Physical Activity Recognition from Accelerometer Data Using Multi-view Aggregation

Aiguo Wang[1], Xianhong Wu[2], Liang Zhao[3], Haibao Chen[3], and Shenghui Zhao[3,*]

[1] *School of Electronic Information Engineering, Foshan University, Foshan, China*
[2] *Shenzhen Zhiwei Sci-Tech Innovation Co., Ltd., Shenzhen, China*
[3] *School of Computer and Information Engineering, Chuzhou University, Chuzhou, China*
[*] *Corresponding author. E-mail: zsh@chzu.edu.cn*

Received: Jan.21,2021, Accepted: Feb.02,2021

Human physical activities play an essential role in many aspects of daily living and are inherently associated with the functional status and wellness of an individual, therefore, automatically and accurately detecting human activities with pervasive computing techniques has practical implications. Although existing accelerometer-based activity recognition models perform well in a variety of applications, most of them typically work by concatenating features of different domains and may fail to capture the multi-view relationships, resulting in degraded performance. To this end, we present a multi-view aggregation model to analyze the accelerometer data for human activity recognition. Specifically, we extract the time-domain and frequency-domain features from raw time-series sensor readings to obtain the multi-view data representations. Afterwards, we train a first-level model for each view and then unify the models with stacking ensemble into a meta-model. Finally, comparative experiments on three public datasets are conducted against other three activity recognition models. Results indicate the superiority of the proposed model over its competitors in terms of four evaluation metrics across different scenarios.

KEYWORDS: Wearable computing; Activity recognition, Multiview aggregation.

http://dx.doi.org/10.6180/jase.202108_24(4).0016

1. Introduction

It has been known that physical activities are inherently associated with the functional status and wellness of an individual and that accurately automating the recognition of activities plays an important role in effectively bridging the gap between the low-level sensor data and high-level daily living applications that range from human computer interaction, sports and exercise to elderly healthcare, smart home, and ambient assisted living [1, 2]. Accordingly, to adapt to different scenarios, researchers have explored various sensing technologies, which can be grouped into wearable sensor-, environment sensor-, and vision-based methods [3, 4]. Different from vision-based methods relying on computer vision techniques and environment sensor-based methods inferring activities according to the interaction between a person and the surrounding objects, wearable sensor-based methods train an activity recognizer with the collected sensor readings to infer human activities. They have the advantage of being suitable for indoor and outdoor scenarios, high adherence, low cost, and high degree of portability [3].

As for wearable devices, commonly used sensing units mainly include accelerometer, gyroscope, magnetometer, heart rate chip, breath rate chip, and light, among which the accelerometer that measure the acceleration is of the first priority and perhaps the most commonly used sensors in building an activity recognition system. For example, Bao and Intille [5] built an activity recognizer by utilizing five small biaxial accelerometers worn on different parts of the body to collect data and extracting four features (mean, energy, frequency-domain entropy, and correlation) to train an activity recognizer. Ravi et al. [6] explored the use of a tri-axis accelerometer to detect eight activities, where they collected sensor data and extracted four features (mean, standard deviation, frequency-domain entropy, and correlation) to train an activity recognition model. Förster et al. [7] presented an accelerometer-based activity recognition model and applied it to the scenarios of fitness track and gesture recognition. They extracted the time-domain features (mean and variance) based on a sliding window. Lu et al. [8]presented an unsupervised model to recognize activities by extracting nineteen time-domain features from the accelerometers. With the increasing power of a smartphone in processing and communication, they are equipped with sensing units such as accelerometer, gyroscope, and Bluetooth and accordingly provide a convenient way for sensing and data collection. For example, Kwapisz et al. [9] used cell phone accelerometers to recognize six simple activities, where they extracted mean, standard deviation, average absolute difference, average resultant acceleration, time between peaks, and binned distribution features to build an activity recognizer. Dernbach et al. [10]used the accelerometer in a smartphone to infer simple and complex activities. In their study, thirty time-domain features (e.g., mean, min, max, standard deviation, zero-cross, and correlation) were extracted to encode the raw sensor data into a feature vector.

