IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 20, 2019, accepted June 10, 2019, date of publication June 14, 2019, date of current version July 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2922974

A Comprehensive Study of Smartphone-Based Indoor Activity Recognition via Xgboost

WENTING ZHANG, XIAOHUI ZHAO, (Member, IEEE), AND ZAN LI, (Member, IEEE)
College of Communication Engineering, Jilin University, Changchun 130012, China

Corresponding author: Zan Li (zanli@jlu.edu.cn)

This work was supported by the Natural Science Foundation of China under Grant 61801189.

ABSTRACT Recently, multi-floor indoor positioning has become increasingly interesting for researchers, in which accurate recognition of indoor activities is critical for the detection of floor changes and the improvement of positioning accuracy according to indoor landmarks. However, we have not found a comprehensive study for recognizing indoor activities related to multi-floor indoor positioning based on a robust machine learning algorithm. In this work, we propose a framework for recognizing five indoor activities, i.e., walking, stillness, stair climbing, escalator, or elevator taking. In this framework, we investigate the relevant sensors and features to improve the recognition accuracy of these activities, especially some specific features in the frequency domain and wavelet domain. We propose to utilize a promising tree-based ensemble learning classifier, XGBoost, to recognize these activities. Based on our dataset created by 40 volunteers, we provide a comprehensive analysis of the proposed framework for indoor activity recognition. Considering both accuracy and computational cost, the XGBoost-based indoor activity recognition algorithm outperforms the other ensemble learning classifiers and single classifiers, and the average recognition F-score of XGBoost reaches 84.41%. In addition, our introduced specific features in the frequency domain and wavelet domain can significantly improve the recognition accuracy. Moreover, we use a publicly available dataset to verify our proposed framework and XGBoost classifier reaches 84.19% that outperforms the other classifiers.

INDEX TERMS Activity recognition, ensemble learning, XGBoost, smartphone.

I. INTRODUCTION

In recent years, indoor positioning has become an emerging research topic and received a lot of attentions due to its huge commercial values. Much research has been conducted to improve positioning accuracy on single floor. For example, WiFi based indoor positioning is often integrated with pedestrian dead reckoning (PDR) to get accurate single-floor indoor positioning, in which the PDR algorithm calculates the user's current position by using a previously determined position based on inertial sensors [1]. However, most buildings in cities are with multiple floors, the traditional PDR algorithm [2] can not locate the users on multiple floors. Hence, in these buildings, detection of floor numbers is a critical research issue to achieve multi-floor indoor positioning. On the other hand, the crowdsensing of indoor walking paths based on crowdsourcing PDR traces have recently become interesting for researchers because it can eliminate the effort of site surveying in fingerprinting based indoor positioning [3]. However, at present, the research on the crowdsensing of walking paths is mainly conducted on single floor [3]–[6]. To achieve multiple-floor crowdsensing of walking paths, we need to further cluster the traces based on their floor numbers [7]. In order to address the aforementioned problems in the multiple-floor indoor positioning and the crowdsensing of walking paths, human activity recognition (HAR) technique can be used to detect the changes of floor number by recognizing and classifying different kinds of indoor activities of users, such as walking, stair climbing, and elevator taking, based on inertial sensor data. In MPiLoc [7], the authors propose to leverage barometer to detect floor changing then partition the user traces into multiple floors. Although this method can detect the floor changes, it can not recognize different types of activities. In ALIMC [4], barometer and accelerometer are adopted to detect the activities of stair climbing and elevator taking for crowdsensing of walking paths. However, the authors do not provide and evaluate their HAR methods explicitly. In iMap [6], the authors design a rule-based algorithm based on barometer and accelerometer to recognize indoor activities of stair climbing, escalator or

The associate editor coordinating the review of this manuscript and approving it for publication was Datong Liu.

elevator taking. Although they claim a high accuracy of the algorithm, the performance of the rule-based algorithm significantly relies on the empirical parameters and is normally not robust to the changes of users and environments. To the best of our knowledge, there is still no robust solution for recognizing indoor activities, i.e., walking, stillness, stair climbing, escalator or elevator taking, which is critical for the multi-floor indoor positioning. Additionally, much work has been conducted in HAR to investigate different user activities, such as falling detection and risk assessment for elderly people [8], assisting security personnel to monitor surrounding environment and detect suspicious activities [9] and inferring transportation modes [10].

In the early stage of HAR, most researches utilize wearable sensors placed in various parts of human body leading to complicated data collection [11] and demanding an extra cost for hardware [12]. In recent years, smartphones become more and more popular in peoples' daily life, which integrate various embedded sensors such as accelerometer, gyroscope, and barometer for diverse applications. Compared with the wearable devices, smartphones are easy to use, available and portable. Therefore, we can use the embedded sensors in a smartphone to conduct HAR due to its convenience, low cost, and ubiquitousness. In embedded sensors, accelerometer and gyroscope in a smartphone are often used to detect common activities such as walking, standing, jogging, etc. [13], [14], which can achieve high recognition accuracy. However, indoor activities are often conducted in three dimensions, which could be difficult to reach high recognition accuracy when we only rely on accelerometer and gyroscope. Fortunately, barometer can measure altitude changes. Hence, for indoor activity recognition especially for the detection of floor changes, barometer can help us significantly enhance the performance of HAR [15], [16]. Besides the types of sensors used for HAR, the applied features also play an important role in the accuracy of HAR and highly correlated with the activities to be analyzed. The most commonly-used features in time domain are mean, variance, minimum, maximum and median for HAR [17]. However, the frequency domain and wavelet domain features are rarely investigated for the recognition of indoor activities.

On the other hand, many machine learning algorithms have been used in HAR, such as Support Vector Machine (SVM) [18], *K*-nearest Neighbor (KNN) [19], Hidden Markov Model (HMMs) [20] and Multi-Layer Perception (MLP). However, it is still challenging to achieve high recognition accuracy by using these machine learning algorithms with single classifiers. To address this problem, ensemble learning which accomplishes learning tasks by combining multiple learners has played a key role in HAR [21]. This ensemble learning often achieves generalized performance that is significantly superior to a single learner. In ensemble learning classifiers, Extreme Gradient Boosting (XGBoost) has recently been proposed as an emerging ensemble learning approach and it can achieve excellent performance in many applications, e.g., web text classification, behavior prediction and malware classification. Compared with the other classifiers, XGBoost has many advantages such as high efficiency, low computational cost, supporting parallelization, robustness to overfitting, etc. It has also made excellent results in data analysis and data mining competitions in recent years.

In this work, we present a framework based on XGBoost to recognize five kinds of indoor human activities, i.e., walking, stillness, stair climbing, escalator or elevator taking to help multi-floor indoor positioning. The main contributions of this work are summarized as follows:

- We build a comprehensive indoor activity dataset collected from 40 subjects based on the aforementioned five indoor activities and four smartphone placement locations, i.e., in trouser pocket, in bag, in hand and holding horizontally. The total length of the dataset is 5.16 million samples collected in 2,400 minutes. Compared with the other datasets, our dataset is the first complete set for indoor activities related to floor change.
- We apply XGBoost to achieve high recognition accuracy of indoor activities. We optimally adapt the parameters in a XGBoost model based on ten-fold cross validation to obtain a robust and accurate model. Note that although XGBoost has shown its excellent capability in many applications, we find little work of using XGBoost in activity recognition based on smartphone sensors, especially in the domain of indoor activities related to floor change.
- Based on the our dataset, we provide a comprehensive analysis of our proposed XGBoost-based indoor activity recognition algorithm and compare it with the other two tree-based ensemble learning classifiers, random forest and GBDT. Additionally, we also compare the performance of the ensemble learning classifiers and the single classifiers. The results show that the ensemble learning classifiers outperform the single classifiers and XGBoost performs the best.
- Moreover, we explore diverse sensors and features in frequency domain and wavelet domain, especially the spectrum entropy and the wavelet energy related to these indoor activities to improve recognition accuracy. Besides the commonly used accelerometer and gyroscope for HAR, we further leverage barometer data in our algorithm to better distinguish the altitude changes for different activities.

The rest of the paper is organized as follows: Section II presents the latest development and research of HAR. Section III gives the preliminaries to ensemble learning based on tree structure. Section IV introduces our main contribution about our XGBoost based indoor activity recognition algorithm. Section V describes the implementation of our system for indoor activity recognition. Section VI provides the experimental results and analysis. Section VII concludes the paper.

II. RELATED WORK

Researchers have been studying HAR for many years, which mainly has two directions: vision-based activity recognition and sensor-based activity recognition. The vision-based activity recognition monitors the subjects in real time through cameras and uses image analysis technology to recognize the activities of subjects [22]–[24]. Although this vision-based activity recognition has shown ideal results in terms of complexity and accuracy, its performance depends on external environments. Particularly it requires sufficient light and background conditions, and it may invade the privacy of subjects. On the other hand, the HAR based on wearable sensors has become popular in recent years due to its robustness to external environments and less privacy issues. However, wearable sensors require extra sensing components which are inconvenient for users. To deal with aforementioned problem, researchers begin to use embedded sensors in smartphone for HAR.

Nowadays, the applications of smartphone become more popular, and many sensors such as accelerometer, gyroscope, magnetometer, and barometer are embedded in a smartphone. Considering portability and ubiquitousness of smartphone, the research for HAR based on smartphone sensors has been conducted. For instance, Jain and Kanhangad [25] use accelerometer and gyroscope sensor data to identify the subject activities. Furthermore, Vaizman *et al.* [26] collect 300,000 minutes sensor data from 60 subjects, using accelerometer, gyroscope, and location, audio, and phone-state sensors to improve context recognition. Considering indoor activity recognition for multi-floor indoor positioning, some people use barometer to recognize activities [15], [16]. However, in these studies, barometer data is analyzed based on the rule-based algorithms. They do not deeply investigate the relevant features in the barometer data related to the activities.

The extracted features are normally application oriented. For different activities, we need to select different relevant features. Xu *et al.* [17] select time domain features in motion signals such as mean, variance, zero crossing rate, etc., to represent human activities. In order to detect the period of an activity and recognize the activities with repeating motion patterns (e.g. walking) from those without a repetition, Ustev *et al.* [27] add frequency domain features and improve recognition accuracy. Wang and Zhang [28] propose a method for extracting wavelet domain features including both time domain and frequency domain features to recognize six activities. From these previous studies, we find that the extracted features are greatly relevant to the analyzed activities.

