Evaluation of Random Forest for Complex Human Activity Recognition Using Wearable Sensors

Aiguo Wang, Huihui Chen, Chundi Zheng
School of Electronic Information Engineering
Foshan University
Foshan, China
wangaiguo2546@163.com, {hhchen, cdzheng}@fosu.edu.cn

Liang Zhao, Jinjun Liu
School of Computer and Information Engineering
Chuzhou University
Chuzhou, China
{leon, jinjunl}@chzu.edu.cn

Lulu Wang
Biomedical Device Innovation Center
Shenzhen Technology University
Shenzhen, China
wanglulu@sztu.edu.cn

***Abstract*—The recognition of human activities plays a central role in bridging the gap between the raw sensor signals and high-level pervasive applications. However, the complex nature of human behavior presents a great challenge to the choice of representative features and a discriminant classification model and further makes it quite difficult to develop an accurate and robust activity recognizer. Random forest benefits from the idea of bootstrap and random feature sampling, while only preliminary experimentation for simple activity recognition has been previously reported and few studies, as far as we know, have comprehensively and systematically discussed its performance, especially in wearable-based complex activity recognition. The objective of this study is to present the first empirical results obtained with random forest and to compare its performance with other commonly used prediction models and dimensionality reduction methods. We conduct extensive comparative experiments on three benchmarked datasets. The results indicate the superiority of random forest-based activity recognizers over five individual classifiers, three ensemble models, one feature extractor as well as two feature selectors in terms of recognition performance across scenarios. Besides, the parameter sensitivity analysis recommends the default setting for the parameters and the time cost analysis indicates its applicability to practical applications.**

Keywords-Complex activity recognition; ensemble learning; random forest; feature selection; generalization

I. INTRODUCTION

The rapid development of sensing technologies, internet of things, and artificial intelligence has greatly advanced the pervasive computing and further facilitated the in-depth understanding of human behavior and the smart provision of various application services that range from smart home and ambient assisted living to healthcare, wellness evaluation, and among others [1]. Accordingly, how to accurately and automatically recognize human activities plays a central role in bridging the gap between the raw sensor signals and high-level activity semantics [2]. However, the complex nature of human behavior makes it difficult to develop an accurate and robust activity recognition model [3].

Compared with the traditional classification tasks such as web mining and image recognition, activity recognition faces challenges from the data, the modelling and evaluation, and also human behavior itself. The typical challenges of human activity recognition are given as follows [3, 4].

1) *Interleaved and concurrent activity execution*. Besides performing activities sequentially, one can switch between the steps of two or more activities and can simultaneously do two or more activities, which presents a complex issue.

2) *Inter-subject and intra-subject variations*. Different individuals probably perform the same activity differently and even an individual may perform an activity in a different way at different places and times. Particularly, human behavior may change over time. This would also degrade the generalization performance of an activity recognizer.

3) *Confusion between similar activities*. For a specific application, there may exist predefined activities that have similar sensor readings, which would potentially reduce the group discrimination and result in degraded accuracy.

Therefore, how to accurately and automatically recognize on-going human activities is a challenging but rewarding topic that deserves further investigation [4, 5].

Accordingly, researchers have been exploring different types of sensing units and proposed a number of activity recognition models. We can broadly group existing methods into three categories based on the sensing units: environment sensor-based, vision-based, and wearable sensor-based methods [4, 6]. Particularly, the miniature and increasing processing and communication power of sensing units makes it possible for an individual to simultaneously take multiple heterogeneous or homogeneous sensing devices. Wearable sensor-based methods are also suitable for both outdoor and indoor scenarios and have advantages of low costs and portability. For activity recognition models, researchers have used generative models, discriminant models, unsupervised learning models, and ensemble models. To better handle the complexity of human behavior, one feasible solution is to utilize ensemble learning that typically combines multiple base classifiers. Particularly, random forest, an ensemble of decision trees, generally obtains satisfactory prediction performance and is robust to noise and outliers in multiple domains. Preliminary experimentation on simple activity recognition has been previously reported [7], while few studies have systematically and comprehensively made an