One common feature of these methods is that they either only use time-domain features or simply use the concatenated features of different domains. The former only utilizes partial information, while the latter probably fails to capture the relationships among different views [11]. To this end, we proposed a multi-view aggregation model to analyze the accelerometer data for activity recognition. Specifically, we first extract time-domain and frequency-domain features from the time-series sensor data to obtain the multi-view data representations. Then, we train a first-level model for each view and unify the models with stacking ensemble to get a meta-model. Table 1. summarizes the comparison between the proposed model and related work. Particularly, the main contributions of this study are as follows. (1) We propose a model to utilize multi-view sensor data under the stacking ensemble framework. We detail its main components and present two algebraic ensemble models for comparison purposes. (2) We conduct extensive comparative experiments on three public datasets. Results demonstrate the better generalization of our model over its competitors, including the *single-view*, *view concatenation*, and two algebraic ensemble models, across different scenarios.

2. Theory and formula

Human activity recognition chain (ARC) typically consists of data collection, segmentation, feature extraction, model training, and prediction [12, 13]. Which features to be extracted from raw sensor data largely determines the performance of an activity recognizer. Generally, we can extract different types of features, among which time-domain and frequency-domain features are the most widely and commonly used in previous studies. The two domains represent different data views and we can analyze it under different learning paradigms. For better illustration, let *TD* and *FD* be the time-domain and frequency-domain features and use C = $\{C1, C2, \ldots, C_{|C|}\}$ to denote a label set with $|C|$ different activities.

2.1. Single-view model

For single-view learning, we only take as input one of the two views (i.e., *TD* or *FD*) to train an activity recognizer, as shown in Fig.1(a). Besides, we can concatenate different views and use it to train an activity recognizer, shown in Fig. 1(b). For convenience, we name it as *view concatenation model*. As we discussed in the introduction, this scheme is commonly adopted.

2.2. Multi-view aggregation

For multi-view aggregation, the task is to train a model for each view and combine their results according to the two views *TD* and *FD*. Herein, we present two different aggregation schemes. To mitigate the voting conflict (e.g., view *TD* labels a sample x as activity *sitting*, while view *FD* labels it as activity *lying*), we use the soft voting. That is, in classifying x, a classifier $\mathrm{cls_h}$ outputs a vector *h*(x), which is the posterior probability output. Specifically,

$$h(x) = [cls_h^1(x), cls_h^2(x), \ldots, cls_h^{|c|}(x)]^T \tag{1}$$

where

$$cls_h^i(x) = p(C_i|x) \in [0,1], 1 \leq i \leq |C|, \sum_{k=1}^{|c|} cls_h^K(x) = 1$$

Obviously, we can combine the results of each view with algebraic combiners such as max and mean, named the *algebraic ensemble model* in Fig.1(c). We here use the maximum and average to make predictions *H(x)*.

$$H(x) = C_{\underset{1 \leq i \leq |C}{\arg\max}\{\max\{cls_{TD}^i(x), cls_{FD}^i(x)\}\}} \tag{2}$$

Table 1. Comparisons between the proposed method and related work.

Studies	Time domain	Frequency domain	Multiview
Bao et al.[5]	√	√	view concatenation
Ravi et al. [6]	√	√	view concatenation
Förster et al.[7]	√	×	single-view
Lu et al. [8]	√	×	single-view
Kwapisz et al. [9]	√	√	view concatenation
Ours	√	√	multi-view

$$H(x) = C_{\underset{1\le i\le |C|}{\mathrm{argmax}}\left\{\frac{1}{|\{TD,FD\}|}\sum_{h\in\{TD,FD\}} \mathrm{cls}_h^i(x)\right\}} \tag{3}$$

where $cls^i_{TD}(x)$ denotes the model trained on TD.

Besides, we can take as input the results obtained on each view to train another model, as shown in Fig. 1(d). Specifically, we first train a first-level classifiers for each view and then train a second-level classifier (also called meta-classifier) with the concatenation outputs of first-level classifiers. For convenience, we call it stacking ensemble model.