Some advanced studies explore various human activities in HAR. Shoaib *et al.* [29] utilize smartphones to recognize activities, like smoking, eating, giving a talk, coffee drinking, walking and biking. Hsu *et al.* [30] present a wearable inertial sensor network to recognize ten common domestic activities in human daily lives and eleven sport activities. Their final recognition accuracy can reach more than 95%. Zubair *et al.* [31] analyze the fusing data of smartphone sensors including 30 subjects and recognize fifteen activities. Besides these daily activities, some research has been conducted in indoor activities related to multi-floor indoor positioning. The authors of [4], [6] recognize stair climbing, escalator or elevator taking as a supplement to construct indoor maps. However, up till now, we have not found a comprehensive study of indoor activities related to multi-floor indoor positioning, i.e., walking, stillness, stair climbing, escalator or elevator taking based on a robust machine learning method.

Different classification algorithms for HAR have been developed. Paul and George [32] use the KNN classification algorithm and Clustered KNN to recognize four activities of walking, running, sitting and standing, and achieve an overall accuracy of 92%. In their study, the Clustered KNN is an improvement of Minimum Distance and *K*-Nearest Neighbor classification algorithms to eliminate the computational complexity of KNN by creating clusters. San-Segundo *et al.* [33] present a human activity sensing system based on HMMs for classifying physical activities i.e., walking, upstair walking, downstair walking, sitting, standing and lying down. They apply a publicly available dataset to conduct experiment and achieve an activity recognition accuracy of 97.5%. Chen *et al.* [34] propose a robust HAR system based on coordinate transformation, principal component analysis (CT-PCA) and online support vector machine (OSVM) to improve the activity recognition accuracy. In addition, to evaluate the performance of each classifier, diverse classification algorithms are compared in [35] and [36]. In these algorithms, XGBoost has become a hot research topic, which has been applied to smartphone authentication, user authorization [37] and electroencephalography (EEG) based brain activity recognition [38]. However, in indoor activity recognition, XGBoost is rarely used. Besides, there are many factors that affect the performance of activity recognition, such as location and orientation of smartphone and sensors type, and they have been investigated in many studies [29], [35], [39].

Some publicly available datasets are used in study of HAR. Saha *et al* [57] utilize the University of Dhaka mobility dataset (DU-MD) which consists of eight daily activities, i.e., jogging, laying down, sitting, descending stairs, ascending stairs, standing, walking and falling. They apply SVM, KNN and ensemble of classifiers to the dataset for activity and fall classification and achieve HAR classification accuracy of 93%, fall detection accuracy of 97% respectively. Ye *et al* [58] propose a novel probabilistic algorithm pFTA for HAR and apply it to three publicly available sdatasets, UCI Daily and Sports Activities, Smartphone-Based Human Activity dataset and MSRActionPairs dataset. The experiments on these three public HAR datasets show the proposed pFTA approach can achieve competitive performance in both accuracy and efficiency. Xu *et al.* [15] use Mobile Health (mHealth) dataset from the UCI Machine Learning

Repository to recognize twelve activities such as standing, walking, cycling, jumping, etc., and present a set of activity recognition algorithm based on decision trees with ensemble approach. From these studies in recent years, we find that there are many datasets widely-used in HAR, but to our knowledge there is still no comprehensive indoor activity dataset, including data of walking, stillness, stair climbing, escalator or elevator taking.

Many researchers study HAR based on data from smartphone-sensors, among them, the closest study to our work is [43]. In [43], Chen and Shen present a performance analysis for HAR via smartphone-sensors. They first collect and preprocess 10 subjects' sensor data (i.e., accelerometer and gyroscope) using smartphone, then extract some features which have been normalized in time, frequency and wavelet domain. They implement three classifiers, KNN, random forest and SVM to recognize five human activities, i.e., descending stairs, ascending stairs, walking, jogging and jumping. They analyze four factors that may influence recognition performance: 1) various smartphone placement settings, 2) different user space, 3) combination of sensors, and 4) impact of data imbalance. Although they present a systematic performance evaluation for HAR, it is still greatly different from our work in this paper. First, in [43], they focus on a comprehensive analysis of HAR based on smartphone-sensor data, and consider some factors that may affect recognition performance, including sensor type, smartphone placement, etc., by using three classifiers widely applied in HAR. However we concentrate on comparing the performance of tree-based ensemble learning classifiers, XGBoost, Random Forest and GBDT to three commonly used single classifiers MLP, SVM and KNN. Specifically, we apply XGBoost, a classifier that is not widely used in indoor activity recognition, and achieve better recognition results. Second, the recognized activities are very different. In [43], they choose five common activities in daily life, i.e., stair descending, stair ascending, walking, jogging and jumping. While in our work, in order to address multiple-floor indoor positioning, we recognize five indoor activities, i.e., walking, stillness, stair climbing, escalator or elevator taking, to detect the changes of floor number comprehensively. Finally, we have different sizes of dataset. In [43], they build a dataset including 27681 sensory samples from 10 male subjects aging from 20 to 23. However, to make data diverse, our dataset contains more than 5 million sensory samples from 40 subjects from different gender, age and region.

In summary, the previous studies have not yet provided a comprehensive analysis for robust machine learning recognition of indoor activities. Moreover, the selected sensors and features are highly correlated with the analyzed activities, and they require to be deeply investigated for improving recognition accuracy. Although XGBoost has revealed excellent performance in many research results, it has not yet been applied in the recognition of the indoor activities related to multi-floor positioning even it has been proved a strong potential classification ability.

III. PREPARATION FOR ENSEMBLE LEARNING CLASSIFIERS BASED ON TREE STRUCTURE

Since the principle of XGBoost is based on two tree-based ensemble learning classifiers, i.e., random forest and GBDT, we give a brief introduction of them in this section.

A. RANDOM FOREST

Leveraging integration thinking, random forest algorithm combines multiple decision trees to improve the accuracy of classification, in which each decision tree is a base classifier. After a random forest model is trained, the prediction results of the model are obtained by a voting approach based on the classification results from all the decision trees. We summarize the main construction process of random forest as follows.

1) n pieces of data are randomly selected from the original training data as the input of random forest.
2) Each decision tree randomly selects m features from the entire feature set with a size of M. In general, m is much smaller than M.
3) In each decision tree, the feature with the smallest Gini index or the largest information gain is used to split the nodes. Other nodes of the decision tree are constructed using the same splitting rules until all training samples of the node belong to the same class or reach the maximum depth of the tree. In our experiment, we choose the feature that maximizes the information gain to split the nodes.
4) Repeat the above steps T times to get a random forest with T decision trees.

Random forest has two prominent advantages. First, the random selection of original training data and features improves the capability of noise resistance and generalization ability. Second, a parallel training model in random forest can speed up the training process. Therefore, random forest can achieve a high and robust recognition accuracy. Additionally, it is suitable for the applications based on the data with high-dimensional features.

B. GRADIENT BOOSTED DECISION TREE

GBDT (Gradient Boosted Decision Tree) combines multiple decision trees in a boosting way. In contrast with the random forest in which each tree is independent, the trees in GBDT are closely related to the previously established decision trees. In general, each new tree is created to reduce the residual of the previous model toward the gradient direction, in which the residual is defined as the difference between the true value and the predicted value. Until the residual is smaller than a threshold or the number of decision trees reaches a certain threshold, the final model has been trained.

Moreover, GBDT uses the negative gradient value of the loss function as an approximation of the residual to fit a decision tree, which means that the tree is established in the direction of the negative gradient of the loss function obtained by the previously established decision trees.

TABLE 1. GBDT algorithm description.

Gradient boosting decision tree
$f_{k0}(\mathbf{x}) = 0, \quad k = 1, \cdots, K$ For $m = 1$ to M do : $p_k(\mathbf{x}) = \exp(f_k(\mathbf{x})) \,/\, \sum_{l=1}^{K} \exp(f_l(\mathbf{x})), \quad k = 1, \cdots, K$ $\tilde{y}_{ik} = -\left[\frac{\partial L(\{y_{il}, f_l(x_i)\})_{l=1}^{K}}{\partial f_k(x_i)}\right]_{\{f_l(x)=f_{l,m-1}(x)\}_1^K}$ $= y_{ik} - p_{k,m-1}(x_i), \quad i = 1, \cdots, N$ $\{R_{jkm}\}_{j=1}^{J} = J$ terminal node tree $(\{\tilde{y}_{ik}, x_i\}_1^N)$ $\gamma_{jkm} = \frac{K-1}{K} \frac{\sum_{x_i \in R_{jkm}} \tilde{y}_{ik}}{\sum_{x_i \in R_{jkm}} \lvert\tilde{y}_{ik}\rvert (1 - \lvert\tilde{y}_{ik}\rvert)}, \quad j = 1, \cdots, J$ $f_{km}(\mathbf{x}) = f_{k,m-1}(\mathbf{x}) + \sum_{j=1}^{J} \gamma_{jkm} 1(\mathbf{x} \in R_{jkm})$ End For

Let $\mathbf{y}$ and $f(\mathbf{x})$ denote the labels of classes and the cumulative output values for all the current decision trees, respectively. In this paper, we select a log-likelihood loss function as the objective function to be optimized, i.e.,

$$L(\mathbf{y}, f(\mathbf{x})) = -\sum_{k=1}^{K} \mathbf{y}_k \log p_k(\mathbf{x}) \tag{1}$$

where $y_k = 1$ if the label of the sample is class k. $p_k(x)$ refers to the probability of the predicted value belonging to class k. The description of the GBDT algorithm is shown in Table 1, where M presents the number of iterations, K represents the number of categories, and N is the number of samples. $\tilde{y}_{ik}$ is the negative gradient of the loss function of the sample i corresponding to class k, and a decision tree $\{R_{jkm}\}_{j=1}^{J}$ consisting of J leaf nodes is obtained from each sample point X and $\tilde{y}_{ik}$. After the decision tree is established, the gain of each leaf node γ_{jkm} is calculated. Finally, the output values $f_{km}(\mathbf{x})$ is updated. It is proved that GBDT can achieve high recognition accuracy with a short time of parameter adjusting [44].

