Human Activity Recognition from Accelerometer Data: Axis-Wise Versus Axes-Resultant Feature Extraction

Aiguo Wang
School of Electronic Information Engineering, Foshan University, Foshan, China
wangaiguo2546@163.com

Shenghui Zhao
School of Computer and Information Engineering, Chuzhou University, Chuzhou, China
zsh@chzu.edu.cn

Guilin Chen*
School of Computer and Information Engineering, Chuzhou University, Chuzhou, China
glchen@chzu.edu.cn

ABSTRACT

Benefitting from the development of pervasive computing, recent years have witnessed a variety of meaningful human-centric applications, where automating the recognition of human activities plays a central role in bridging the gap between sensing data and high-level services. Accelerometer-based activity recognizer often remains a priority due to its recognition performance, low costs, and portability, however, few studies systematically investigate how to extract and use features from the time-series sensor data and further compare their discriminant power. To this end, we herein propose two different ways of extracting features and exploring their combinations. Specifically, we take as a resultant axis or separate channels the accelerometer axes and then extract axes-resultant and axis-wise features. Afterwards, we evaluate the cases where the two feature sets are used separately or jointly. Finally, we conduct comparative experiments on two public activity recognition datasets with five different classification models in terms of four performance metrics. Results show that the use of axis-wise features outperforms its competitor in the majority across the datasets and that their joint use generally leads to enhanced accuracy.

CCS CONCEPTS

• **Human-centered computing**; • **Ubiquitous and mobile computing**; • **Ubiquitous and mobile computing theory, concepts and paradigms**; • **Ubiquitous computing**;

KEYWORDS

Activity recognition, Accelerometer, Information fusion

ACM Reference Format:
Aiguo Wang, Shenghui Zhao, and Guilin Chen*. 2021. Human Activity Recognition from Accelerometer Data: Axis-Wise Versus Axes-Resultant Feature Extraction. In *The 5th International Conference on Computer Science and Application Engineering (CSAE 2021), October 19–21, 2021, Sanya, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3487075.3487152

CSAE 2021, October 19–21, 2021, Sanya, China

ACM ISBN 978-1-4503-8985-3/21/10…$15.00
https://doi.org/10.1145/3487075.3487152

1 INTRODUCTION

The rapid development of pervasive computing, internet of things, artificial intelligence, and sensor technology facilitates a wealth of human-centric applications that range from smart homes, ambient assisted living, and elderly health care, to human computer interaction, rehabilitation, and smart city [1], where the automatic recognition of human activities plays an essential role in bridging the gap between the raw sensor data and high-level applications [2, 3]. However, since human behavior is typically characterized by the inherent complexity, it remains a critical challenge to develop a robust activity recognizer [4, 5]. For example, different people could perform one activity in different ways (i.e., *inter-subject variation*) and even an individual could do the same activity differently at different places and time (i.e., *intra-subject variation*). Besides, there exist *similar*, *concurrent*, and *interleaved* activities [6].

Accordingly, researchers have explored a variety of sensing units and designed a wealth of recognition models to adapt to different application scenarios. According to the used sensing units, we can group them into vision-, ambient sensor-, and wearable sensor-based methods [2], where wearable sensor-based methods have the advantage of high portability, low costs, robustness to environments, and being suitable for both indoor and outdoor scenarios in contrast to ambient sensor- and vision-based methods. Particularly, the miniature and increasing processing power of sensing units further extends their applicable scope. Commonly used sensing units include WIFI, gyroscope, accelerometer, Bluetooth, and heart rate chip, where accelerometer is perhaps the most commonly used one in implementing an activity recognition supported system. For example, Kwapisz et al. used the cell phone accelerometer to design an activity recognizer [7]. Zappi et al. proposed an accelerometer-based activity recognizer to infer manipulative gestures of a car maintenance assembly-line worker [8].

