

Increasing accuracy of classification physical activity based on smartphone using ensemble logistic regression with boosting method

by Lawi Armin

Submission date: 19-Feb-2020 11:51AM (UTC+0700)

Submission ID: 1259987110

File name: Lawi_2019_J._Phys.__Conf._Ser._1341_042002.pdf (873.87K)

Word count: 3370

Character count: 18558

PAPER • OPEN ACCESS

Increasing accuracy of classification physical activity based on smartphone using ensemble logistic regression with boosting method

To cite this article: A Lawi ⁹ *et al* 2019 *J. Phys.: Conf. Ser.* **1341** 042002

View the [article online](#) for updates and enhancements.

**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices
to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of
every title for free.

Increasing accuracy of classification physical activity based on smartphone using ensemble logistic regression with boosting method

A Lawi¹, F Aziz² and S L Wungo³

¹Department of Computer Science, Universitas Hasanuddin Makassar, Indonesia. 90245

²Faculty of Mathematics and Natural Sciences, Universitas Pancasakti Makassar, Indonesia. 90132

³Program Study Informatics Engineering, STMIK Kreatindo Monokwari, Indonesia. 98300

E-mail: firmanaziz88@gmail.com

Abstract. At present, the smartphone is equipped with several sensors such as Accelerometer, Gravity sensor, and Gyroscope which can be used to recognize human physical activities such as walking upstairs and walking downstairs etc. Machine learning is needed to group data and get information. Statistical methods have poor performance in classifying because procedures must be met. To overcome this, an ensemble technique was used. This study proposes the application of the gradientboost ensemble method to classify walking upstairs and walking downstairs. The Android-based .apk system is designed for data retrieval using a smartphone. Then, the dataset will be partitioned into 70% training data and 30% test data. The results show that the performance of the ensemble boosting method produces 81.82% accuracy, 86.11% sensitivity and 77.50% specificity.

1. Introduction

Physical activity is a body movement that uses energy such as walking, running, walking upstairs and walking downstairs, the benefits of which are very good for health. Currently, walking upstairs and walking downstairs is very popular for maintaining health and is very easy to adopt because it is very suitable for modern urban life which consumes almost all of its activities in the room. Its application is very convenient in everyday life and just need to make it a habit to be repeated in order to obtain significant health benefits. According to stepjockey, an increase in heart rate by walking upstairs and walking downstairs can help protect against high blood pressure, weight gain and blockage of blood vessels. In addition, it also trains bones and muscles to increase strength, bone density, and muscle tone. Current technological developments allow a system to classify human physical activity using cameras, sensors smartphones. smartphones that have an Accelerometer, Gravity sensor and Gyroscope can be used to recognize human physical activity by placing the smartphone on the head, chest, arms and thighs. to recognize human physical activity data produced by sensors can use data mining and data analytics which is a computational process by understanding patterns in data sets using methods such as artificial intelligence, machine learning, statistics etc. [1].

Muhammad Shoaib, et al. Analyzed the role of accelerometer, gyroscope, and magnetometer sensors on smartphones in the introduction of activity. four body positions were evaluated using seven classifiers namely Naive Bayes, Support Vector Machines, Neural Networks, Logistic Regression, Nearest Neighbor, Rule Based Classifiers, and Decision Trees to recognize six physical activities. The overall



Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

results of the proposed method show that accelerometers and gyroscopes complement each other in the process of recognition of activity. Classifiers with the best performance in all activities are Neural Networks and Decision Trees [2]. Aiguo Wang, et al. Evaluated the triaxial accelerometer and gyroscope built-in smartphone in recognizing human physical activity using K-Nearest Neighbors and Naive Bayes. The results show that the combination of accelerometer and gyroscope data contributes to better recognition performance than using single source data [3]. Warren Triston and Kavitha proposed six classifiers namely K-Nearest Neighbors, Naive Bayes, Support Vector Machines, Conditional Inference Tree, J48 and Random Forest for activity recognition. The results show that the best performing classifiers are Random Forest, Naïve Bayes and J48 [4]. Ankita Jain and Vivek Kanhangad analyzed the performance of multiclass Support vector machines and K Nearest Neighbor. Both methods are evaluated in two datasets namely, UCI HAR dataset and physical activity sensor data. The results show that the two methods provide the best performance with an average classification accuracy of 96.83% [5]. Johan Wannenburg and Reza Malekian proposed seven classifiers namely Support vector machines, Multilayer Perceptron, Naive Bayes, K Nearest Neighbor, Bagging, J48, and kStar to find the best overall performance and the best performance for certain classes of activities. The results show that the classifiers with the best performance are Nearest Neighbor and kStar [6].