IV. HUMAN ACTIVITY RECOGNITION BASED ON XGBOOST

In this section, we introduce our proposed indoor activity recognition algorithm based on XGBoost. We first present the sensors and features used in our work related to the indoor activities of walking, stillness, stair climbing, escalator or elevator taking, in which some specific features have been extracted. Then, we present the XGBoost classifier for the indoor activity recognition.

A. SENSOR FEATURES

The selected features for HAR are closely correlated with the activities to be analyzed. Since the indoor activities of walking, stillness, stair climbing, escalator or elevator taking are rarely investigated in the previous work, we need to investigate the effective features for these activities, where we use n samples, and we denote each sample value as x_i.

1) TIME DOMAIN FEATURES

Time domain features are the most commonly used features for HAR. We first introduce the time-domain features used in our work.

Mean: It is an indicator of trends in the signal.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2}$$

Variance: It is a measure of the differences between the source signals and the mean.

$$\sigma = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{3}$$

Max and Min: It is a reflection or a range of signal.

Median: It is a value in the middle of signal that is not affected by maximum and minimum values.

Interquartile Range: It is an dispersion indication of signal in middle 50%. It becomes larger if the signals are more concentrated in the middle. The interquartile range (IQR) is defined as

$$\mathrm{IQR} = \mathrm{Q}_3 - \mathrm{Q}_1 \tag{4}$$

where Q_3 is 75% of the values in the sample and Q_1 is 25% of the values.

Kurtosis: It is an indicator to reflect the sharpness or flatness of the distribution curve comparing with the normal distribution. The value of zero indicates the distribution same as the normal distribution. The value larger than zero shows that the distribution is steeper than the normal distribution. Otherwise, the distribution is flat compared with the normal distribution. Kurtosis is defined as

$$Kurt = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} \tag{5}$$

2) FREQUENCY DOMAIN FEATURES

The indoor activities to be analyzed show different frequency characteristics. Hence, in this work, we apply some features in frequency domain to improve recognition accuracy. Fast Fourier Transform (FFT) converts time domain signals into frequency domain signals which show the information in different frequencies.

Values and Indexes of Top and Second Peaks: We take stair climbing and escalator taking as an instance to introduce the feasibility of these frequency features for indoor activity recognition. A user climbing stairs will make periodic changes of accelerometer data which generate a large amplitude with his frequency. However, escalator taking normally generates more stable accelerometer data and the peak value in frequency domain may be small. Therefore, in consideration of the indoor activities to be analyzed, the peak values in frequency domain and their corresponding indexes can be extracted as features. Therefore, we extract the first two peaks in the frequency domain as the features since they have very different values for different activities.

Spectrum Entropy: For different activities, the complexity of signals in frequency domain is very different. For example, the complexity of accelerometer data in frequency domain for elevator taking is much lower than that for stair climbing. We know that the spectrum entropy can represent uncertainty and complexity of recognition signals. The more complex the signal, the larger the spectral entropy. Spectrum Entropy is defined as

$$H = -\sum_{i=1}^{n} p[i] \times \log(p[i]) \tag{6}$$

where

$$p[i] = P[i] / \sum_{i=1}^{n} P[i] \tag{7}$$

and $P[i]$ is the square of amplitude of the frequency domain signal.

3) WAVELET DOMAIN FEATURES

As introduced in the aforementioned subsections, the indoor activities show the specific features in both time domain and frequency domain. The traditional Fourier transform can only describe the frequency characteristics of data but can not provide the relevant features in time domain. Hence, we further apply wavelet analysis to investigate the features in both frequency domain and time domain.

Wavelet Energy: In wavelet decomposition, each layer decomposes low frequency signals from upper layer into low frequency and high frequency signals. There are many wavelet functions can be used in wavelet decomposition, e.g., Haar (haar), Daubechies (db), Symlets (sym), Coiflets (coif), Biorthogonal (bior), and Reverse biorthogonal (rbio). In this work, we take three-order Daubechies wavelet (db3) at four levels to decompose data. Then, signals are decomposed into a sum of subbands as

$$x = D_1 + D_2 + D_3 + D_4 + A_4 \tag{8}$$

where D_i is the high frequency signals obtained by decomposition of the i th layer and A_i is the low frequency signals. Here, we apply wavelet energy to extract the following feature

$$E = \|D_3\|^2 + \|D_4\|^2 \tag{9}$$

B. XGBOOST

In many pioneering works of HAR, the most commonly used classifiers are the traditional single classifiers, e.g., Decision Tree, SVM, MLP, and KNN. The performance of these single classifiers can be improved by increasing the amount of training data. However, the improvement of the algorithms gets marginal after the training data increases to a certain extent. In other word, even if a single classifier is trained with a large amount of data, its recognition accuracy cannot meet our requirement. As a result, we consider to combine several models to improve the recognition accuracy.

As mentioned in Section III, GBDT is a boosting learner based on multiple decision trees and uses the gradient boosting to iteratively form a strong learner. In this case, the algorithm often runs a certain number of trees to achieve satisfactory recognition accuracy. However, when the dataset is large and complex, it is probable to take thousands of iterations to train a GBDT model resulting in a computational bottleneck to use the algorithm. In order to address this problem, a recent scalable machine learning system for tree boosting eXtreme Gradient Boosting, namely XGBoost, is proposed in [45].

XGBoost can be efficiently implemented compared with GBDT. Its objective function is

$$Obj(\phi) = L(\mathbf{y}, f(\mathbf{x})) + \sum_{m} \Omega(f_m) \tag{10}$$

$L(\mathbf{y}, f(\mathbf{x}))$ represents the loss function and $\Omega(f)$ is the regularization item indicating the complexity of the model. The additional regularization item in XGBoost is added to avoid overfitting and simplify the model compared with GBDT. The objective function of the mth iteration is calculated by

$$\begin{aligned} Obj^{(m)} &= \sum_{i=1}^{N} L(y_i, \hat{y}_i^{(m)}) + \sum_{i=1}^{m} \Omega(f_i) \\ &= \sum_{i=1}^{N} L(y_i, \hat{y}_i^{(m-1)} + f_m(x_i)) + \Omega(f_m) \end{aligned} \tag{11}$$

A second order Tailor expansion is conducted on Equation (11) as

$$Obj^{(m)} \simeq \sum_{i=1}^{N} \left[L(y_i, \hat{y}_i^{(m-1)}) + g_i f_m(x_i) + \frac{1}{2} h_i f_m^2(x_i) \right] + \Omega(f_m) \tag{12}$$

where g_i and h_i are the first and second order gradient statistics on the loss function

$$\begin{aligned} g_i &= \partial_{\hat{y}^{(m-1)}} L(y_i, \hat{y}^{(m-1)}) \\ h_i &= \partial^2_{\hat{y}^{(m-1)}} L(y_i, \hat{y}^{(m-1)}) \end{aligned} \tag{13}$$

After removing the constant terms, we can obtain

$$Obj^{(m)} \simeq \sum_{i=1}^{N} \left[g_i f_m(x_i) + \frac{1}{2} h_i f_m^2(x_i) \right] + \Omega(f_m) \tag{14}$$

According to the CART theory, $f_m(x)$ can be determined by the structure of the tree q and its leaf weights w, i.e.,

$$f_m(x) = w_{q(x)}, \quad w \in \mathbf{R}^T, \quad q : \mathbf{R}^d \rightarrow \{1, 2, \cdots T\} \tag{15}$$

where T represents the number of leaf nodes in a tree, $q(x)$ is a mapping used to map a sample to a leaf node, that is to represent the structure of the tree. And the regularization item can be described as

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \tag{16}$$

where γ is the complexity of each leaf and λ is a parameter to scale the penalty.

We can rewrite (14) according to (15) and (16) as

$$\begin{aligned} Obj^{(m)} &\simeq \sum_{i=1}^{N}\left[g_i w_{q(x_i)} + \frac{1}{2}h_i w_{q(x_i)}^2\right] + \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2 \\ &= \sum_{j=1}^{T}\left[\left(\sum_{i\in I_j} g_i\right) w_j + \frac{1}{2}\left(\sum_{i\in I_j} h_i + \lambda\right) w_j^2\right] + \gamma T \\ &= \sum_{j=1}^{T}\left[G_j w_j + \frac{1}{2}\left(H_j+\lambda\right) w_j^2\right] + \gamma T \qquad (17) \end{aligned}$$

where

$$\begin{aligned} G_j &= \sum_{i\in I_j} g_i \\ H_j &= \sum_{i\in I_j} h_i \qquad (18) \end{aligned}$$

and $I_j = \{i|q(x_i) = j\}$ is defined as the sample set of leaf j.

When $q(x)$ is fixed,we can compute the optimal weight w_j^* of leaf j and the corresponding value of objective function

$$w_j^* = -\frac{G_j}{H_j+\lambda} \qquad (19)$$

$$Obj^* = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j+\lambda} + \gamma T \qquad (20)$$

TABLE 2. Greedy algorithm description.

Exact greedy algorithm for split finding
Input: I, instance set of current node
Input: d, fearture dimension
$gain \leftarrow 0$
$G \leftarrow \sum_{i\in I} g_i, H \leftarrow \sum_{i\in I} h_i$
For $q = 1$ to Q do :
$G_L \leftarrow 0,\ H_L \leftarrow 0$
For $j\ in\ sorted(I,\ by\ x_{jq})$ do
$G_L \leftarrow G_L + g_j, H_L \leftarrow H_L + h_j$
$G_R \leftarrow G - G_L, H_R \leftarrow H - H_L$
$score \leftarrow \max(score, \frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{G^2}{H+\lambda})$
End For
End For
Output:Split with max score

Obj^* can evaluate the structure of a tree. For searching the best structure of tree, we can apply the greedy strategy, that is each feature and all its values are traversed, and the feature and its value corresponding to the maximum gain are selected to perform node splitting. The description of the greedy algorithm is shown in Table 2 [45]. And the gain after the split is

$$Gain = \left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right] - \gamma \qquad (21)$$

For searching the best splitting point, the exact greedy algorithm needs to be enumerated over all possible splitting leaf nodes greedily. In order to improve the computational efficiency, a more efficient algorithm than the greedy strategy is proposed in XGBoost. The general idea is to enumerate several candidate points that may become the splitting points according to the percentile method, and then to find the best splitting point from the candidate points. This approach can greatly reduce the computational cost. Another advantage of XGBoost is its support for parallelism. In the learning process of a tree, features need to be sorted by the loss function to determine the optimal splitting point. Before training, XGBoost sorts the features in advance and saves them as a block [46]. This block is used repeatedly in iterations which greatly reduces the amount of calculation.