3. Experimental setup

To evaluate the proposed model, we conduct extensive comparative experiments on three public activity recognition datasets. The first dataset BAFitness contains the sensor readings of six activities (i.e., *flick kicks, knee lifts, jumping jacks, superman jumps, high knee runs, and feet back runs*) that were collected with ten tri-axis accelerometers working at a sample rate of 64Hz and placed on the right leg [7]. For each activity, about thirty seconds of sensor data were collected. A constant eight-second sliding window with two thirds overlap between consecutive segments is used to divide the time-series data. The task of the second dataset HCI is to recognize *triangle pointing up, square, circle, infinity, and triangle pointing down*. The sensor data were collected with eight accelerometers under a 96Hz sample rate. The raw data were manually segmented to contain a single activity [7]. For the two datasets, we only use one accelerometer for experiments as did in [7], which is sufficient for the application. The third dataset WISDM contains sensor signals of *walking, jogging, upstairs, downstairs, sitting*, and *standing* that were collected with a tri-axis accelerometer in a smartphone. The sample rate is 20Hz and we divide the time series sensor data with a ten-second sliding window without overlap between consecutive segments [9].

Afterwards, we extract various features from each segment to form a feature vector for subsequent activity recognition model training. Specifically, according to [14], for a tri-axis accelerometer with readings a_x, a_y and a_z from the axes X, Y, and Z, we obtain the resultant acceleration $a = \sqrt{a_x^2 + a_y^2 + a_z^2}$ and then extract features from a. For time-domain features, we use *average, standard deviation, mode* (the value that occurs most frequently), *maximum, minimum, range* (difference between *maximum* and *minimum*), and *mean crossing rate* features. For frequency-domain features, we apply the fast Fourier transform (FFT) algorithm on a to transform it into frequency domain and then extract the *direct component*, the *first five peaks, frequencies of the five peaks, energy*, four *shape* features (*mean, standard deviation, skewness*, and *kurtosis*), and four *amplitude* features. In total, there are twenty-seven features from the time domain and frequency domain.

The models shown in Fig. 1 are general frameworks, so we can take as the building blocks different classification models. Herein, we adopt four models that have different metrics, i.e., naïve Bayes (NB), k nearest neighbor (KNN), decision tree (DT), and support vector machine (SVM). They are commonly used in previous studies [3, 15, 16]. Particularly, for the stacking ensemble model, different models can be used in the first-level learner and meta-learner. We call it the *homogeneous* mode in the case of the same model and name it as *heterogeneous* mode in the case of different models. In this study, we use SVM in the meta-learner because of its discrimination capability and use other classifiers at the first level. We also evaluate the combination of different models and present corresponding results in the following section.

To create independent training set and test set, we use stratified three-fold cross validation on BAFitness and HCI, where two thirds of the data are used to train an activity recognizer. For WSIDM, we randomly partition it into training set and test set with the ratio of 7:3 and repeat the process ten times. Finally, we report the mean results. *Accuracy (Acc), precision (Prec), recall (Rec)*, and *F1* are used as performance metrics. For illustration purpose, Table 2 presents an ex-