Based on several studies concerning the best performance average physical human activity when using machine learning methods with excellent performance results. In contrast to statistical methods that provide poor results due to the existence of procedures that must be met such as the number of populations must have the same variant, measurement of variables in the interval scale, and linear data. But with the development of research in recent years the weaknesses of the statistical method can be overcome by performing ensemble techniques [7]. This study proposes a classification of human physical activity. not only using the accelerometer and gyroscope, this research adds a sensor that is a gravity sensor to maintain gravity in changing the orientation of the smartphone. The focus of this research is to conduct ensemble boosting techniques by combining logistic regression as the initial classification and gradientboost to correct classification errors from logistic regression [8].

2. Literature Review

2.1. Accelerometer

Accelerometer is a device that serves to measure the right acceleration by identifying the phenomenon of heavy acceleration experienced by the test mass on an accelerometer device [9] but the accelerometer does not have the accuracy of coordinates when measuring precise acceleration [10]. The accelerometer sensor on a smartphone has a function to determine the degree of slope of a smartphone that performs three axis readings from different directions. Figure 1. Illustration of an accelerometer sensor on a smartphone.

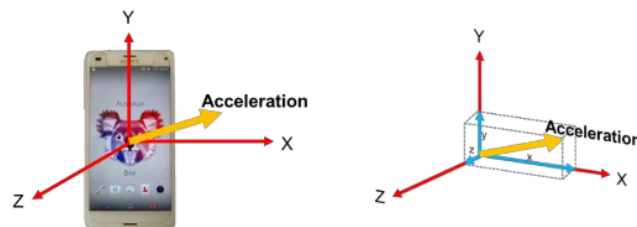


Figure 1. Accelerometer vector and axes

3.2. Gravity Sensor

The gravity sensor is a virtual sensor that comes from a 3-axis acceleration sensor. The 3-axis gravitational component provides a measure of the effects of Earth's gravity observed on the device's reference axis.

The gravitational components measured on the device vary based on changes in the orientation of the device, and hence provide the size of rotation that is targeted by the device [11].

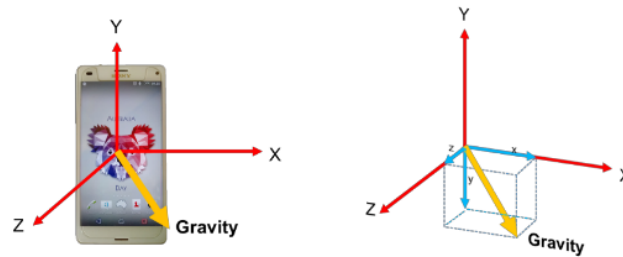


Figure 2. Gravity sensor vector and axes

Gravity sensor displays 4 values: 3 values of Cartesian axis and time stamp. The gravity sensor measures and returns the axis value in "m/s²" (meters per second squared). When the device is rotated towards ± X, ± Y, or ± Z, the corresponding output will increase (+) or decrease (-).

2.3. Gyroscope

Gyroscopes are devices for measuring orientation, based on angular momentum [12]. Gyroscope has a working principle when rotating clockwise on the Z axis, the output voltage will be smaller (-Z), whereas when rotating counterclockwise on the Z axis, the output voltage will increase (+Z) and when not rotate or be at rest, the output voltage will match the offset value of the gyrosensor. Figure 3. Shows the illustration of the gyroscope sensor on a smartphone.