978-1-7281-8954-3/20/$31.00 ©2020 IEEE
DOI 10.1109/NaNA51271.2020.00060

empirical evaluation in the context of wearable-based complex activity recognition on several aspects. For example, *how does it perform when compared with other widely used individual classification models and ensemble learning models? How does it perform in comparison with the feature selection and feature extraction methods? What are the empirically recommended default values for its parameters (e.g., the number of used trees and the number of features randomly selected at each node of a tree)? Is the random forest-based activity recognizer suitable for wearable devices that have limited resources?* Herein, we aim to systematically study the random forest-based activity recognizer and compare it with commonly used models. This work, as far as we know, is the first to conduct an extensive empirical evaluation on this topic. Particularly, the main contributions of this study are as follows. (1) We present the random forest-based activity recognizer under the framework of activity recognition chain, where we detail its components and discuss its mechanism in returning an accurate and diverse set of trees. (2) We conduct extensive experiments on three datasets and compare it with five individual classifiers, three ensemble learning methods, and two feature selectors and one feature extraction method. The comprehensive suite of experiments shows its superiority over its competitors. (3) The results of parameter sensitivity analysis indicate the recommended empirically values of the parameters. Besides, we initially analyze its time costs in making predictions, which indicates its applicability to practical applications.

The reminder of this paper is organized as follows. In Section II, we briefly discuss related work on sensors and activity recognition model. Section III presents the wearable-based activity recognition framework. Experimental results are given in Section IV, followed by the conclusion section.

II. Related Work

Researchers have used different types of sensing units and presented a large number of activity recognition models. Generally, we can divide them into three groups according to the used sensing units: vision-based, environment sensor-based, and wearable-based methods. Compared with the vision-based methods that utilize a camera to capture and detect human activities and the environment sensor-based methods that infer the on-going activity by placing sensing units on household objects and capturing the interaction between an individual and the environment [6, 8], wearable-based methods record the sensor signals and recognize activities from them when one performs activities. Benefiting from the miniature and ever-increasing processing and communication power of sensing units, they have advantages of a high degree of portability, low costs, and a wide range of applications. Representative wearable sensing units include, but not limited to, gyroscope, accelerometer, temperature and light sensors, electrocardiograph, and radio frequency identification (RFID) [9, 10, 11]. Particularly, one could take multiple sensing units on different parts of the human body. For example, Zappi et al. collected the sensing data of ten manipulative gestures from a car maintenance assembly-line worker, where 2x10 USB sensors with accelerometers were worn on the left and right upper and lower arm [12]. Kim et al. developed an RFID-based healthcare monitoring system, which used the RFID tags to detect the location and activities of the elderly [10]. Besides, there are researches that simultaneously use different types of sensing units for activity recognition. For example, Wang et al. exploited the accelerometer and gyroscope in a smartphone to train an activity recognizer [2]. Results show that the fusion of both accelerometer and gyroscope data leads to higher accuracy than the case with single source data.

As for the activity recognizers, there are a wealth of models and we can broadly group them into knowledge-driven and data-driven methods. Compared with knowledge-driven methods that heavily rely on an abstract model of domain knowledge (e.g., logical modelling, reasoning, and ontology) to define the activity specification, data-driven methods use the sensor data to construct and optimize an activity recognizer and then use it to make predictions. Accordingly, researchers have used many classification models that include generative methods (e.g., hidden Markov model and naïve Bayes), discriminant methods (e.g., support vector machine and decision tree), unsupervised learning methods, and ensemble learning methods (e.g., boosting, bagging, and stacking). For example, Wang et al. utilized naïve Bayes and *k*-nearest-neighbor to train activity recognizers [2]. Zappi et al. trained a hidden Markov model-based activity recognizer [12]. Xu et al. used the random forest to recognize six simple activities (*walking*, *go-upstairs*, *go-downstairs*, *jump*, *run*, and *static*) [7]. Feng et al. proposed a random forest-based ensemble to recognize human activities [13]. As an ensemble model, random forest tends to obtain satisfactory performance and is robust to noise and outliers. Previous researches have used it for simple human activity recognition, while few studies, to our knowledge, have systematically and comprehensively studied its performance and there is no related work that empirically recommends the default values of its values for complex activity recognition. In this study, we present the first work that systematically investigates its performance and addresses related issues on this topic. Specifically, we compare it with other widely used classification models (including five individual classifiers and three ensemble methods) and three feature reduction methods (i.e., principal component analysis, reliefF, and minimum-redundancy maximum-relevancy), and evaluate its sensitivity to the parameter values and time costs in making predictions. This study is expected to guide users in choosing and optimizing a robust activity recognizer for practical application scenarios.