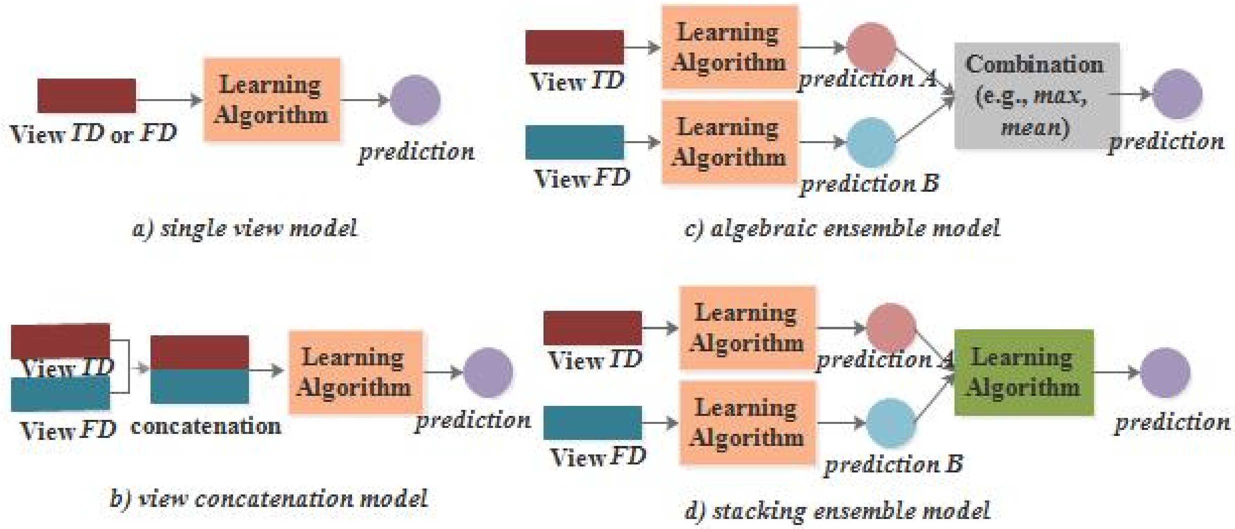
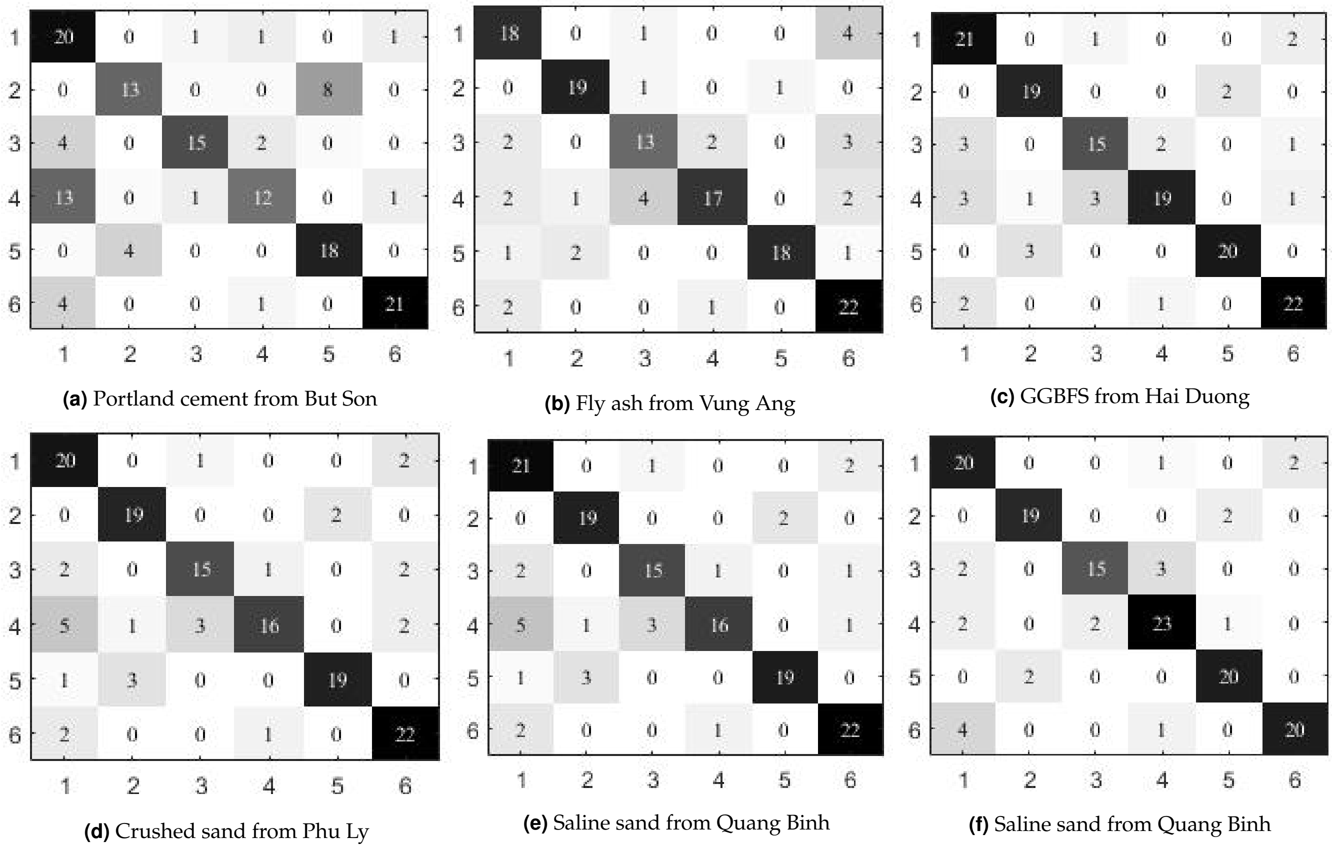


Fig. 1. Activity recognition models with different learning schemes. (a) *single-view model*; (b) *view concatenation model*; (c) *algebraic ensemble model*; (d) *stacking ensemble model*.