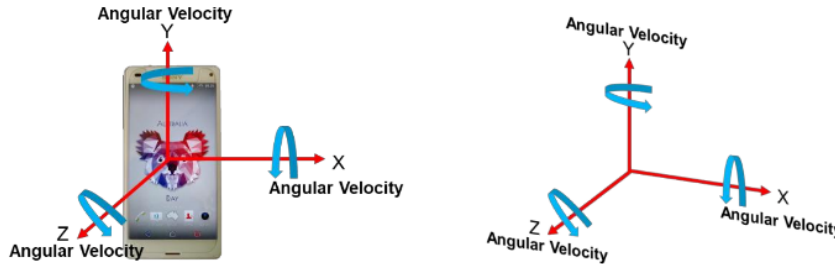


Figure 3. Gyroscope vector and axes

18

2.4. Logistic Regression

Binary logistic regression is one approach used to analyze the relationship between response variables with a set of binary predictor variables [13]. Parameter predictions in the logistic regression model can be done using the Maximum Likelihood Estimation (MLE) method. The general logistic regression model is:

$$\pi(x) = \frac{\exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)}{1 + \exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)} \quad (1)$$

with $0 \leq \pi(x) \leq 1$

The function $\pi(x)$ is a non-linear function so it is necessary to do a logit transformation to obtain a linear function. By performing a transformation from $\logit \pi(x)$, it can be expressed as $g(x)$, namely:

$$g(x) = \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) \quad (2)$$

Where :

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k \quad (3)$$

2.5. Gradient Boosting

Ensemble is the process of combining several sets of models to get a more accurate model by conducting learning and training from the original data set and constructing hypotheses from the data being trained. The training set process is chosen based on the performance of the previous classifiers. Where samples that are not correctly predicted by classifiers will be selected compared to samples that are predicted correctly. Thus, ensembles try to produce new base classifiers that are better for predicting samples on previous base classifiers that have poor performance [14].

Gradient boosting is a machine learning technique for regression and classification problems by matching simple parameter functions that produce prediction models in ensemble form from weak models.

In the case of the estimation function one system consists of a random 'output' or 'response', variable y and a random set of 'inputs' variables $x = \{x_1, \dots, x_n\}$. Given the 'training' sample $\{y_i, \mathbf{X}_i\}_1^N$ which is known the value $(x; y)$, the aim is to find the function $F^*(x)$ which divides x to y , so that through distribution to all values $(y; x)$ [15].

Algorithm Gradient Boosting

$$F_0(\mathbf{x}) = \arg \min_{\gamma} \sum_{i=1}^N \Psi(y_i, \gamma).$$

for $m = 1$ to M do:

$$\tilde{y}_{im} = - \left[\frac{\partial \Psi(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)} \right]_{F(\mathbf{x})=F_{m-1}(\mathbf{x})}, i = 1, N$$

$$\{R_{lm}\}_1^L = L - \text{terminal node } tree(\{\tilde{y}_{im}, \mathbf{x}_i\}_1^N)$$

$$\gamma_{lm} = \arg \min_{\gamma} \sum_{\mathbf{x}_i \in R_{lm}} \Psi(y_i, F_{m-1}(\mathbf{x}_i) + \gamma)$$

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + v \cdot \gamma_{lm} \mathbf{1}(\mathbf{x} \in R_{lm})$$

endfo

17. Performance Evaluation

Performance evaluation of the classification method is calculated based on the value of misclassification using confusion matrix. [16].