To achieve an optimal performance of XGBoost for the indoor activity recognition, we need to correctly set the parameters in XGBoost. In general, XGBoost has the following parameters required to be optimized by cross-validation.

- **Learning rate:** Lower learning rate improves the robustness of the model and the optimum model can be accurately obtained. However, a lower learning rate requires more iterations to find the optimal model.
- **Number of subtrees:** It is the number of iterations in training. A XGBoost model with more subtrees normally has better performance, but it will require more training time.
- **Gamma:** In XGBoost, a node is only split if the value of the loss function decreases after splitting. Gamma specifies the minimum drop value of the loss function for the node splitting. An algorithm becomes more conservative with the increasing Gamma.
- **L1 regularization and L2 regularization weights:** These two parameters prevent overfitting.
- **Maximum depth of a tree:** It controls the complexity of a model.
- **Minimum weight sum of leaf node sample:** It is also used to prevent overfitting.

V. IMPLEMENTATION

In this section, we introduce the implementation of the proposed indoor activity recognition system. The overview of the system is shown in Figure 1.

A. DATA COLLECTION AND PREPROCESSING

1) DATA COLLECTION

We build a dataset for indoor activities by using a self-designed Android application for data collection running on smartphone. Using six kinds of phones as shown in Table 3, we collect diverse sensor data in the application such as accelerometer, gyroscope, magnetometer, proximity sensor, barometer, light sensor, GPS, and WiFi RSSI. In this work,

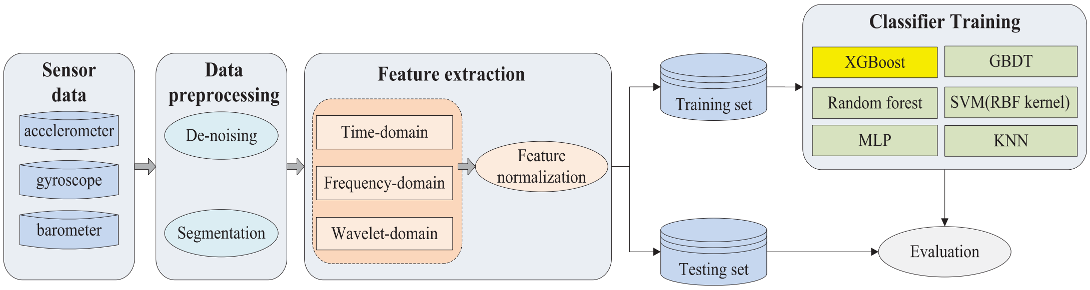


FIGURE 1. Basic process framework of HAR.

TABLE 3. Sampling rate of barometer in different phones.

Smartphone Type	Sample Rate
Xiaomi-mi note	5Hz
Xiaomi-mi 5	25Hz
Xiaomi-mi 5s	25Hz
Xiaomi-mi Mix	25Hz
Samsung-9350	25Hz
Xiaomi-mi 4	50Hz

we only use accelerometer, gyroscope and barometer to recognize the indoor activities. In these sensors, the accelerometer and gyroscope are sampled with 50 Hz. The barometer is sampled according to the type of smartphone given in Table 3. We store the values of sensors together with their sampling timestamp in milliseconds. Then, we synchronize the values from different sensors with their timestamps.

To build the dataset, we collect the sensor data from 40 subjects performing five daily activities, i.e., walking, stillness, stair climbing, escalator or elevator taking. Moreover, the sensor data are collected with the smartphone placed in different locations, i.e., in trouser pocket, in bag, swinging in hand and holding horizontally. For each activity, the sensor data are collected around three minutes for each subject. To ensure the natural movement of each subject, we do not control the speed and the way of their movements. We collect about 60 minutes of data for each subject. The total length of the dataset we have built is 2,400 minutes with about 5.16 million samples. All the data are uploaded to a cloud server for processing.

2) DE-NOISING

The raw data collected from the sensors are normally noisy which will significantly degrade the performance of activity recognition. Filtering techniques are often used to mitigate the influence of these unwanted noise [47]. The most common filtering methods include mean filter, low-pass filter, Gaussian filter, wavelet filter and Kalman filter, etc. In this work, we apply the five-spot triple smoothing algorithm [43] to smooth the raw data as

$$\begin{cases} \bar{x}_{-2} = (69x_{-2} + 4x_{-1} - 6x_0 + 4x_1 - x_2)/70 \\ \bar{x}_{-1} = (2x_{-2} + 27x_{-1} + 12x_0 + 4x_1 - x_2)/35 \\ \bar{x}_0 = (-3x_{-2} + 12x_{-1} + 17x_0 + 12x_1 - 3x_2)/35 \\ \bar{x}_1 = (2x_{-2} - 8x_{-1} + 12x_0 + 27x_1 + 2x_2)/35 \\ \bar{x}_2 = (-x_{-2} + 4x_{-1} - 6x_0 + 4x_1 - 69x_2)/70 \end{cases} \tag{22}$$

where $(x_{-2}, x_{-1}, x_0, x_1, x_2)$ are the five time-adjacent points for a sensor data series and $(\bar{x}_{-2}, \bar{x}_{-1}, \bar{x}_0, \bar{x}_1, \bar{x}_2)$ are the smoothed samples. Compared with the other filtering methods, the five-spot triple smoothing algorithm is more time efficient and consumes less cache.

3) DATA SEGMENTATION

After de-noising, in order to determine the activity of a subject over a period of time, we need to slice the data. Inspired by the studies of [48], [49], we extract a sliding window in 5 seconds overlapped 10% from the previous window.

B. FEATURE EXTRACTION AND NORMALIZATION

We extract the features introduced in Section IV-A in each sliding window. Because different ranges of features will influence the final recognition results, we normalize all the features to a certain range. The normalization can eliminate the adverse effects caused by singular sample data and avoid the recognition results dominated by certain features with a wide range of values. Normalization techniques, such as z-normalization (zNorm), batch normalization (BN), and pressure mean subtraction (PMS) are widely utilized in HAR [50]. In our work, we apply z-normalization as

$$z = \frac{x - \mu}{\sigma} \tag{23}$$

where μ and σ refer to the mean and standard deviation over all available training data.

C. CLASSIFIER IMPLEMENTATION

In this work, we implement XGBoost to achieve indoor activity recognition with high accuracy by adjusting the parameters as introduced in Section IV-B. Moreover, we also implement other five commonly used classifiers, i.e., random forest, GBDT, MLP, SVM and KNN for comparison. Note that this work focus on applying XGBoost to recognize indoor activities and the other five classifiers are only as the references during our evaluation. Moreover, we adopt a ten-fold cross validation method to find the optimal combination of parameters in XGBoost to prevent overfitting. We divide the training set into ten parts, nine of which are used as the training data and the rest as the testing data. we obtain a F-score for each test and calculate the average of F-scores over ten times. The parameters with the largest means are selected as the optimal parameters to train the model.

VI. EXPERIMENTS AND RESULTS EVALUATION

In this section, we present our evaluation method and the analysis of results. In the evaluation, we analyze the following factors that affect the recognition results: 1) various classifiers, 2) sensors combination, 3) feature extraction, 4) parameter tuning of XGBoost, 5) positions to carry smartphone, 6) number of subject and 7) model scalability.

A. PERFORMANCE EVALUATION METHOD

In order to evaluate the performances of recognition algorithms, e.g., accuracy and generalization, the dataset is divided into the training and the testing one. The training set is used to build and fit the model, while the testing set is utilized to evaluate the discriminating ability of the classifier for new samples. In this work, we employ ten-fold cross validation to evaluate the performance. In the validation process, we divide the dataset with 40 subjects into ten parts, each of which contains 4 subjects and is labeled with i, $i = 1, 2, \cdots, 10$. Each time we select the ith set as the testing set and the rest as the training set to train a model. In this way, ten models are established and the prediction results of these ten models are averaged to get the final results for avoiding the effects of individual abnormal data on the results. In order to evaluate the generalization of the classifiers, we provide two performance measures, accuracy and F-score. The F-score is a comprehensive consideration of precision and recall. Moreover, we also investigate the training time of the models and their prediction time, which are important for the design of a real-world system.

B. RESULTS AND ANALYSIS

1) ACTION RECOGNITION VIA DIVERSE CLASSIFIERS

As we know that machine learning algorithms are normally application depended, we compare the performance of our proposed XGBoost to the other five commonly-used classifiers, i.e., random forest, GBDT, MLP, SVM, and KNN, based on the the dataset we built for indoor activities.

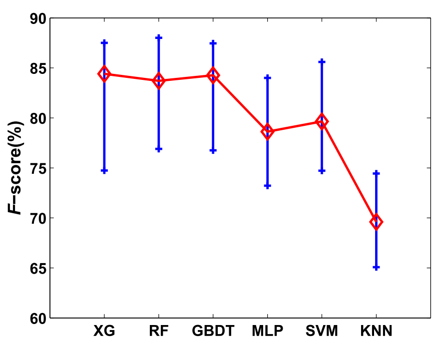


FIGURE 2. Average *F*-scores of various classifiers.

Figure 2 gives the measured F-scores and their averages over ten times (ten-fold cross validation as introduced in Section VI-A). Table 4 shows the detailed results. As shown in Figure 2, compared with the single classifiers of MLP, SVM and KNN, the ensemble learning classifiers behave better, in which the F-score increases by 5%. Additionally, both training and prediction time are greatly shortened for the ensemble learning compared with those of MLP and SVM. As for KNN, although the training time is short, we note that the prediction time is long and the recognition performance is inferior to those of the other classifiers. The average F-score is only 69.60%.