Furthermore, from the perspective of underlying models, we can broadly group existing activity recognizers into data-driven and knowledge-driven models [2]. The latter typically use expert domain knowledge (such as ontology, semantic map, and logical reasoning) to describe human activities. Hence, they have limited capacity in handling new cases. In contrast, data-driven models use the collected (un)annotated sensor data to train an activity recognizer for classifying test data. Particularly, based on the activity recognition chain, how to extract features from raw sensor data and how to use these features largely determine the performance of an activity recognizer [9]. Even though deep learning model has the capability of automatically learning features from raw data, it demands huge amounts of computational resources and also suffers

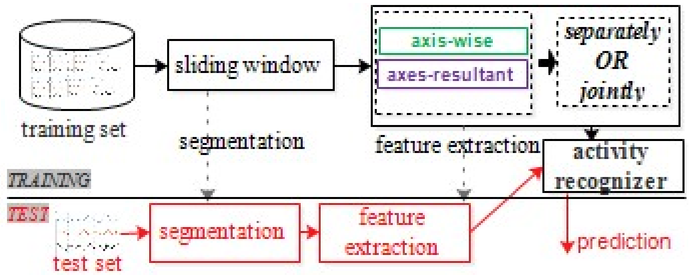

Figure 1: Proposed Activity Recognition Model.

from interpretability [10]. Besides, there are studies demonstrating better accuracy of using hand-crafted features than using deep learning.

Accordingly, researchers have proposed to extract features from different domains (e.g., time domain, frequency domain, as well as wavelet) with the aim to capture the important characteristics of human activities [11]. For example, Kwapisz et al. used *mean*, *standard deviation*, *time between peaks*, *mean resultant acceleration*, *mean absolute difference*, and *binned distribution features*. Except the mean resultant acceleration, they extracted features from each axis and then concatenate them for use [7]. Lübbe et al. calculated the magnitude of the sensor values and obtained the *mean*, *root mean square*, *signal energy*, *standard deviation*, *autocorrelation*, *signal magnitude area*, and *spectral entropy* [11]. Although researchers have done considerable work in obtaining and using features, few studies, to the best of our knowledge, systematically study how to extract features from the multi-axis accelerometer data and further compare their discriminant power from the perspective of information fusion. Therefore, in this study, we propose two different ways of extracting features from raw signals and evaluate their use in activity recognition. The main contributions of this work include: (1) We present two different ways of extracting features from the raw time-series accelerometer readings (i.e., *axis-wise* versus *axes-resultant* methods) and further evaluate their power in the situations where they are used separately or jointly. In addition, we compare the effectiveness of the time-domain and frequency-domain features. These help researchers to better extract and exploit features in optimizing an activity recognizer. (2) We conduct comparative experiments on two activity recognition datasets with five classification models to avoid selection bias. Results show that the use of axis-wise features generally achieves better recognition performance compared to the use of resultant accelerometer features and that their joint use remains a priority in enhancing accuracy.

The rest of this paper is structured as follows. Section 2 illustrates the proposed activity recognition model and how to extract features. Section 3 details the experimental setup and results, followed by the conclusion section.

2 PROPOSED ACTIVITY RECOGNITION MODEL

2.1 Activity Recognition Model

Figure 1 presents the data-driven human activity recognition chain (ARC) that consists of the training stage and prediction stage. In the training stage, we first divide the raw sensor data into segments using the sliding window technique and then extract various features from each segment to return a feature vector for training an activity recognizer *AR*. Obviously, feature extraction plays an important role in ARC and we here explore different ways of extracting and using features, as illustrated in next subsection. During prediction, the test data are first segmented using the same sliding window as the one used in the training stage and then the features are extracted and organized into a feature vector. Finally, the activity recognizer *AR* infers the activity label.

2.2 Feature Extraction

After segmenting the time-series sensor readings, we extract features from each segment to return a feature vector. For an accelerometer with s axes $\{ax_1, ax_2, \ldots, ax_s\}$, we adopt the following steps to analyze the segment.

2.2.1 Axis-wise Feature Extraction. We take as a channel each axis of the accelerometer, extract features from each axis, and then concatenate them for use. Particularly, time-domain and frequency-domain features can be extracted. For time-domain features, we use *median*, *maximum*, *minimum*, *mean*, *standard deviation*, *maximum-minimum*, *median absolute deviation*, *zero crossing rate*, *twenty-five percent quantile*, and *seventy-five percent quantile*. In addition, we use the Pearson correlation coefficient between each two of the s axes.

Afterwards, we apply the fast Fourier transform (FFT) to get the frequency coding of raw signals. We extract the *direct component*, the five *peaks* and corresponding *positions* of the five peaks, *energy*, four *amplitude features* and four *shape features* (i.e., *skewness*, *kurtosis*, *mean*, and *standard deviation*).