III. Activity Recognition Framework

Wearable sensor-based activity recognition systems work by first recording sensor recordings during the on-going activities and training an activity recognizer offline and then using it to infer the activities. Fig. 1 presents the typical activity recognition chain (ARC), which mainly includes four stages to get an activity recognizer. First, the streaming raw sensor signals are collected and segmented with a static or dynamic sliding window with/without overlap between two adjacent segments. Second, we extract the time domain, frequency domain, and time-frequency features, such as

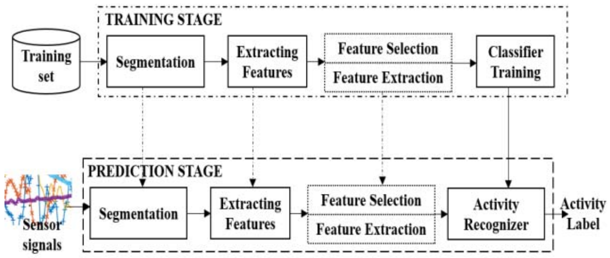


Figure 1. Wearable-based activity recognition chain.

mean, standard deviation, maximum, minimum, cross zero rate, entropy, signal magnitude area, autoregression coefficients, energy and correlation coefficient from each segment to form the feature vector. Third, we use feature extraction and feature selection methods to reduce the feature space by removing irrelevant and redundant features or projecting the original data into a low-dimensional space. It is noteworthy that this step is optional. Fourth, we train an activity recognizer with the training set. Afterwards, in the prediction phase, the test sensor data are first segmented using the same processing techniques as the training phase, features as those used in the training phase are extracted from the segment to return a feature vector, and then the activity recognizer outputs the corresponding activity label.

Random forest is an ensemble of unpruned classification trees and uses the majority voting rule to make predictions. It has advantages of getting satisfactory results and robustness to noise and outliers. The key to returning a set of diverse base classifiers is to randomly select bootstrap samples from the training set and to randomly get a feature subset at each node of each tree. We present the pseudo-code of random forest-based activity recognizer in Algorithm 1, where *numTrees* means the number of trees and *numFeatures* refers to the number of features to sample for each tree node. As for the ARC, a random forest is trained in step 4 of the training phase and used to classify test samples in the prediction phase. That is, it is a component of the ARC.

Algorithm 1: Random Forest-based Human Activity Recognition Model

Input: a train set D with activity labels, parameters *numTrees*, *numFeatures*, test set x
Output: the activity label lx of x

// construct the random forest-based activity recognizer
1. **for** i = 1: *numTrees* **do**
    1.1) randomly bootstrap a dataset D_i from D, and $|D_i| = |D|$
    1.2) construct a Tree T_i on D_i, where *numFeatures* features are randomly selected for each node
    1.3) RF.add(T_i) // add T_i to the forest RF

// predict the label of x
2. $lx = RF(x)$; // aggregate the votes of all trees T_i
3. **return** lx

IV. Experimental Setting and Results

A. Experimental Setting

In this study, we perform extensive experiments on three publicly available datasets for activity recognition. Table I presents the activities of interest. *PAMAP* has six human activities and *PAMAP2* contains fifteen activities [14]. Both

TABLE I. Experimental Data Description

Dataset	*Activities of Interest*
PAMAP	*lie, sit/stand, walk, run, cycle, Nordic walk*
PAMAP2	*stand, lie, sit, walk, run, cycle, Nordic walk, ascending stairs, descending stairs, fold laundry, rope jump, vacuum clean, iron, play soccer, clean house*
SkodaMiCP	*open hood, close hood, open left front door, check gaps on the front, open and close trunk, check trunk gaps, close both left doors, close left front door, checking steering wheel, write on notepad*