For the three ensemble learning classifiers, XGBoost and GBDT provide the best performance from the perspective of recognition F-score, whose F-scores reach 84.41% and 84.26% compared with random forest of 83.71%. To well evaluate performance of a classifier, it is based on not only its F-score, but also the computational cost reflected by training and prediction time. The running time of a program determines CPU time of a system. The longer the running time, the more resources a CPU uses. As shown in Table 4, the training time of GBDT is the longest, whose average training time is 985.475 seconds. For random forest and XGBoost, the training time are comparable, which are 296.579 seconds and 476.343 seconds respectively. We known from Section III-B that GBDT is difficult to train in parallel due to the dependence among the weak classifiers. Additionally, the loss function and its gradient under all features in GBDT are calculated each time for splitting the feature nodes, which is extremely time consuming for high-dimensional training data with diverse features. By contrast, the parallel training supported XGBoost and random forest can accelerate the training rate, which is suitable for high-dimensional data processing. Consequently, we conclude that XGBoost is the best choice in these six commonly-used classifiers for the indoor activity recognition considering both recognition F-score and computation cost.

2) SENSOR COMBINATION

Accelerometer and gyroscope are the most commonly-used sensors for HAR. For the indoor activities in this work, we leverage barometer as another sensor to improve the recognition F-score. The barometer measures the changes of atmospheric pressure at altitude. In this work, we develop two ways of sensor combination, i.e., the combination of accelerometer and gyroscope, and the combination of accelerometer, gyroscope, and barometer. Figure 3 shows F-scores of these two combinations and their averages over ten times. Table 5 presents the detailed results.

In Figure 3, we can see that the performance of all the classifiers deteriorates after removing barometer. For example, the recognition F-score of XGBoost degrades from 84.41% to 76.98%, but XGBoost still outperforms the other classifiers. Thus, we can conclude that barometer is indispensable for indoor activity recognition. The reason is that some of the indoor activities, e.g., stillness and elevator taking, are difficult to be recognized only by accelerometer and gyroscope, while they can be distinguished by the altitude changes measured by barometer.

TABLE 4. Recognition performance of various classifiers.

	$F_{\min}$	$F_{\max}$	F_{AVG}	$Acc_{\min}$	$Acc_{\max}$	Acc_{AVG}	$T-tra_{AVG}(s)$	$T-pre_{AVG}(s)$
XGBOOST	74.74%	87.50%	84.41%	75.04%	87.47%	84.56%	476.343	0.645
Random Forest	76.90%	88.01%	83.71%	77.14%	88.05%	83.88%	296.579	0.251
GBDT	76.76%	87.45%	84.26%	77.11%	87.45%	84.42%	985.475	0.337
MLP	73.22%	84.00%	78.63%	73.18%	84.06%	78.72%	1671.324	1.007
SVM	74.72%	85.60%	79.65%	74.78%	85.68%	79.75%	1782.887	131.604
KNN	65.08%	74.44%	69.60%	65.11%	74.67%	69.64%	1.746	731.022

Note: '$F_{\min}$' - minimum F-score among 10 results, '$F_{\max}$' - maximum F-score, F_{AVG} - average F-score, $Acc_{\min}$ - minimum accuracy, $Acc_{\max}$ - maximum accuracy, Acc_{AVG} - average accuracy, $T-tra_{AVG}$ - average training time, $T-pre_{AVG}$ - average prediction time.

TABLE 5. Recognition performance without barometer.

	$F_{\min}$	$F_{\max}$	F_{AVG}	$Acc_{\min}$	$Acc_{\max}$	Acc_{AVG}	$T-tra_{AVG}(s)$	$T-pre_{AVG}(s)$
XGBOOST	70.80%	81.74%	76.98%	70.95%	82.32%	77.18%	419.030	0.648
Random Forest	68.90%	82.67%	76.23%	69.08%	83.41%	76.49%	262.475	0.253
GBDT	69.48%	80.29%	75.91%	69.68%	81.20%	76.14%	858.081	0.313
MLP	66.63%	77.87%	72.41%	66.83%	78.37%	72.49%	2170.054	1.000
SVM	68.97%	81.42%	74.54%	69.09%	81.81%	74.68%	2089.323	148.350
KNN	62.35%	70.31%	65.92%	62.32%	70.96%	65.95%	1.265	493.131

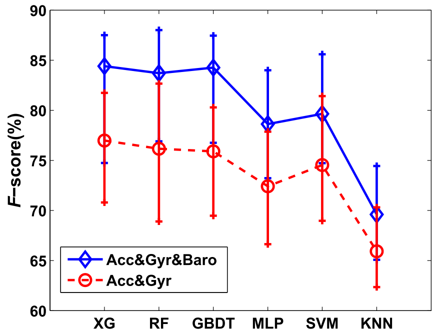


FIGURE 3. Average *F*-scores with or without barometer.

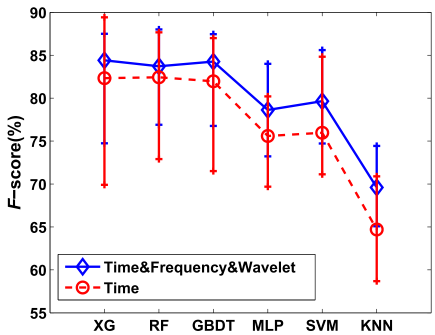


FIGURE 4. Average *F*-scores v.s. various features.

3) IMPACT OF FEATURE SELECTION

In Section IV-A, we introduce some specific features in frequency domain and wavelet domain, i.e., the top and second peak values and indexes, the spectrum entropy and the wavelet energy, to improve recognition F-score of the indoor activities. In order to find the effects, we compare the performance of the classifiers with and without these features. The experimental results are shown in Figure 4 and Table 6.

The recognition F-score clearly deteriorates after removing the frequency domain and wavelet domain features. Considering the XGBoost classifier, we find that the F-score degrades by 2%. Therefore, for the indoor activities in this work, our selected features are applicable, especially the specific features in frequency domain and wavelet domain are beneficial to improve the recognition performance.

4) PARAMETER TUNING

In order to acquire higher recognition F-score, it is necessary to tune parameters for an algorithm. Since XGBoost has more than thirty hyper-parameters, we only select the following seven parameters with higher impact on the optimization performance, i.e., learning rate, number of subtrees, gamma, L1 regularization, L2 regularization, maximum depth of a tree and minimum weight sum of leaf node sample, which are introduced in Section IV-B. Table 7-Table 11 show the detailed results of the parameters tuned by ten-fold cross validation.

Figure 5 gives the confusion matrix for the XGBoost classifier. According to Figure 5, the indoor activities of stillness, escalator or elevator taking are most difficult to distinguish, whose the average F-scores are all below 80%, resulting in a lower overall F-score. The reason is that the moving styles of these activities are very similar, and correspondingly their features are with high similarity. Compared with these activities, the recognition accuracy of stair climbing (96.2%) is much higher.

5) IMPACT OF SMARTPHONE PLACEMENT LOCATIONS

In daily life, people normally place their smartphone variously, so that we need to study the impact of phone placement locations on the recognition results. For same activity, the sensor data collected from different locations may be very

TABLE 6. Recognition performance without frequency and wavelet domain features.

	$F_{\min}$	$F_{\max}$	F_{AVG}	$Acc_{\min}$	$Acc_{\max}$	Acc_{AVG}	$T-tra_{AVG}(s)$	$T-pre_{AVG}(s)$
XGBOOST	69.90%	89.41%	82.33%	70.79%	89.79%	82.56%	284.476	0.621
Random Forest	72.90%	87.671%	82.43%	73.42%	88.21%	82.64%	269.196	0.237
GBDT	71.51%	87.00%	81.97%	72.69%	87.57%	82.22%	564.873	0.210
MLP	69.69%	80.20%	75.59%	69.71%	80.16%	75.54%	1707.71	0.948
SVM	71.132%	84.83%	75.97%	71.34%	85.39%	76.16%	850.082	98.154
KNN	58.68%	70.89%	64.70%	58.96%	71.92%	64.86%	0.971	232.348

TABLE 7. F_{mean} result v.s. learning rate.

Learning rate	F_{mean}
0.01	0.79974
0.05	0.83764
0.07	0.84253
0.1	0.84431
0.2	0.83156

TABLE 8. F_{mean} result v.s. number of subtree.

Number of subtree	F_{mean}
100	0.84431
150	0.84006
200	0.83571
250	0.83281
300	0.83011
350	0.82752
400	0.82597
450	0.82504

TABLE 9. F_{mean} result v.s. gamma.

Gamma	F_{mean}
0	0.84431
0.1	0.84549
0.2	0.84495
0.3	0.84493
0.4	0.8441
0.5	0.84296
0.6	0.84281

different. Taking stair climbing as an instance, a smartphone in a bag may have wide shaking or rotation compared with in hands. Hence, it is necessary to study the recognition performance of the activities regarding diverse locations of smartphone placement. Here, we investigate different classifiers on four different locations of smartphone placement, i.e., in trouser pocket, in bag, hand, and holding horizontally. The results are shown in Figure 6 and Table 12-Table 15.

From Figure 6, we can see that the locations of smartphone placement have an impact on HAR. The recognition algorithm achieves the best F-score when a smartphone is placed in trouser pocket, because the smartphone is fixed with human body and can better represent the body movement. On the other hand, the leg movement is the most significant

TABLE 10. F_{mean} result v.s. L1 regularization and L2 regularization.