2.2.2 Axes-resultant Feature Extraction. We first calculate the resultant axis and then extract features. That is, for a s-axis accelerometer, we obtain the resultant acceleration,

$$a = \sqrt{\sum\nolimits_{i=1}^{s} a_i^2} \tag{1}$$

and then extract time-domain and frequency-domain features as did in the above subsection (but not including the correlation coefficient between two axes).

2.2.3 Feature Fusion. We can combine the axis-wise and axes-resultant features from different views (i.e., different axes) and different domains (i.e., time-domain and frequency-domain), such as the joint use of axis-wise and axes-resultant time-domain features, and the joint use of time-domain and frequency-domain features from the resultant axis.

3 EXPERIMENTAL SETUP AND ANALYSIS

3.1 Experimental Setup

To evaluate the power of different feature sets in training an activity recognizer, we perform comparative experiments on two publicly available activity recognition datasets. The first dataset UCI-HAR was collected by thirty volunteers who carried a waist-mounted smartphone embedded with a tri-axis accelerometer and gyroscope and performed six predefined activities (*sitting*, *standing*, *lying*, *walking*, *go-upstairs*, and *go-downstairs*) [12]. For the purpose of this study, we only use the accelerometer. We divide UCI-HAR data into segments with a 2.56s half-overlap sliding window and extract features from each segment to get a feature vector for subsequent analysis. UCI-HAR contains separate training set and test set, where the former is used to train an activity recognizer and the latter is used to test its power. The second dataset SKODA was collected by tri-axis accelerometers from a car maintenance assembly-line worker [8]. Its task is to recognize ten manipulative gestures (*write on notepad*, *close hood*, *check gaps on the front*, *open hood*, *close left front door*, *close both left doors*, *checking steering wheel*, *open left front door*, *open and close trunk*, and *check trunk gaps*). The sensors have a sampling rate of 96 Hz, and we use one accelerometer placed on the right upper arm and segment the raw sensor data with a 2s half-overlap sliding window. We then extract features from each segment to obtain a feature vector. The five-fold cross validation is used to generate independent training sets and test sets and we report the average results.

As for feature extraction, we extract those features as discussed in subsection 2.2. Five commonly used classification models with different metrics are utilized to train activity recognizers in order to avoid selection bias. Those models include one generative model (naïve Bayes (NB)), two discriminant models (k nearest neighbor with $k = 1$ (KNN) and decision tree (DT)), and two ensemble models (AdaBoost and random forest (RF)). We take as the performance metrics the accuracy (*Acc*), precision (*Prec*), recall (*Rec*), and F1 score, where F1 = 2*precision*recall/(precision + recall) [13, 14].

3.2 Experimental Results

Extensive comparative experimental results of the two activity recognition datasets are given in Tables 1 and 2. The first column refers to the used features, where *time* denotes time-domain, *freq* denotes frequency-domain, *tf* means the time-domain and frequency-domain, *rslt* denotes the resultant axis, *sprt* means axis-wise one, and *comb* refers to the use of both resultant axis and axis-wise features. For example, *rslt_time* means the time domain features of resultant axis, *sprt_freq* stands for the frequency domain features obtained from each axis, and *comb_tf* indicates the joint use of axis-wise and axes-resultant time- and frequency- domain features. For better presentation, we give accuracy, precision, and F1, and results are grouped by the used classification models. Typically, for each classifier, the best accuracy in each domain is bolded and the best accuracy for each classification model is underlined.

Tables 1 and 2 correspond to the results of UCI-HAR and SKODA, respectively, from which we observe that the use of axis-wise features generally obtains higher accuracy than the use of axes-resultant features. For example, on UCI-HAR, RF obtains the accuracy of 75.34% for *rslt_time*, 81.62% for *sprt_time*, and 78.55% for *rslt_freq*, 82.42% for *sprt_freq*. This is mainly because there exists loss of information in obtaining the resultant acceleration from separate axes. Second, we see that the joint use of axis-wise and axes-resultant features generally outperforms their single use. This indicates that the axis-wise features and resultant acceleration features possibly contain complementary information. Third, mixed results are obtained in comparing the power of time-domain features and frequency-domain features. Specifically, the use of axis-wise frequency-domain features outperforms the use of axis-wise time-domain features on UCI-HAR in the majority, while it is the opposite case on SKODA. Fourth, we observe that RF remains a priority in training an activity recognizer compared with NB, KNN, DT, and AdaBoost.