are obtained by asking nine subjects wearing one heart rate monitor and three inertial measurement units (IMU) to do activities. The sampling frequency of the IMU is 100 Hz and the sampling rate of the heart rate monitor is 9 Hz. The data is processed using a sliding window of 5.12 seconds with a shifting of 1 second. The third dataset *SkodaMiCP* consists of the sensor signals of ten manipulative gestures of a car maintenance assembly-line worker, who wore sensors on the right and left lower and upper arm [12]. It was collected for about 3 hours. Each sensor is a 3-axis accelerometer that has a 64 Hz sampling rate. The sensor data were segmented by a 1s sliding window with 50% overlap between two adjacent segments. For the three datasets, time-domain, frequency-domain and time-frequency domain features, such as mean, maximum, minimum, variance, skewness, 25th percentile, and 75th percentile, are extracted from the segments to form a feature vector.

Besides using random forest (RF), we explore other five individual models having different metrics (i.e., k nearest neighbor with k = 1 (1NN) and k = 3 (3NN), naïve Bayes (NB), support vector machine with linear kernel (SVM), and decision tree (DT)), as well as three ensemble methods (i.e., AdaBoost, Bagging, and Subspace learning). RF essentially utilizes bagging (bootstrap sample) and the random subspace method (random feature selection) to construct randomized trees. Furthermore, we include three dimensionality reduction methods as a comparison. Principal component analysis (PCA), a feature extractor, projects the high-dimensional data into a new lower dimensional subspace. The reliefF individually evaluates the goodness of each feature and ranks the features based on the concepts of near-hit and near-miss. The minimum redundancy maximum relevancy (mRMR) captures the higher-order statistics and considers redundancy among features.

For the performance metrics, five-fold cross validation is used and we use accuracy and F1 to evaluate the activity recognizers. F1 is the harmonic mean of recall and precision to deal with class imbalance, as given in (1).

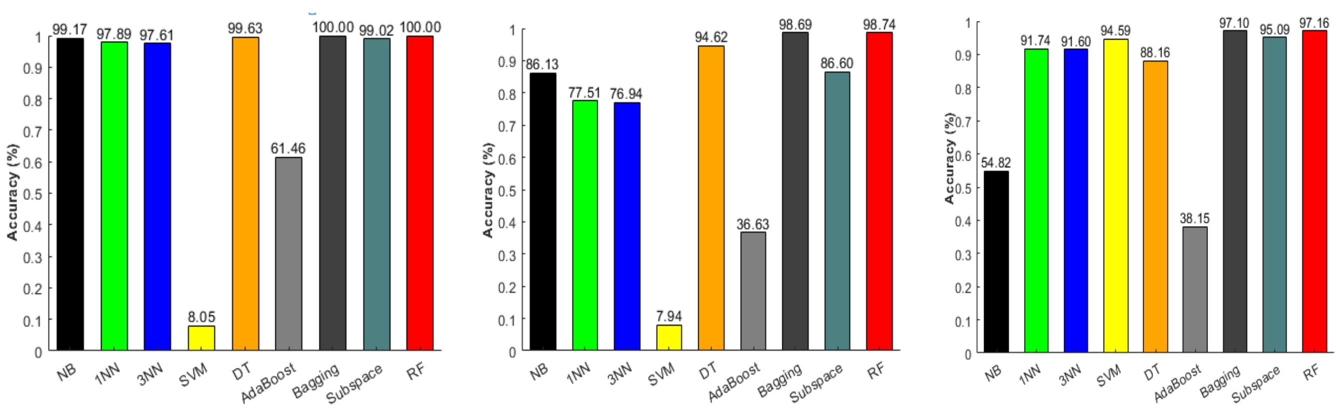
$$F1 = \frac{2 * precision * recall}{precision + recall}\ . \tag{1}$$