L1 \ L2	0.05	0.1	1	2	3
0.05	0.84487	0.84428	0.84515	0.84433	0.84568
0.1	0.84463	0.84282	0.84466	0.84561	0.84602
1	0.84514	0.84368	0.84583	0.84569	0.84583
2	0.84477	0.84522	0.84490	0.84376	0.84413
3	0.84486	0.84463	0.84418	0.84465	0.84248

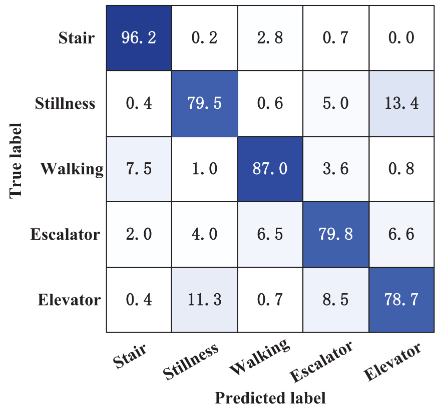


FIGURE 5. Confusion matrix for XGBoost classifier.

in human body for the analysis of indoor activities. For example, during walking and stair climbing, the degree of leg bending is quite different. For smartphone held horizontally, the smartphone can also well represent human movement but the obtained features are not as significant as those when a smartphone is in trouser pocket. For the other locations, i.e., in bags and hands, a smartphone may be flipped or moved at random, which results in that the collected data contains the noise not related to the human activities. Therefore, the F-score with these two smartphone placement locations are lower.

6) SUBJECT COMPOSITION

The movement styles are normally personalized due to different age, height, weight, gender, sports habits, health status, and motion frequency. Hence, the values of features extracted from different persons may also be different even for the same activities. If we only apply a dataset with a limited number of subjects for training, the result classifier may be very personalized. Therefore, we further analyze the performance of the recognition algorithms with different composition of

TABLE 11. F_{mean} **result v.s. maximum depth of tree and minimum weight sum of leaf node sample.**

$depth_{max}$ \ $weigh_{min}$	1	2	3	4	5
3	0.83351	0.83459	0.8370	0.83280	0.83338
4	0.84325	0.84377	0.84273	0.84351	0.84260
5	0.84431	0.84512	0.84463	0.84618	0.84434
6	0.84383	0.84387	0.84309	0.84419	0.84358
7	0.84065	0.84053	0.83884	0.83982	0.84121
8	0.83666	0.83740	0.83528	0.83716	0.83594
9	0.83383	0.83477	0.83552	0.83485	0.83620

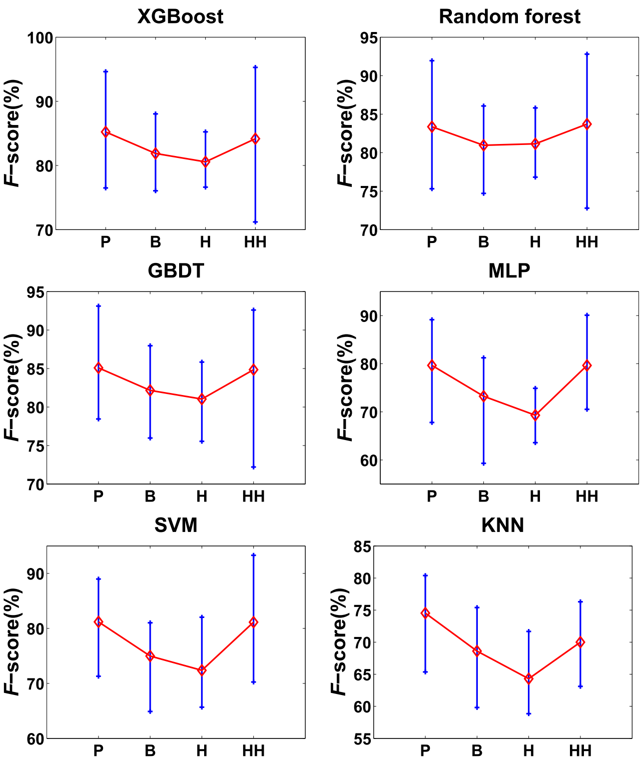


FIGURE 6. **Average *F*-scores for diverse settings of smartphone-position.The x-axis labels are 'P' - trouser pocket, 'B' - bag, 'H'- hand, and 'HH'- hold smartphone horizontally, respectively.**

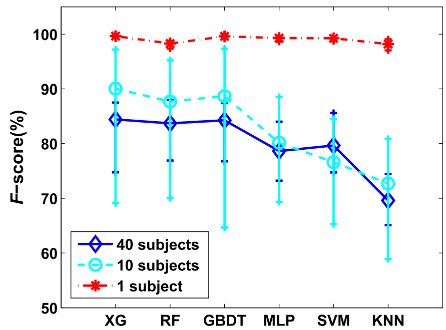


FIGURE 7. **Average *F*-scores v.s. various subjects.**

subjects. We set the number of subjects in the dataset to 1, 10 and 40 respectively. Figure 7 reports the *F*-scores and their averages over ten times. In Figure 7, three fold lines represent different sizes of dataset, i.e., 40 subjects, 10 subjects and 1 subject. Among them, the dark blue solid line indicates the recognition results using the dataset of 40 subjects, the cyan dotted line is the recognition results of 10 subjects, and the red dot dash line indicates the results of training and prediction using single subject's data. The detailed results are shown in Table 16 and Table 17.

According to Figure 7, it is evident that as the number of subjects increases, the *F*-score decreases. Taking XGBoost classifier as an instance, the average *F*-scores are 99.63%, 90.03% and 84.41% with the number of subjects being 1, 10 and 40 respectively. The reason is that, due to various characteristics and behavioral habits of subjects, the behaviors even for the same activity are also different. This may make the data collected by the sensors vary greatly. If the classifier is trained only based on the data from a single subject, its *F*-score is the highest with its own data for prediction, because all the personalized data are included in both training and prediction. Moreover, for a single subject, the performances of all the six classifiers are very similar and with *F*-scores higher than 98%. Nevertheless, in this case, the number of subjects is too small, leading to low generalization ability of the model, which is prone to overfitting. Although the *F*-score decreases with the increasing number of subjects, the performance of the ensemble learning classifiers, especially XGBoost, is more robust compared with the single classifiers.

7) CONSISTENCE AND SCALABILITY OF THE ALGORITHM

The aforementioned evaluation is conducted based on the dataset whose data are collected by a self-designed data collection application running on smartphones. Moreover, in order to verify the consistence and scalability of our proposed model, we further conduct evaluations on a publicly available dataset. To best of our knowledge, there is currently no comprehensive indoor activity dataset related to indoor positioning, we choose a publicly available dataset Smartphone-Based Recognition of Human Activities and Postural Transitions dataset (HAPT) from the UCI Machine Learning Repository [51], in which the physical activities are similar to the activities that we recognize. The dataset includes 30 volunteers between ages of 19 and 48, and they perform a protocol of activities composed of six basic activities including standing, sitting, lying, walking, walking downstairs and walking upstairs. During the experiments, data are collected by wearing a smartphone (Samsung

TABLE 12. Recognition performance when smartphone in trouser pocket.

	$F_{\min}$	$F_{\max}$	F_{AVG}	$Acc_{\min}$	$Acc_{\max}$	Acc_{AVG}	$T-tra_{AVG}(s)$	$T-pre_{AVG}(s)$
XGBOOST	76.50%	94.63%	85.23%	76.87%	94.64%	85.32%	104.582	0.418
Random Forest	75.31%	91.95%	83.36%	76.54%	91.96%	83.61%	52.310	0.056
GBDT	78.45%	93.11%	85.09%	78.70%	93.12%	85.15%	169.181	0.033
MLP	67.79%	89.15%	79.10%	67.85%	89.38%	79.13%	272.099	0.369
SVM	71.30%	88.98%	81.20%	71.06%	89.06%	81.29%	50.278	7.063
KNN	65.36%	80.38%	74.53%	65.50%	80.13%	74.56%	0.246	40.113

TABLE 13. Recognition performance when smartphone in bag.

	$F_{\min}$	$F_{\max}$	F_{AVG}	$Acc_{\min}$	$Acc_{\max}$	Acc_{AVG}	$T-tra_{AVG}(s)$	$T-pre_{AVG}(s)$
XGBOOST	76.06%	88.04%	81.87%	75.94%	88.29%	82.04%	108.616	0.154
Random Forest	74.71%	86.06%	80.96%	76.58%	86.05%	81.30%	56.542	0.057
GBDT	75.97%	87.96%	82.16%	75.59%	88.16%	82.40%	169.756	0.033
MLP	59.30%	81.23%	73.24%	61.39%	81.43%	73.68%	260.103	0.368
SVM	64.90%	81.02%	74.98%	65.93%	81.90%	75.37%	56.796	7.672
KNN	59.82%	75.40%	68.61%	60.23%	75.36%	68.96%	0.277	51.212

TABLE 14. Recognition performance when smartphone in hand.

	$F_{\min}$	$F_{\max}$	F_{AVG}	$Acc_{\min}$	$Acc_{\max}$	Acc_{AVG}	$T-tra_{AVG}(s)$	$T-pre_{AVG}(s)$
XGBOOST	76.62%	85.24%	80.57%	77.11%	85.26%	80.24%	106.766	0.148
Random Forest	76.82%	85.81%	81.15%	77.22%	86.72%	81.51%	54.614	0.062
GBDT	75.55%	85.84%	81.40%	76.41%	86.31%	81.66%	164.334	0.032
MLP	59.30%	81.23%	73.24%	61.39%	81.43%	73.68%	260.103	0.368
SVM	65.68%	82.04%	72.38%	66.62%	83.04%	72.72%	61.033	8.348
KNN	58.55%	71.67%	64.30%	58.29%	72.14%	64.53%	0.281	39.972

TABLE 15. Recognition performance when holding smartphone horizontally.

	$F_{\min}$	$F_{\max}$	F_{AVG}	$Acc_{\min}$	$Acc_{\max}$	Acc_{AVG}	$T-tra_{AVG}(s)$	$T-pre_{AVG}(s)$
XGBOOST	71.20%	95.29%	84.17%	72.16%	95.35%	84.47%	102.877	0.145
Random Forest	72.79%	92.80%	83.71%	73.35%	92.92%	83.83%	52.078	0.057
GBDT	72.21%	92.60%	84.85%	73.21%	92.73%	85.01%	164.412	0.030
MLP	70.53%	90.09%	79.64%	70.79%	90.17%	79.87%	301.561	0.340
SVM	70.25%	93.30%	81.14%	70.77%	93.38%	81.22%	58.050	8.383
KNN	63.11%	76.30%	70.00%	63.37%	76.87%	70.17%	0.258	44.498

TABLE 16. Recognition performance of 10 subjects.