Furthermore, we present the confusion matrix to evaluate the effectiveness of different feature sets in classifying human activities. Due to limited space, we only show the confusion matrix of UCI-HAR where RF is used for illustration purpose, as shown in Figure 2. The key-value pairs are defined as: {1: *walking*, 2: *upstairs*, 3: *downstairs*, 4: *sitting*, 5: *standing*, 6: *lying*}. The columns indicate the true activity labels and the rows are the inferred activity labels. From Figure 2, we observe that the joint use of axis-wise and axes-resultant features generally better discriminates activities having similar sensor readings. Similar results can also be observed for SKODA.

Besides, we compare the proposed method with relevant previous studies. On UCI-HAR, compared with the 88.29% accuracy of hierarchical feature selection [6] and the 94.79% accuracy of deep learning-based model [15], the proposed method obtains the accuracy of 87.07%. Notably, though lower, only the accelerometer is used in our study rather than the use of both accelerometer and gyroscope. On SKODA, compared with the 92.91% accuracy [13], 91.2% accuracy and 93.10% accuracy obtained with deep learning [16], the proposed method obtains 96.30% accuracy, which indicates its power and potentially motivates users to carefully exploit features.

4 CONCLUSIONS

Accelerometer-based activity recognition plays an essential role in a variety of pervasive computing applications such as human computer interaction, healthcare, smart homes, and security, where the extraction and use of features greatly determines the performance of an activity recognizer, which remains a great challenge to current study. Thus, we conduct a systematical study to investigate different ways of extracting features from raw sensor signals (*axis-wise* versus *axes-resultant*) in terms of different views (i.e., *time-domain* and *frequency-domain*) and to further compare their combinations. Finally, comparative experiments are conducted on two public activity recognition datasets. Results show that compared to the use of features from resultant axis, the use of axis-wise features achieves better recognition accuracy in the majority of cases and also show that the joint use of them generally leads to improved performance,

Table 1: Experimental Results on UCI-HAR with Different Features and Classifiers

Features	NB			KNN			DT			AdaBoost			RF		
	Acc	Prec	F1	Acc	Prec	F1	Acc	Prec	F1	Acc	Prec	F1	Acc	Prec	F1
rslt_time	67.02	66.64	67.95	68.29	67.79	68.38	70.73	70.56	70.91	75.94	75.70	76.24	75.34	75.08	75.61
sprt_time	79.65	78.77	79.92	**68.83**	68.26	68.87	79.72	79.42	79.55	82.33	82.07	82.39	81.62	81.34	81.61
comb_time	**80.69**	79.80	80.98	66.62	66.28	66.89	**80.12**	79.79	79.83	**83.53**	83.36	83.64	**83.43**	83.22	83.43
rslt_freq	56.70	57.99	61.65	61.18	61.21	62.09	72.67	72.20	72.60	78.32	77.67	78.24	78.55	77.94	78.56
sprt_freq	77.58	77.02	78.92	79.99	79.48	80.24	81.59	81.02	81.31	82.96	82.74	83.07	82.43	82.13	82.46
comb_freq	**80.09**	79.49	81.21	**80.15**	79.87	80.44	**82.96**	82.59	82.73	**87.64**	87.41	87.59	**87.34**	87.08	87.26
rslt_tf	70.67	70.59	72.09	59.74	60.05	60.94	73.61	73.45	73.78	79.35	79.17	79.63	79.89	79.69	80.12
sprt_tf	81.52	80.85	82.27	79.52	79.11	79.92	81.56	81.39	81.53	86.40	86.22	86.42	86.37	86.20	86.42
comb_tf	**83.70**	83.02	84.25	**80.32**	79.99	80.77	**81.66**	81.28	81.66	**86.80**	86.67	86.88	**87.07**	87.00	87.18