B. Comparison with Other Classifiers

We first compare the RF-based activity recognizer with other five individual classifiers (i.e., NB, 1NN, 3NN, SVM, and DT) and three ensemble classifiers (i.e., AdaBoost, Bagging, and Subspace). For AdaBoost and Bagging, we use the decision tree as the base classifiers. For the subspace learning, we use KNN as the base classifier. The number of base classifiers of the ensemble model equals that of the used random forest. According to our preliminary experimental results, in this study, the number of trees *numTrees* is 50 and the number of features *numFeatures* equals the square root of the total number of features. Figs. 2 and 3 present results. We observe that RF-based activity recognizer generally outperforms its competitors on all the datasets. Particularly, compared with AdaBoost, Bagging and Subspace, RF-based activity recognizer performs better, which is mainly because of the simultaneous sampling of both samples and features. Obviously, this contributes to the return of a set of accurate and diverse base classifiers of a random forest. Furthermore, we notice that the use of ensemble methods may fail to obtain satisfactory results. For example, AdaBoost obtains the 61.46% accuracy on the *PAMAP*, which is lower than the 99.17%, 97.89%, and 97.61% accuracy of NB, 1NN, and 3NN, respectively. Among the five individual classifiers, we see the performance of SVM is unstable across applications and DT remains a priority.

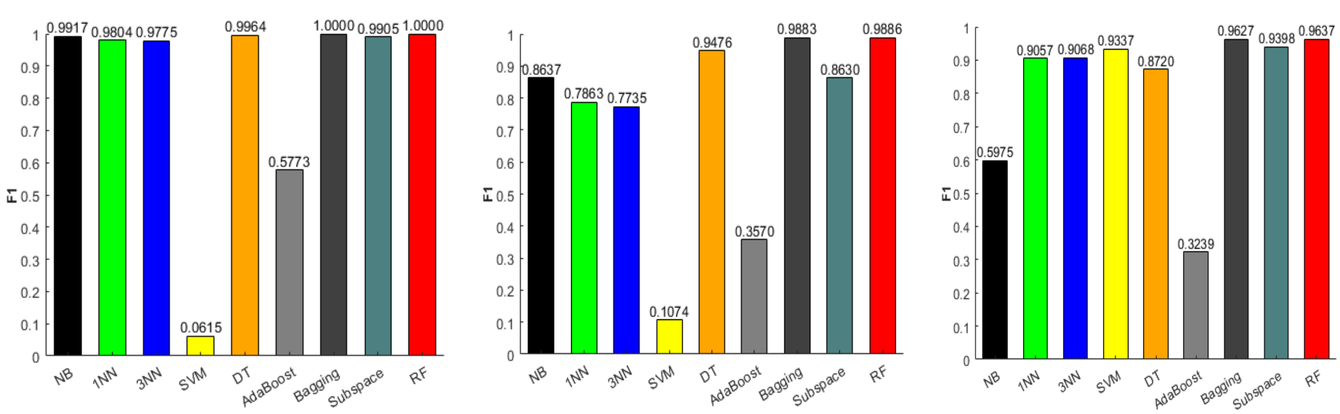
C. Comparision with Dimensionality Reduction Methods

In this subsection, we compare the RF-based activity recognizer with three feature reduction methods (i.e., PCA, reliefF, and mRMR). For PCA, we choose to keep 99.0% of the total variance. Both reliefF and mRMR belong to feature ranking methods, and we experimentally select the top twenty-five features to return an optimal feature subset. Since PCA, reliefF, and mRMR are filter methods, we use a classification model to evaluate the finally obtained feature subset. We herein utilize the widely used NB, 1NN, and SVM. Particularly, for each of the feature reduction methods, we combine it with the models under the framework of Fig. 1. Accordingly, Fig. 4 presents the experimental results on the three datasets. The X-axis indicates different methods and *fs*(*cls*) refers to the use of feature reduction method *fs* and classification model *cls*. The Y-axis shows the accuracy. From Fig. 4, we observe that RF-based activity recognizer performs better than its competitors on all the three datasets. Specifically, the random forest obtains 100%, 98.74%, and 97.16% accuracy on the three datasets, respectively. For the three feature reduction methods, they obtain mixed results. For example, reliefF(SVM) gets the best accuracy of 98.46% on *PAMAP*, mRMR(NB) has the best accuracy of 82.64% on *PAMAP2*, and reliefF(KNN) obtains the best accuracy of



(a) *PAMAP* (b) *PAMAP2* (c) *SkodaMiCP*

Figure 2. Comparison of accuracy with other five individual classifiers and three ensemble methods.



(a) *PAMAP* (b) *PAMAP2* (c) *SkodaMiCP*

Figure 3. Comparison of F1 with other five individual classifiers and three ensemble methods.