	$F_{\min}$	$F_{\max}$	F_{AVG}	$Acc_{\min}$	$Acc_{\max}$	Acc_{AVG}	$T-tra_{AVG}(s)$	$T-pre_{AVG}(s)$
XGBOOST	69.10%	97.19%	90.03%	69.67%	97.20%	90.18%	105.174	0.120
Random Forest	70.03%	95.21%	87.68%	71.36%	95.23%	88.09%	49.989	0.047
GBDT	64.66%	97.34%	88.67%	68.73%	97.36%	89.18%	169.114	0.318
MLP	69.31%	88.52%	80.15%	70.94%	88.60%	80.72%	262.841	0.353
SVM	65.29%	84.52%	76.59%	67.21%	84.83%	77.43%	44.913	6.137
KNN	58.89%	80.86%	72.70%	59.38%	80.87%	73.09%	0.252	43.550

Galaxy S II) on waist, and they capture the embedded accelerometer and gyroscope of the smartphone as sensor data. In order to better apply the dataset to our proposed model, the original raw data instead of pre-processed data are utilized as the input data of our framework because the pre-processed data have already been de-noised, sampled within sliding windows and extracted from specified feature, which is repeated with our data processing. Figure 8 shows F-scores of our dataset and HAPT and their averages over ten times. The detailed results are presented in Table 18.

As shown in Figure 8, for HAPT, the ensemble learning classifiers, XGBoost, random forest and GBDT outperforms

TABLE 17. Recognition performance of 1 subject.

	$F_{\min}$	$F_{\max}$	F_{AVG}	$Acc_{\min}$	$Acc_{\max}$	Acc_{AVG}	$T-tra_{AVG}(s)$	$T-pre_{AVG}(s)$
XGBOOST	99.53%	99.84%	99.63%	99.38%	99.84%	99.63%	9.211	0.010
Random Forest	97.35%	98.76%	98.29%	97.35%	98.75%	98.28%	3.345	0.096
GBDT	99.38%	99.84%	99.61%	99.38%	99.84%	99.61%	13.897	0.005
MLP	98.60%	99.69%	99.28%	98.60%	99.84%	99.28%	38.592	0.037
SVM	98.75%	98.91%	99.24%	98.75%	99.69%	99.24%	0.769	0.080
KNN	97.03%	99.07%	98.17%	97.03%	99.06%	98.17%	0.164	0.530

TABLE 18. Recognition performance of using HAPT.

	$F_{\min}$	$F_{\max}$	F_{AVG}	$Acc_{\min}$	$Acc_{\max}$	Acc_{AVG}	$T-tra_{AVG}(s)$	$T-pre_{AVG}(s)$
XGBOOST	81.29%	87.84%	84.19%	81.23%	87.78%	84.20%	208.377	0.164
Random Forest	78.22%	86.03%	82.64%	78.18%	85.99%	82.63%	76.750	0.097
GBDT	80.76%	87.94%	83.85%	80.70%	87.88%	83.86%	216.351	0.046
MLP	79.59%	84.70%	81.96%	79.52%	84.70%	81.99%	140.341	0.177
SVM	75.15%	84.93%	81.86%	75.16%	85.02%	81.91%	6.986	1.159
KNN	77.65%	83.11%	80.80%	77.67%	83.20%	80.85%	0.078	7.398

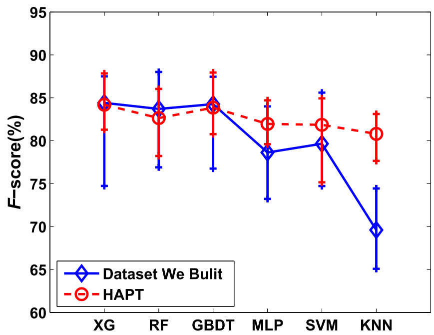


FIGURE 8. Average *F*-scores v.s. various datasets.

the single classifiers of MLP, SVM and KNN, in which F-scores improve by 3%. This finding is consistent with our evaluation conducted on our dataset. Moreover, comparing with other two ensemble learning classifiers, XGBoost also performs best, whose F-score reaches 84.19%, while random forest and GBDT reach 82.64% and 83.85%, respectively. For the computational cost reflected by training time and prediction time, according to Table 18, we can see that among the three ensemble learning classifiers, the training time of GBDT is also the longest, whose average training time reach 216.351 seconds, while XGBoost and random forest are 208.377 seconds and 76.750 seconds, respectively. Although the training time of random forest is shorter than that of XGBoost, the average F-score reduces by nearly 2% compared to that of XGBoost. For the single classifiers, the prediction times are all longer than those of the ensemble learning classifiers, especially SVM and KNN require the longest time with prediction time of 1.159 seconds and 7.398 seconds, respectively. Therefore, we can conclude that XGBoost achieves the best recognition performance not only on our dataset but also the HAPT dataset, which proves the consistence and scalability of the model.

Additionally, in contrast with the recognition results based on our dataset, the performance of the single classifiers, MLP, KNN and SVN, gets better in the HAPT dataset, which approaches that of the ensemble learning classifiers. All the six classifiers are almost similar, whose F-scores range from 80.80% to 84.19%. For MLP, KNN and SVM, they are prone to the overfitting problems. Even if the training errors on the training set reduce, it is probable to result in large predicting errors on the testing set because of the overfitting. There are two reasons that for the similiar recognition performance of these six classifiers. First, in the HAPT dataset, the data collection is conducted in a more standardized way and all participants complete all activities in same order. In contrast, in the data collection process, we give every subject enough freedom to ensure his natural movement, which may make the data vary greatly. Hence the single classifiers prone to overfitting problem performs better in the HAPT dataset than in our dataset. This finding further demonstrate that the ensemble learning approaches are robust to the overfitting problem. Second, in the HAPT dataset, all participants wear their own smartphone on the waist durning data collection which offers more significant features for recognizing different activities. However, in our dataset, the phone placement locations are more diverse. The features may not be as significant as those on the waist. Therefore, for the single classifiers, it is more challenging to achieve high accuracy using our dataset.

VII. CONCLUSION

In this work, we propose an approach based on XGBoost to recognize five kinds of indoor activities related to indoor positioning, i.e., walking, stillness, stair climbing, escalator or elevator taking. Moreover, besides some commonly used features in time domains, we analyze some specific features in frequency domain and wavelet domain extracted from the data of accelerometer, gyroscope and barometer to improve recognition accuracy. Based on a dataset created by 40 volunteers, we conduct comprehensive analysis of our recognition algorithms on the indoor activities. We achieve a recognition accuracy of 84.41% based on XGBoost. In consideration of both the accuracy and computation cost, the results show that the model based on XGBoost outperforms the other classifiers. Moreover, our proposed specific features in frequency and wavelet domains can well represent the characteristic of these indoor activities and help us to improve the recognition accuracy. We also study the impact of subject composition and the different smartphone placement locations on the recognition results, and find that less subjects may cause overfitting and the best location of smartphone placement is in trouser pocket for the recognized activities.

In this work, we observe that the placement of smartphone has an impact on the recognition performance. A two-step activity recognition algorithm, i.e., placement recognition and then activity recognition, is considered as an effective solution to this problem in our future work. In this way, each activity on each placement location is trained by a classification model. After recognizing the placement of smartphone, a corresponding model can be selected for activity recognition. Moreover, deep learning has become a powerful tool in many fields, we plan to conduct our research in indoor activity recognition via deep learning methods in the near future.

REFERENCES

[1] Z. Li, D. B. Acuña, Z. Zhao, J. L. Carrera, and T. Braun, "Fine-grained indoor tracking by fusing inertial sensor and physical layer information in WLANs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, May 2016, pp. 1–7.

[2] F. Li, C. Zhao, G. Ding, J. Gong, C. Liu, and F. Zhao, "A reliable and accurate indoor localization method using phone inertial sensors," in *Proc. ACM Conf. Ubiquitous Comput.*, Pittsburgh, PA, USA, Sep. 2012, pp. 421–430.

[3] Z. Li, X. Zhao, and H. Liang, "Automatic construction of radio maps by crowdsourcing PDR traces for indoor positioning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.

[4] B. Zhou, Q. Li, Q. Mao, W. Tu, X. Zhang, and L. Chen, "ALIMC: Activity landmark-based indoor mapping via crowdsourcing," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2774–2785, Oct. 2015.

[5] Y. He, J. Liang, and Y. Liu, "Pervasive floorplan generation based on only inertial sensing: Feasibility, design, and implementation," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1132–1140, May 2017.

[6] C. Luo, H. Hong, L. Cheng, K. Sankaran, and M. C. Chan, "iMap: Automatic inference of indoor semantics exploiting opportunistic smartphone sensing," in *Proc. IEEE 12th Annu. Int. Conf. Sens., Commun., Netw. (SECON)*, Seattle, WA, USA, Jun. 2015, pp. 489–497.

[7] C. Luo, H. Hong, M. C. Chan, J. Li, X. Zhang, and Z. Ming, "MPiLoc: Self-calibrating multi-floor indoor localization exploiting participatory sensing," *IEEE Trans. Mobile Comput.*, vol. 17, no. 1, pp. 141–154, Jan. 2018.

[8] W. Zijlstra and K. Aminian, "Mobility assessment in older people: New possibilities and challenges," *Eur. J. Ageing*, vol. 4, no. 1, pp. 3–12, 2007.

[9] S.-R. Ke, H. L. U. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, and K.-H. Choi, "A review on video-based human activity recognition," *Computers*, vol. 2, pp. 88–131, Jun. 2013.

[10] K. Y. Chen, R. C. Shah, J. Huang, and L. Nachman, "Mago: Mode of transport inference using the Hall-effect magnetic sensor and accelerometer," in *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 2, Jun. 2017, Art. no. 8.

[11] A. M. Khan, Y.-K. Lee, S. Y. Lee, and T.-S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 5, pp. 1166–1172, Sep. 2010.

[12] Z. Chen, C. Jiang, and L. Xie, "A novel ensemble ELM for human activity recognition using smartphone sensors," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2691–2699, May 2019.

[13] A. Wang, G. Chen, J. Yang, S. Zhao, and C.-Y. Chang, "A comparative study on human activity recognition using inertial sensors in a smartphone," *IEEE Sensors J.*, vol. 16, no. 11, pp. 4566–4578, Jun. 2016.