Table 2: Experimental Results on SKODA with Different Features and Classifiers

	NB			KNN			DT			AdaBoost			RF		
	Acc	Prec	F1	Acc	Prec	F1	Acc	Prec	F1	Acc	Prec	F1	Acc	Prec	F1
rslt_time	42.84	48.34	47.30	55.57	57.74	58.67	57.42	59.48	59.78	68.81	70.96	70.85	68.44	70.57	70.55
sprt_time	76.76	78.54	79.12	92.14	92.07	91.97	87.44	86.86	86.87	95.62	95.16	95.09	95.44	94.81	94.77
comb_time	**76.81**	78.83	79.29	**92.75**	92.83	92.70	**87.56**	86.94	87.09	**96.14**	95.69	95.67	**96.17**	95.52	95.48
rslt_freq	41.94	46.47	46.92	29.91	31.64	31.96	62.20	63.96	63.98	72.20	73.39	73.63	72.11	73.17	73.36
sprt_freq	63.13	66.30	67.56	59.85	62.04	62.74	85.33	85.02	85.14	93.37	93.10	93.17	93.11	92.65	92.67
comb_freq	**65.73**	68.47	69.55	**72.31**	72.78	73.15	**85.77**	85.46	85.46	**93.70**	93.27	93.32	**93.92**	93.59	93.63
rslt_tf	45.64	50.24	50.64	29.91	31.64	31.96	65.81	67.24	67.28	78.04	78.70	78.82	78.24	78.91	78.94
sprt_tf	74.60	76.55	77.57	59.85	62.04	62.74	**88.04**	87.53	87.63	96.12	95.78	95.79	95.95	95.44	95.44
comb_tf	**76.08**	77.87	78.87	**72.31**	72.78	73.15	87.64	87.65	87.75	**96.14**	95.70	95.72	**96.30**	95.80	95.82

which indicates that axis-wise and axes-resultant features contain complementary views to a certain extent.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Nos. 61902068, 61972092), Department of Education of Guangdong Province (No. 2021KTSCX117), Anhui Provincial Major Scientific and Technological Special Project (No. 201903A06020026), Key Research and Development Project of Anhui Province (Nos. 201904A05020073, KJ2020A0725, KJ2019ZD44), Science Foundation of Foshan University (No. CGZ0802502), and Research Project of Chuzhou University (No. 2020QD13).

REFERENCES

[1] A Haque, A Milstein and L Fei-Fei (2020). Illuminating the dark spaces of healthcare with ambient intelligence. Nature, 585(7824), 193-202.
[2] A Wang, S Zhao, C Zheng, J Yang, G Chen and CY Chang (2021). Activities of daily living recognition with binary environment sensors using deep learning: A comparative study. IEEE Sensors Journal, 21(4), 5423-5433.
[3] A Bulling, U Blanke and B. Schiele (2014). A tutorial on human activity recognition using body-worn inertial sensors. ACM Computing Surveys, 46(3), 1-33.
[4] A Wang, G Chen, J Yang, S Zhao and CY Chang (2016). A comparative study on human activity recognition using inertial sensors in a smartphone. IEEE Sensors Journal,16(11), 4566-4578.
[5] X Zhou, W Liang, I Kevin, K Wang, H Wang, LT Yang and Q Jin (2020). Deep-learning-enhanced human activity recognition for Internet of healthcare things. IEEE Internet of Things Journal, 7(7), 6429-6438.
[6] A Wang, G Chen, X Wu, L Liu, N An and CY Chang (2018). Towards human activity recognition: A hierarchical feature selection framework. Sensors, 18(11), 3629.
[7] J Kwapisz, G Weiss and S Moore (2011). Activity recognition using cell phone accelerometers. ACM SigKDD Explorations Newsletter, 12(2), 74-82.
[8] 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. Proceeding of European Conference on Wireless Sensor Networks, 2008, pp. 17-33.
[9] S Rosati, G Balestra and M Knaflitz (2018). Comparison of different sets of features for human activity recognition by wearable sensors. Sensors, 18(12), 4189.
[10] A Ignatov (2018). Real-time human activity recognition from accelerometer data using convolutional neural networks. Applied Soft Computing, 62, 915-922.
[11] C Lübbe, B Friedrich, S Fudickar, S Hellmers and A Hein. Feature based random forest nurse care activity recognition using accelerometer data. Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, 2020, pp. 408-413.
[12] D Anguita, A Ghio, L Oneto, X Parra and JL Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. Proceedings of 21st European Symposium on Artificial Neural Networks, 2013, pp. 437–442.
[13] A Wang, S Zhao, C Zheng, H Chen, L Liu and G Chen (2021). HierHAR: Sensor-based data-driven hierarchical human activity recognition, IEEE Sensors Journal, 21(3), 3353-3365.
[14] A Wang, X Wu, L Zhao, H Chen, S Zhao (2021). Physical activity recognition from accelerometer data using multi-view aggregation. Journal of Applied Science and Engineering, 24(4), 611-620.
[15] C Ronao and S Cho (2016). Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications, 59, 235-244.
[16] H Ma, W Li, X Zhang, S Gao and S Lu. AttnSense: Multi-level attention mechanism for multimodal human activity recognition. International Joint Conference on Artificial Intelligence, 2019, pp. 3109-3115.