Table 1. Confusion Matrix

Prediction	Walking Upstair	Walking Downstair
Walking Upstair	True Positif	False Negatif
Walking Downstair	False Positif	True Negatif

The performance of each classification was evaluated based on Accuracy, Sensitivity, and Specificity. Measurement uses the following equation:

$$accuracy = \frac{TP+TN}{TN+FP+FN+TP} \quad (4)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (6)$$

3. Experimental Results and Discussion

3.1. Ensemble Logistic Regression with GradientBoost Algorithm

- Load dataset
- Identification label attribute and class
- Determine training set and testing set
- Form classification
 - Initialization model
Classification *Logistic Regression* with approach *maximum likelihood*
 - Use *likelihood* functions and logarithms
 - Differentiate the *likelihood* equation
 - Iteration $m = 1$ to M
 - Calculate *pseudo-residuals*
 - Train *pseudo-residuals* using training set
 - Count *multiplier*
 - Update model
- Output model

3.2. Dataset

In this study, data collection uses an accelerometer, gravity sensor, and Gyroscope that has been embedded in a smartphone with an Android-based system that has been designed to read the sensor. Android smartphone is placed on the human right thigh with a duration of ± 12 seconds. Data Accelerometer, Gravity sensor, and Gyroscope will be automatically stored in the smartphone storage space in *.xls format (walking upstairs = 0, walking downstairs = 1). Data in the form of x, y, z coordinates of Accelerometer, Gravity sensor, and Gyroscope with descriptions as in Table 2.

Table 2. Sensor Description

Sensor	Description
Accelerometer	Measure the acceleration force that applied to the device, including force of gravity
Gravity sensor	Measure the force of the gravity that applied to the device, in three axes (x,y,z)
Gyroscope	Measure the device's rotation in three axes (x,y,z)

These three sensors produce thousands of records that are used as datasets for the classification process. Reducing records in the data collection is done for the same class representation between the 'walking upstairs' class and 'walking downstairs'. Total data of 9150 records with representations of each class of 4575 records. Of the total 9150 records partitioned into 70% training data and 30% test data.

Table 3. Dataset

16	Accelerometer			Gravity sensor			Gyroscope			Class
	x	y	z	x	y	z	x	y	z	
1.6829224	9.595764	-0.651001	-0.070755005	-0.2472229	0.034194946	1.6026108	9.672657	-0.20429027	0	
1.6829224	9.595764	-0.651001	0.16217041	-0.03843689	-0.008407593	1.6026108	9.672657	-0.20429027	0	
1.2088623	9.825607	-1.0420532	0.16217041	-0.03843689	-0.008407593	1.6026108	9.672657	-0.20429027	0	
1.2088623	9.825607	-1.0420532	0.33758545	0.00061	-0.0289917	1.6026108	9.672657	-0.20429027	0	
1.2088623	9.825607	-1.0420532	0.33758545	0.00061	-0.0289917	1.4840026	9.683384	-0.44742778	0	
1.0460663	9.471268	-1.7172241	0.33758545	0.00061	-0.0289917	1.4840026	9.683384	-0.44742778	0	
1.0460663	9.471268	-1.7172241	-0.038085938	-0.026016235	-0.073028564	1.4840026	9.683384	-0.44742778	0	
1.290268	9.677155	-1.4698181	-0.038085938	-0.026016235	-0.073028564	1.4840026	9.683384	-0.44742778	0	
0.9311371	9.699509	-0.9271393	-0.038085938	-0.026016235	-0.073028564	1.4840026	9.683384	-0.44742778	0	
0.9311371	9.699509	-0.9271393	-0.038085938	-0.026016235	-0.073028564	1.3473698	9.633099	-1.2483561	0	
1.3285828	9.58139	-1.3181915	0.006637573	0.008071899	0.001907349	1.3304217	9.6267	-1.3141541	1	
1.3285828	9.58139	-1.3181915	0.002029419	-0.007904053	0.001190186	1.3304217	9.6267	-1.3141541	1	
1.3445282	9.586182	-1.3134003	0.002029419	-0.007904053	0.001190186	1.3304217	9.6267	-1.3141541	1	
1.3445282	9.586182	-1.3134003	0.010192871	0.018005371	-0.006988525	1.3304217	9.6267	-1.3141541	1	
1.2854767	9.568619	-1.41716	0.010192871	0.018005371	-0.006988525	1.3304217	9.6267	-1.3141541	1	
1.2854767	9.568619	-1.41716	0.010192871	0.018005371	-0.006988525	1.3343246	9.62437	-1.3272034	1	
1.2854767	9.568619	-1.41716	0.010192871	0.018005371	-0.006988525	1.3322059	9.62332	-1.3369117	1	
1.2854767	9.568619	-1.41716	0.005218506	0.002380371	-0.003082275	1.3322059	9.62332	-1.3369117	1	
1.3190002	9.590973	-1.2942505	0.005218506	0.002380371	-0.003082275	1.3322059	9.62332	-1.3369117	1	
1.3190002	9.590973	-1.2942505	0.005218506	0.002380371	-0.003082275	1.315864	9.622038	-1.3620875	1	