[14] Y. Zheng, W. K. Wong, X. Guan, and S. Trost, "Physical activity recognition from accelerometer data using a multi-scale ensemble method," in *Proc. AAAI 27th Conf. Artif. Intell.*, 2013, pp. 1575–1581.

[15] Z. Xu, J. Wei, J. Zhu, and W. Yang, "A robust floor localization method using inertial and barometer measurements," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sapporo, Japan, Sep. 2017, pp. 1–8.

[16] C. Lang and S. Kaiser, "Classifying elevators and escalators in 3D pedestrian indoor navigation using foot-mounted sensors," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Nantes, France, Sep. 2018, pp. 1–7.

[17] W. Xu, M. Zhang, A. A. Sawchuk, and M. Sarrafzadeh, "Robust human activity and sensor location corecognition via sparse signal representation," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 11, pp. 3169–3176, Nov. 2012.

[18] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *Proc. 4th Int. Workshop Ambient Assist. Living*, Vitoria-Gasteiz, Spain, 2012, pp. 216–223.

[19] Y. Saez, A. Baldominos, and P. Isasi, "A comparison study of classifier algorithms for cross-person physical activity recognition," *Sensors*, vol. 17, no. 1, p. 66, Dec. 2016.

[20] Y.-S. Lee and S.-B. Cho, "Activity recognition using hierarchical hidden Markov models on a smartphone with 3D accelerometer," in *Proc. Int. Conf. Hybrid Artif. Intell. Syst.* Berlin, Germany: Springer, May 2011, pp. 460–467.

[21] Z. Feng, L. Mo, and M. Li, "A Random Forest-based ensemble method for activity recognition," in *Proc. IEEE 37th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Milan, Italy, Aug. 2015, pp. 5074–5077.

[22] C. Tran and M. M. Trivedi, "3-D posture and gesture recognition for interactivity in smart spaces," *IEEE Trans. Ind. Informat.*, vol. 8, no. 1, pp. 178–187, Feb. 2012.

[23] N. Zerrouki, F. Harrou, Y. Sun, and A. Houacine, "Vision-based human action classification using adaptive boosting algorithm," *IEEE Sensors J.*, vol. 18, no. 12, pp. 5115–5121, Jun. 2018.

[24] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 790–808, Nov. 2012.

[25] A. Jain and V. Kanhangad, "Human activity classification in smartphones using accelerometer and gyroscope sensors," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1169–1177, Feb. 2018.

[26] Y. Vaizman, K. Ellis, and G. Lanckriet, "Recognizing detailed human context in the wild from smartphones and smartwatches," *IEEE Pervasive Comput.*, vol. 16, no. 4, pp. 62–74, Oct./Dec. 2017.

[27] Y. E. Ustev, O. D. Incel, and C. Ersoy, "User, device and orientation independent human activity recognition on mobile phones: Challenges and a proposal," in *Proc. ACM Conf. Pervasive Ubiquitous Comput. Adjunct Publication*, Zürich, Switzerland, Sep. 2013, pp. 1427–1436.

[28] C. Wang and W. Zhang, "Activity recognition based on smartphone and dual-tree complex wavelet transform," in *Proc. 8th Int. Symp. Comput. Intell. Design (ISCID)*, Hangzhou, China, Dec. 2015, pp. 267–270.

[29] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "Complex human activity recognition using smartphone and wrist-worn motion sensors," *Sensors*, vol. 16, no. 4, p. 426, 2016.

[30] Y.-L. Hsu, S.-C. Yang, H.-C. Chang, and H.-C. Lai, "Human daily and sport activity recognition using a wearable inertial sensor network," *IEEE Access*, vol. 6, pp. 31715–31728, 2018.

[31] M. Zubair, K. Song, and C. Yoon, "Human activity recognition using wearable accelerometer sensors," in *Proc. IEEE Int. Conf. Consum. Electron.-Asia (ICCE)*, Oct. 2016, pp. 1–5.

[32] P. Paul and T. George, "An effective approach for human activity recognition on smartphone," in *Proc. IEEE Int. Conf. Eng. Technol. (ICETECH)*, Coimbatore, India, Mar. 2015, pp. 1–3.

[33] R. San-Segundo, J. D. Echeverry-Correa, C. Salamea, and J. M. Pardo, "Human activity monitoring based on hidden Markov models using a smartphone," *IEEE Instrum. Meas. Mag.*, vol. 19, no. 6, pp. 27–31, Dec. 2016.

[34] Z. Chen, Q. Zhu, S. Y. Chai, and L. Zhang, "Robust human activity recognition using smartphone sensors via CT-PCA and online SVM," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 3070–3080, Dec. 2017.

[35] Y.-C. Huang, C.-W. Yi, W.-C. Peng, H.-C. Lin, and C.-Y. Huang, "A study on multiple wearable sensors for activity recognition," in *Proc. IEEE Conf. Dependable Secure Comput.*, Taipei, Taiwan, Aug. 2017, pp. 449–452.

[36] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Evaluation of supervised classification algorithms for human activity recognition with inertial sensors," in *Proc. IEEE SENSORS*, Glasgow, U.K., Oct./Nov. 2017, pp. 1–3.

[37] Z. Salam Patrous, "Evaluating XGBoost for user classification by using behavioral features extracted from smartphone sensors," Ph.D. dissertation, School Elect. Eng., Comput. Sci., KTH Royal Inst. Technol., Stockholm, Sweden, Jul. 2018.

[38] X. Zhang, L. Yao, D. Zhang, X. Wang, Q. Z. Sheng, and T. Gu, "Multi-person brain activity recognition via comprehensive EEG signal analysis," in *Proc. MobiQuitous*, Melbourne, VIC, Australia, Nov. 2017, pp. 28–37.

[39] A. Henpraserttae, S. Thiemjarus, and S. Marukatat, "Accurate activity recognition using a mobile phone regardless of device orientation and location," in *Proc. Int. Conf. Body Sensor Netw.*, Dallas, TX, USA, May 2011, pp. 41–46.

[40] S. S. Saha, S. Rahman, M. J. Rasna, T. B. Zahid, A. K. M. M. Islam, and M. A. R. Ahad, "Feature extraction, performance analysis and system design using the DU mobility dataset," *IEEE Access*, vol. 6, pp. 44776–44786, 2018.

[41] J. Ye, G. Qi, N. Zhuang, H. Hu, and K. A. Hua, "Learning compact features for human activity recognition via probabilistic first-take-all," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.

[42] S. Khatun and B. I. Morshed, "Fully-automated human activity recognition with transition awareness from wearable sensor data for mHealth," in *Proc. IEEE Int. Conf. Electro/Inf. Technol. (EIT)*, Rochester, MI, USA, May 2018, pp. 934–938.

[43] Y. Chen and C. Shen, "Performance analysis of smartphone-sensor behavior for human activity recognition," *IEEE Access*, vol. 5, pp. 3095–3110, 2017.

[44] C. Pan, J. Tan, and D. Feng, "Identification of power quality disturbance sources using gradient boosting decision tree," in *Proc. Chin. Automat. Congr. (CAC)*, Xi'an, China, Nov./Dec. 2018, pp. 2589–2592.

[45] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. ACM 22nd SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jun. 2016, pp. 785–794.

[46] Z. Wen, B. He, R. Kotagiri, S. Lu, and J. Shi, "Efficient gradient boosted decision tree training on GPUs," in *Proc. IEEE Int. Parallel Distrib. Process. Symp. (IPDPS)*, Vancouver, BC, Canada, May 2018, pp. 234–243.

[47] S. Anitha, L. Kola, P. Sushma, and S. Archana, "Analysis of filtering and novel technique for noise removal in MRI and CT images," in *Proc. Int. Conf. Elect., Electron., Commun., Comput., Optim. Techn. (ICEECCOT)*, Mysuru, India, Dec. Dec. 2017, pp. 1–3.

[48] X. Qi, G. Zhou, Y. Li, and G. Peng, "RadioSense: Exploiting wireless communication patterns for body sensor network activity recognition," in *Proc. IEEE 33rd Real-Time Syst. Symp.*, San Juan, Puerto Rico, Dec. 2012, pp. 95–104.

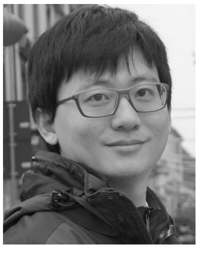
[49] F. Al Machot, H. C. Mayr, and S. Ranasinghe, "A windowing approach for activity recognition in sensor data streams," in *Proc. 8th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Vienna, Austria, Jul. 2016, pp. 951–953.

[50] S. Münzner, P. Schmidt, A. Reiss, M. Hanselmann, R. Stiefelhagen, and R. Dürichen, "CNN-based sensor fusion techniques for multimodal human activity recognition," in *Proc. ACM Int. Symp. Wearable Comput.*, Maui, Hawaii, Sep. 2017, pp. 158–165.

[51] L. Oneto, J. L. R. Ortiz, and D. Anguita, "Constraint-aware data analysis on mobile devices: An application to human activity recognition on smartphones," in *Adaptive Mobile Computing*. New York, NY, USA: Elsevier, 2017, pp. 127–149.

WENTING ZHANG received the bachelor's degree in communication engineering from Jilin University, China, in 2017, where she is currently pursuing the master's degree with the College of Communication Engineering. Her research interests include wireless communication networks and human activity recognition.

XIAOHUI ZHAO received the Ph.D. degree in control theory from the Universite de Technologie de Compiegne, France, in 1993. He is currently a Full Professor with the College of Communication Engineering, Jilin University, China. His current research interests include signal processing and wireless communication, and cognitive radio.



ZAN LI received the Ph.D. degree from the University of Bern, Switzerland, in 2016. He was with the Alibaba Group, China, as a Senior Algorithm Engineer, from 2016 to 2017. Since 2017, he has been with the College of Communication Engineering, Jilin University, as a Lecturer. He was a recipient of the Best Paper Award from the WMNC 2014. His Ph.D. dissertation received the Fritz-Kutter Award 2016 (the Best Ph.D. Dissertation in Swiss Universities).