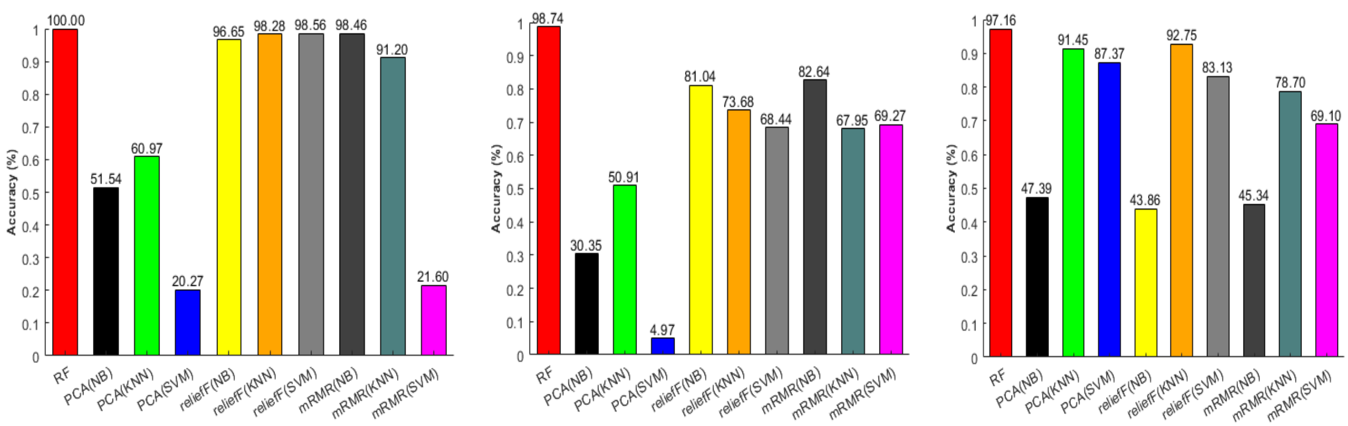
92.75% on *SkodaMiCP*. For PCA, reliefF, and mRMR, the results show the priority of reliefF. Furthermore, according to Fig. 2 and Fig. 4, we see that the use of feature selection and extraction methods fails to consistently obtain improved accuracy. For example, on *PAMAP*, NB obtains 99.17% accuracy, which is higher than the 51.54% of PCA(NB), the 96.65% of reliefF(NB), and the 98.26% of mRMR(NB). For the complex *SkodaMiCP*, SVM gets the accuracy of 94.59% compared with the 87.37% of PCA(SVM), the 83.13% of reliefF(SVM), and the 69.10% of mRMR(SVM). This indicates the necessity for a further study of designing effective feature reduction methods to derive an optimal feature subset.

D. Parameter Sensitivity Analysis

The parameters *numTrees* and *numFeatures* in building a random forest are two crucial parameters in lowering the correlation between individual trees of the forest and they largely influence the generalization ability of a random forest. We herein evaluate their effects by varying the values of *numTrees* and *numFeatures*. The values tested for the *numTrees* parameter are 1, 5, 10, 20, 30, 50, 90, and 120, and the values tested for the *numFeatures* are *mtry*/4, *mtry* /2, *mtry*, 2**mtry*, 4**mtry*, where *mtry* is the square root of the number of features. Fig. 5 presents corresponding results on the three datasets. The X-axis represents the candidate values of *numFeatures*, Y-axis denotes *numTrees*, and Z-axis shows the recognition performance of F1. From Fig. 5, we observe that the two parameters indeed influence the performance of RF-based activity recognizer and that *numTrees* has a greater impact on the performance than *numFeatures*. That is, the activity recognizer is less sensitive to the number of splitting features. From the view of *numTrees*, there is a general trend that the tendency to overfit decreases with the increase of the number of trees, and the use of 50 trees is tradeoff between the time costs and recognition accuracy.

