A.1. IEEE Permission