(a) Portland cement from But Son

(b) Fly ash from Vung Ang

(c) GGBFS from Hai Duong

(d) Crushed sand from Phu Ly

(e) Saline sand from Quang Binh

(f) Saline sand from Quang Binh

Fig. 2. Confusion matrix on BAFitness of the six models.

emplar confusion matrix of three classes for calculating the metrics.

Precision is the weighted average of the correctly classified sample for each class.

Table 2. Confusion matrix of three classes.

		True labels			
		C1	C2	C3	sum
Predicted labels	C1	T11	FP12	FP13	NP1
	C2	FP21	T22	FP23	NP2
	C3	FP31	FP32	T33	NP3
	sum	NT1	NT2	NT3	total

$$\text{Precision} = \frac{1}{|C|}\sum_{i=1}^{|C|}\frac{T_{ii}}{NP_i} \tag{4}$$

where $\mathrm{T_{ii}}$ is the number of samples from class $\mathrm{C_i}$ that are correctly classified, and NP_{i} is the number of samples predicted with class $\mathrm{C_i}$. *Recall* is the percentage of correctly retrieved samples for each class.

$$\text{Recall} = \frac{1}{|C|}\sum_{i=1}^{|C|}\frac{T_{ii}}{NT_i} \tag{5}$$

where NT_{i} equals the number of instances from class C_{i}. $F1$ is the harmonic mean of *precision* and *recall*.

$$F1 = \frac{2^{*}\ \text{precision}\ ^{*}\ \text{recall}}{\text{precision}\ +\ \text{recall}} \tag{6}$$

4. Result discussions

Tables 3 4 5 present the experimental results on the three datasets. We use "*TDV*", "*FDV*" and "*TFDV*" to denote the recognizers that are trained on the time-domain, frequency-domain, and concatenation views, respectively. "*EnMax*" and "EnAvg" correspond to the algebraic ensemble models with maximum and mean aggregators, respectively. "*stacking*" refers to the proposed stacking ensemble model. The results are organized by the used first-level model. and the best results of each group are shown in bold and the second best is underlined.

From Tables 3 to 5, we see that the fusion of time-domain and frequency-domain features obtains better performance than that of only using time-domain or frequency-domain features in the majority of cases in terms of accuracy, precision, recall, and F1. For example, in the case of NB on BAFitness, *TFDV* obtains the accuracy of 82.14% compared to the 70.24% accuracy of *TDV* and 76.91% accuracy of *FDV*. This indicates that the two domains contain information that is complementary to each other. However, we also see that concatenating the two views may get degraded performance. For instance, *TFDV* only obtains 16.67% accuracy compared to the 41.19% accuracy of *TDV* when using SVM on BAFitness. This is possibly because the view concatenation may confound the characteristics of each view and result in information redundancy. As for *TDV* and FDV, they obtain mixed results. For example, on WISDM, *FDV* outperforms *TDV* when using NB, KNN and DT, but has lower accuracy with SVM. Second, we can observe that the proposed stacking-based activity recognizer generally performs better than its competitors and generalizes better across classifiers and datasets. For example, on HCI, *TFDV* obtains the best results in the case of NB, but it performs worst with SVM. Investigating the results on BAFitness, HCI and WISDM, we observe the robustness of Stacking. Third, as for the ensemble models, we observe that *stacking* generally outperforms *EnMax* and *EnAvg*. This is possibly because statistically aggregating the results of different views may fail to capture the latent relationships among views.

Afterwards, we present the confusion matrix to investigate the performance improvement of the proposed model. We here only show the results on BAFitness for illustration purposes. Fig. 2 presents the results of NB. The rows give the actual labels and the columns indicate the predicted results. From Fig. 2, we observe that the stacking-based activity recognition model tends to obtain better performance than its competitors.

Finally, we investigate the use of different classifiers in the stacking model, where NB, KNN, DT, and SVM can be used at the first level and also the second level. Tables 6-8 show the corresponding results in terms of accuracy, precision, recall, and F1 on the three experimental datasets, respectively, where the results are grouped by the used first-level classifier. The best results within each group are shown in bold. From Tables 6-8, we observe that the use of SVM at the second level generally performs better, which is consistent with the previous results.