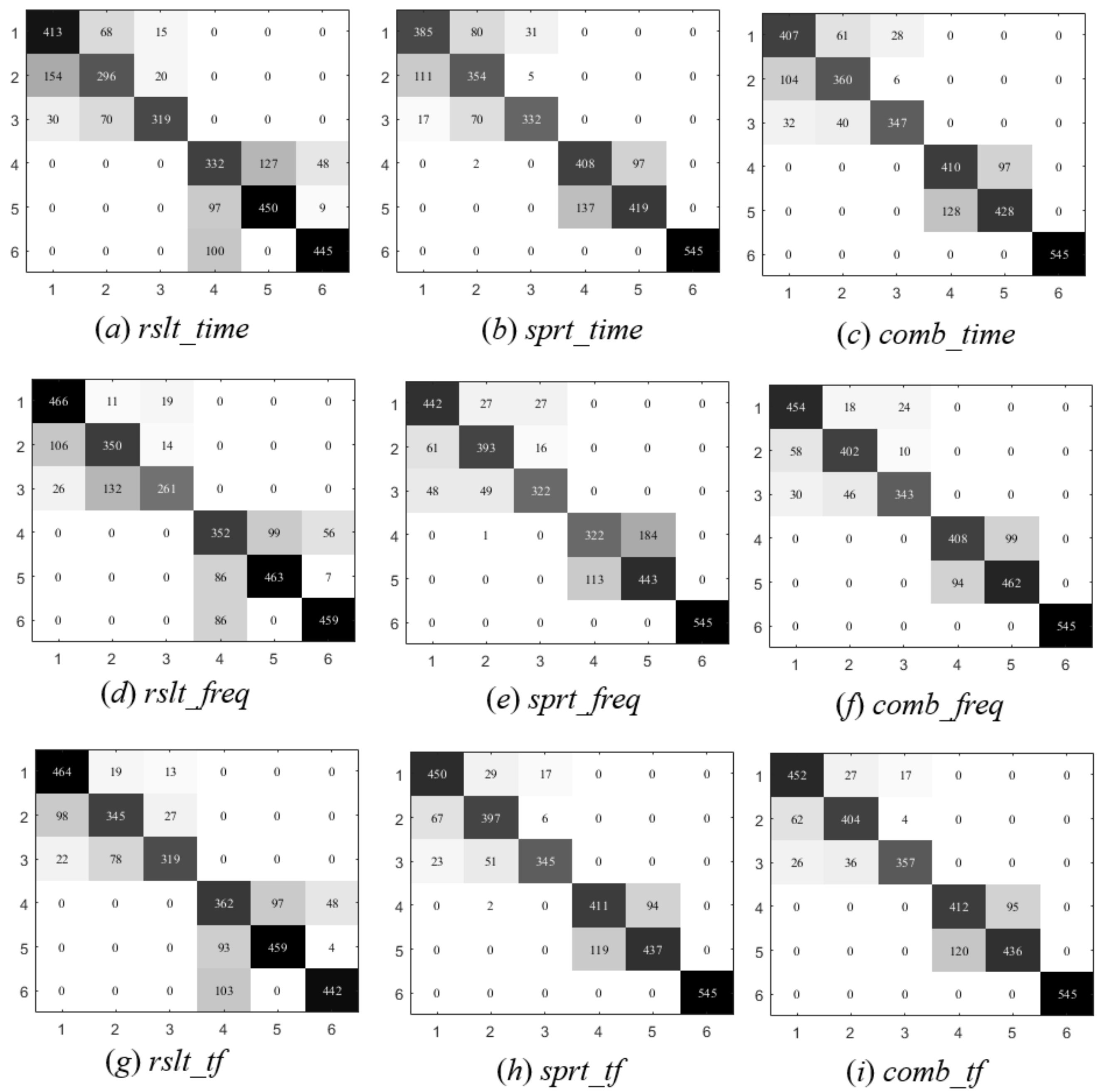

Figure 2: Confusion Matrix on UCI-HAR with Different Types of Features Using RF.