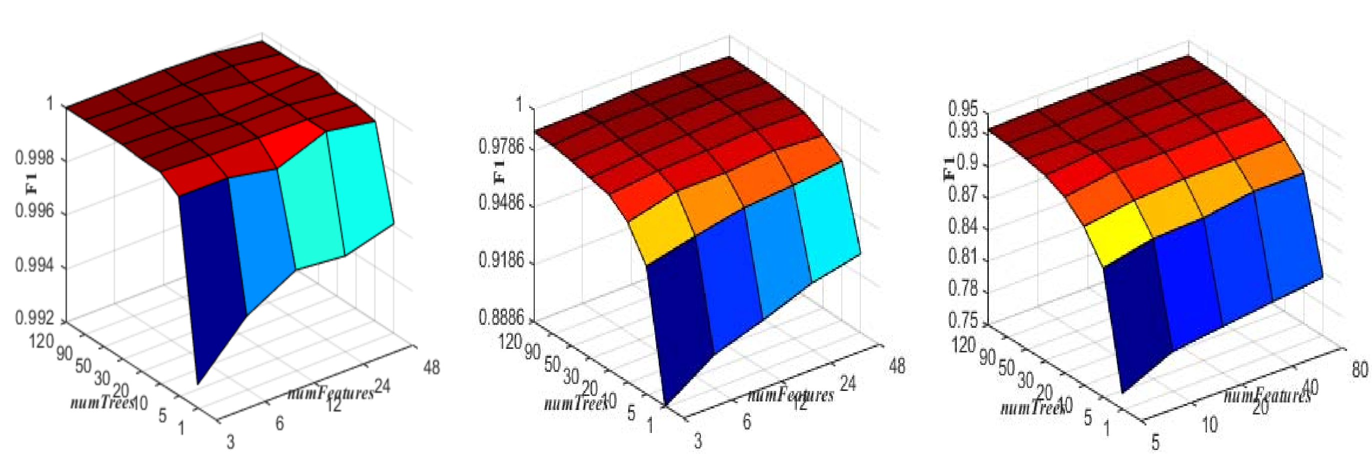
E. Time Costs

In addition to accuracy, time cost in making predictions is another factor where consideration is needed, especially for wearable sensor-based activity recognition systems that have limited resources. We herein investigate the time cost of RF-based activity recognizer and conducted experiments on a computer with a Core i5 3.2GHz CPU and 4G RAM.



(a) *PAMAP* (b) *PAMAP2* (c) *SkodaMiCP*

Figure 4. Comparison of accuracy with two feature selectors and one feature extractor.



(a) *PAMAP* (b) *PAMAP2* (c) *SkodaMiCP*

Figure 5. Parameter sensitivity analysis with regards to the number of trees and number of splitting features.

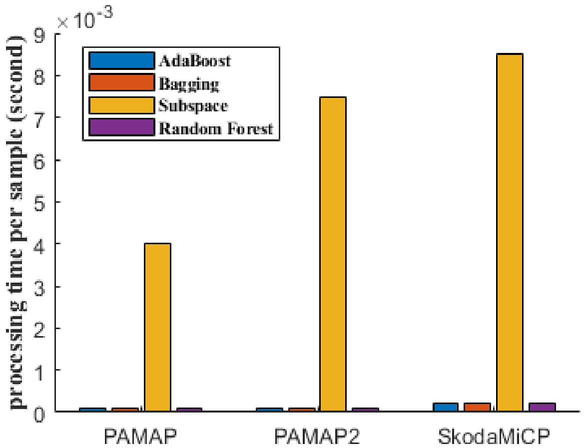


Figure 6. Time cost comparison.

Fig. 6 presents the comparative experimental results with the ensemble models on the datasets. We observe that RF-based activity recognizer has a comparable time cost to the those of AdaBoost and Bagging and can process more than one thousand samples per seconds on all the datasets, which demonstrates its applicability to practical applications. This indicates that the RF remains a priority towards a tradeoff between time cost and accuracy.