5. Conclusions

Accurately automating the recognition of human physical activities plays key roles in daily living and effectively bridges the gap between the sensor data and high-level applications that range from human computer interaction and sports and exercise to elderly healthcare, smart home, and ambient assisted living. Accordingly, researchers have explored various sensing technologies to adapt to different scenarios, among which the accelerometer is of the priority and the most commonly used in building a wearable sensor-based activity recognition system. In the activity recognition chain, how to utilize the extracted features largely determines the performance of an activity recognizer. To this end, we here present a multi-view aggregation model to analyze the accelerometer data for activity recognition. Specifically, we first extract time-domain and frequency-domain features to obtain the multi-view data and train a model for each view and then unify the models with stacking ensemble. We compare the proposed model with the *single-view, view concatenation*, and *algebraic ensemble* models on three datasets in terms of four performance metrics. Results demonstrate the superiority of the proposed model over its competitors across classification models and datasets.

For the future work, we plan to conduct researches along the following lines. First, one limitation of this study is that we evaluated the model in an offline way, which may be different from the online real-time cases. This motivates us to conduct further study. Second, we here only consider time-domain and frequency-domain features. Other features such as the time-frequency-domain features and deep learning learned features can also be used. Third, besides accelerometer, there are other sensors such as gyroscope, sound and light available for use. However, cross-modal differences pose a challenge for heterogeneous data analyses [17]. Hence, how to fuse multi-modal multi-view sensor data requires further study.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (No. 61902068), Anhui Provincial Major Scientific and Technological Special Project (No. 201903A06020026), Key Research and Development Project of Anhui Province (No. KJ2019ZD44), Foshan Self-funded ST Innovation Plan (No. 1920001001001), Program of Natural Science Foundation of Department of Education of Anhui Province (No. KJ2018B05, KJ2019A0647), and Research Project of Chuzhou University (No. 2020QD13).

Table 3. Recognition performance of the six models on BAFitness.

Model	NB				KNN				DT				SVM			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
TDV	70.24	70.89	75.79	73.19	74.76	74.13	74.98	74.55	**90.00**	**89.67**	90.34	90.00	41.19	42.32	40.52	41.40
FDV	76.91	77.06	79.18	78.10	46.67	46.01	46.53	46.26	87.14	87.18	87.66	87.42	16.67	16.67	16.67	16.67
TFDV	82.14	82.37	83.96	83.15	46.67	46.01	46.53	46.26	88.10	87.97	88.76	88.36	16.67	16.67	16.67	16.67
EnMax	79.52	80.06	81.61	80.82	69.05	68.32	69.51	68.91	89.29	89.39	89.68	89.53	19.52	16.67	19.52	17.98
EnAvg	79.76	80.30	82.14	81.20	71.19	70.60	71.80	71.20	89.29	89.39	89.68	89.53	19.52	16.67	19.52	17.98
stacking	**84.76**	**84.83**	**86.18**	**85.49**	**77.14**	**76.33**	**77.28**	**76.80**	89.76	89.60	**90.69**	**90.14**	**81.19**	**80.75**	**82.78**	**81.74**

Table 4. Recognition performance of the six models on HCI.

Model	NB				KNN				DT				SVM			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
TDV	65.15	65.19	66.36	65.76	31.44	31.57	28.22	29.7	57.96	57.87	59.53	58.66	62.50	62.64	66.48	64.48
FDV		87.12	87.12	87.65	87.39	45.08	45.14	46.51	45.79	84.47	84.40	85.36	84.88	32.58	32.94	30.22
31.26																
TFDV	**87.88**	**87.87**	**88.43**	**88.15**	45.08	45.14	46.51	45.79	84.47	84.40	85.37	84.88	32.58	32.94	30.22	31.26
EnMax	87.12	87.10	87.51	87.31	44.70	44.73	44.00	44.35	77.65	77.58	78.15	77.86	54.92	55.05	55.40	55.17
EnAvg	87.12	87.10	87.54	87.32	45.83	45.86	45.90	45.88	77.65	77.58	78.15	77.86	63.26	63.40	64.31	63.85
stacking	86.74	86.73	87.27	87.00	**47.35**	**47.39**	**48.05**	**47.71**	**87.50**	**87.39**	**87.99**	**87.69**	**69.32**	**69.30**	**71.47**	**70.36**

Table 6. Results on BAFitness of the combination of different classifiers.

Metric	NB				KNN				DT				SVM			
	NB	KNN	DT	SVM	NB	KNN	DT	SVM	NB	KNN	DT	SVM	NB	KNN	DT	SVM
Acc	34.76	84.29	83.10	**84.76**	75.00	77.14	70.71	**77.14**	35.00	74.05	85.71	**89.76**	71.91	65.00	70.95	**81.19**
Prec	32.28	84.36	83.02	**84.83**	74.08	**76.86**	70.41	76.33	35.46	74.72	85.72	**89.60**	71.87	65.55	70.93	**80.75**
Rec	77.21	85.48	83.41	**86.18**	75.52	**78.21**	71.48	77.28	47.25	75.91	88.03	**90.69**	74.32	69.91	72.89	**82.78**
F1	45.50	84.91	83.21	**85.49**	74.79	**77.53**	70.93	76.80	38.09	75.31	86.85	**90.14**	73.07	67.65	71.89	**81.74**

Table 5. Recognition performance of the six models on WSIDM.

Model	NB				KNN				DT				SVM			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
TDV	74.43	62.93	65.48	64.18	68.67	57.84	58.50	58.17	78.18	68.07	68.62	68.34	73.39	**59.07**	62.79	**60.81**
FDV	85.59	75.30	76.67	75.98	79.14	67.84	68.30	68.07	84.39	74.61	75.02	74.82	35.71	35.94	36.10	35.91
TFDV	85.45	75.23	76.58	75.90	79.23	**67.93**	68.45	**68.19**	84.00	73.80	74.38	74.09	38.29	43.22	43.92	43.42
EnMax	85.06	74.01	76.44	75.2	1 76.75	60.18	**70.96**	65.12	83.79	72.18	74.95	73.54	**74.52**	54.01	62.20	57.73
EnAvg	85.29	74.27	76.63	75.43	76.75	60.18	**70.96**	65.12	83.88	72.31	75.02	73.64	73.83	51.63	64.60	57.30
stacking	**86.11**	**76.43**	**77.62**	**77.02**	**79.43**	67.91	68.77	**68.33**	**87.68**	**78.96**	**79.90**	**79.43**	73.39	**59.07**	62.79	**60.81**

Table 7. Results on HCI of the combination of different classifiers.

Metric	NB				KNN				DT				SVM			
	NB	KNN	DT	SVM	NB	KNN	DT	SVM	NB	KNN	DT	SVM	NB	KNN	DT	SVM
Acc	35.99	**86.74**	82.2	**86.74**	15.15	43.56	41.67	**47.35**	46.59	85.23	80.68	**87.50**	57.96	62.12	60.23	**69.32**
Prec	36.25	**86.78**	82.07	86.73	15.21	43.60	41.68	**47.39**	46.73	85.10	80.57	**87.39**	58.04	62.07	60.33	**69.30**
Rec	77.43	87.04	83.04	**87.27**	20.50	44.57	43.20	**48.05**	52.37	85.88	81.91	**87.99**	59.57	62.99	61.41	**71.47**
F1	49.26	86.91	82.55	**87.00**	17.34	44.07	42.42	**47.71**	49.31	85.49	81.23	**87.69**	58.79	62.52	60.86	**70.36**

Table 8. Results on WSIDM of the combination of different classifiers.

Metric	NB				KNN				DT				SVM			
	NB	KNN	DT	SVM	NB	KNN	DT	SVM	NB	KNN	DT	SVM	NB	KNN	DT	SVM
Acc	44.58	82.19	83.89	**86.11**	70.63	75.34	79.37	**79.43**	40.46	78.95	80.55	**87.68**	65.94	70.32	69.66	**78.11**
Prec	29.58	70.28	72.13	**76.43**	43.49	58.41	67.10	**67.91**	42.29	66.22	69.36	**78.96**	52.84	56.64	**58.58**	58.01
Rec	50.04	70.72	72.66	**77.62**	56.66	60.90	69.16	**68.77**	45.35	66.57	69.63	**79.90**	55.40	57.14	59.02	**74.43**
F1	36.99	70.49	72.39	**77.02**	49.04	59.57	68.11	**68.33**	43.54	66.39	69.49	**79.43**	54.07	56.89	58.79	**65.17**

References

[1] Jing Yu, Hang Li, Shou Lin Yin, Qingwu Shi, and Shahid Karim. Dynamic gesture recognition based on deep learning in human-to-computer interfaces. *Journal of Applied Science and Engineering*, 23(1):31–38, 2020. ISSN 15606686.