V. CONCLUSION

Accurately recognizing human activities plays a central role in a variety of real-world applications that range from smart home and ambient assisted living systems to human-computer interaction and the elderly healthcare. Among the sensing units, the use of wearable sensors remains a priority in building activity recognizers and developing pervasive applications. However, the inherent complexity of human behavior makes it difficult to infer the on-going activities and further requires researchers to design effective and robust activity recognition models with good generalization ability. In this study, we present a wearable sensor-based activity recognizer using the random forest classifier and systematically investigate its performance from multiple aspects. We perform extensive experiments on three publicly available datasets to explore several aspects, such as its comparison with other widely used classifiers and feature selection and extraction methods, and the recommended parameter values. Experimental results show that the RF-based activity recognizer outperforms other five individual classifiers, three ensemble models, two feature selectors, and one feature extractor across different scenarios. Besides, the parameter sensitivity analysis and time cost analysis show its stability and applicability to practical applications. Future work includes its implementation in a wearable device and the exploration of effective feature selection methods. Particularly, different human activities may have a different subset of features to reflect its unique characteristics, and this requires further researches in obtaining the activity label dependent feature subsets [15].

ACKNOWLEDGMENT

This work was supported in part by the Natural Science Foundation of China (Nos. 61902068, 61701159, 61972092), the Key Research and Development Project of Anhui Province (No. KJ2019ZD44), the Major Special Projects of Anhui Province (No. 201903A06020026), the Guangdong Basic and Applied Basic Research Foundation (No. 2020A1515011499), the Anhui Provincial Natural Science Foundation (No. 1908085MF211), and the Research Project of Chuzhou University (No. 2020QD13).

REFERENCES

[1] M. Z. Uddin, M. M. Hassan, A. Alsanad, and C. Savaglio, "A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare," Inform. Fusion, vol. 55 pp. 105-115, Mar, 2020.

[2] 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, pp. 4566-4578, Jun. 2016.

[3] N. C. Krishnan and D. J. Cook, "Activity recognition on streaming sensor data," Pervasive Mob. Comput., vol. 10, pp. 138-154, Feb. 2014.

[4] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," IEEE Commun. Surv. Tut., vol. 15, pp. 1192-1209, Nov. 2012.

[5] Q. Li, R. Gravina, Y. Li, S. H. Alsamhi, F. Sun, and G. Fortino, "Multi-user activity recognition: Challenges and opportunities," Inform. Fusion, vol. 63, pp. 121-135, Nov. 2020.

[6] R. Poppe, "A survey on vision-based human action recognition," Image Vision Comput., vol. 28, pp. 976-990, Jun. 2010.

[7] L. Xu, W. Yang, Y. Cao, and Q. Li, "Human activity recognition based on random forests," Proc. 13th Interna. Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), IEEE Press, Jul. 2017, pp. 548-553.

[8] G. Chen, A. Wang, S. Zhao, L. Liu, and C.-Y. Chang, "Latent feature learning for activity recognition using simple sensors in smart homes," Multimed. Tools Appl., vol. 77, pp. 15201-15219, Jun. 2018.

[9] A. Wang, G. Chen, X. Wu, L. Liu, N. An, and C.-Y. Chang, "Towards human activity recognition: A hierarchical feature selection framework," Sensors, vol. 18, p. 3629, October 2018.

[10] S. Kim, Y. Jeong, and S. Park, "RFID-based indoor location tracking to ensure the safety of the elderly in smart home environments," Pers. Ubiquit. Comput., vol. 17, pp. 1699-1707, Dec. 2013.

[11] Z. Qin, Y. Zhang, S. Meng, Z. Qin, and K.-K. Choo, "Imaging and fusing time series for wearable sensor-based human activity recognition," Informa. Fusion, vol. 53, pp. 80-87, Jan. 2020.

[12] P. Zappi, C. Lombriser, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Tröster, "Activity recognition from on-body sensors: Accuracy-power trade-off by dynamic sensor selection," Proc. European Wireless Sensor Networks (EWSN), Springer, Jan. 2008.

[13] Z. Feng, L. Mo, and M. Li, "A Random Forest-based ensemble method for activity recognition," Proc. 37th Annual International Conference IEEE Engineering Medicine Biology Society (EMBC), IEEE Press, pp. 5074-5077, Aug. 2015.

[14] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," Proc. IEEE Symp. Wearable Computers (ISWC 12), IEEE Press, pp. 108-109, Jun. 2012.

[15] A. Wang, N. An, G. Chen, L. Liu, and G. Alterovitz, "Subtype dependent biomarker identification and tumor classification from gene expression profiles," Knowl.-Based Syst., vol. 146, pp.104-117. Apr. 2018.