[2] Aiguo Wang, Shenghui Zhao, Chundi Zheng, Jing Yang, Guilin Chen, and Chih Yung Chang. Activities of Daily Living Recognition with Binary Environment Sensors Using Deep Learning: A Comparative Study. *IEEE Sensors Journal*, 21(4):5423–5433, 2021. ISSN 15581748.

[3] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*, 46(3), jan 2014. ISSN 03600300.

[4] Aiguo Wang, Shenghui Zhao, Chundi Zheng, Huihui Chen, Li Liu, and Guilin Chen. HierHAR: Sensor-Based Data-Driven Hierarchical Human Activity Recognition. *IEEE Sensors Journal*, 21(3):3353–3365, 2021. ISSN 15581748.

[5] Ling Bao and Stephen S. Intille. Activity recognition from user-annotated acceleration data. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3001:1–17, 2004. ISSN 16113349.

[6] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael Littman. Activity recognition from accelerometer data. In *17th Conference on Innovative Applications of Artificial Intelligence*, pages 1541–1546, 2005.

[7] Kilian Förster, Daniel Roggen, and Gerhard Tröster. Unsupervised classifier self-calibration through repeated context occurences: Is there robustness against sensor displacement to gain? In *Proceedings - International Symposium on Wearable Computers, ISWC*, pages 77–84, 2009. ISBN 9780769537795.

[8] Yonggang Lu, Ye Wei, Li Liu, Jun Zhong, Letian Sun, and Ye Liu. Towards unsupervised physical activity recognition using smartphone accelerometers. *Multimedia Tools and Applications*, 76(8):10701–10719, 2017. ISSN 15737721.

[9] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, 12(2):74–82, mar 2011. ISSN 1931-0145.

[10] Stefan Dernbach, Barnan Das, Narayanan C. Krishnan, Brian L. Thomas, and Diane J. Cook. Simple and complex activity recognition through smart phones. In *Proceedings - 8th International Conference on Intelligent Environments, IE 2012*, pages 214–221, 2012. ISBN 9780769547411.

[11] Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43–54, 2017. ISSN 15662535.

[12] Aiguo Wang, Guilin Chen, Jing Yang, Shenghui Zhao, and Chih Yung Chang. A Comparative Study on Human Activity Recognition Using Inertial Sensors in a Smartphone. *IEEE Sensors Journal*, 16(11):4566–4578, 2016. ISSN 1530437X.

[13] Zimin Xu, Guoli Wang, and Xuemei Guo. Sensor-based activity recognition of solitary elderly via stigmergy and two-layer framework. *Engineering Applications of Artificial Intelligence*, 95, 2020. ISSN 09521976.

[14] Lisha Hu, Yiqiang Chen, Jindong Wang, Chunyu Hu, and Xinlong Jiang. OKRELM: online kernelized and regularized extreme learning machine for wearable-based activity recognition. *International Journal of Machine Learning and Cybernetics*, 9(9):1577–1590, sep 2018. ISSN 1868808X.

[15] Haodong Guo, Ling Chen, Yanbin Shen, and Gencai Chen. Activity recognition exploiting classifier level fusion of acceleration and physiological signals. In *UbiComp 2014 - Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 63–66. Association for Computing Machinery, Inc, 2014. ISBN 9781450330473.

[16] Aiguo Wang, Ning An, Guilin Chen, Lian Li, and Gil Alterovitz. Accelerating wrapper-based feature selection with K-nearest-neighbor. *Knowledge-Based Systems*, 83(1): 81–91, 2015. ISSN 09507051.

[17] David Stromback, Sangxia Huang, and Valentin Radu. Mm-fit Multimodal deep learning for automatic exercise logging across sensing devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4 (4), dec 2020. ISSN 24749567.