3.3. Performance Evaluation

Python 2.7 programming language is used to obtain the overall classification results from the logistic regression method and gradientboost ensemble.

Table 4. Confusion Matrix of Logistic Regression

	0	1
0	1016	360
1	493	876

Table 5. Confusion Matrix of ensemble Logistic Regression with GradientBoost

	0	1
0	1185	191
1	308	1061

Table 6. Performance method

Method	Accuracy	Sensitivity	Specificity
Logistic Regression	68.92	73.83	63.98
Ensemble Logistic Regression with GradientBoost	81.82	86.11	77.50

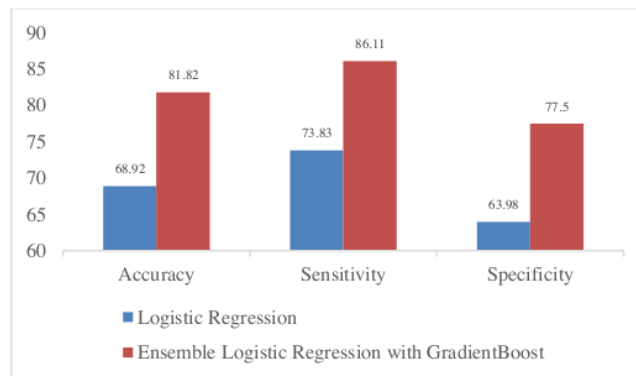


Figure 4. Improved method performance

The success of the proposed method is based on improving performance in terms of accuracy and low sensitivity and specificity. Fig. 4. Shows the performance improvement of logistic regression using ensemble techniques with the gradientboost algorithm. Logistic regression gives accuracy 68.92%, sensitivity 73.83%, and specificity 63.98% while the ensemble logistic regression with gradientboost gives accuracy 81.82%, sensitivity 86.11, and specificity 77.5. the overall performance of the method is shown in table 6. Ensemble logistic regression with gradientboost managed to increase the accuracy of about 12.90% and sensitivity around 12.28% from logistic regression but the performance of the ensemble logistic regression with gradientboost was very bad in terms of specificity because it gave a high value even though it should give lower values of logistic regression. The success of the ensemble logistic regression with gradientboost in increasing accuracy and sensitivity is based on improving classification errors from logistic regression. As shown in tables 4 and 5. the value of Negative False and Positive False from logistic regression was successfully reduced and succeeded in increasing the value of Positive True and Negative True.

4. Conclusion

In this study, we propose an ensemble technique for the classification of human physical activity using an innate sensor from a smartphone, logistic regression is used as an initial classification and to improve the performance of the logistic regression classification the gradientboost method is used. Data retrieval uses Accelerometer, Gravity sensor and Gyroscope readings from a smartphone. The results show that the proposed method succeeded in improving performance by correcting misclassification of previous calculations and providing significant improvements. The results of the experimental evaluation of the ensemble logistic regression with gradientboost showed that the achievement of performance was very good by increasing accuracy of 12.90% and sensitivity 12.28% in the dataset of walking upstairs and walking downstairs, although it was found that specificity had poor performance. The future will be investigated to improve the performance of the ensemble gradientboost in terms of specificity and testing in various data partitions to determine the performance of the ensemble gradientboost.

References

- [1] Chen, M.S., Han, J. and Yu, P.S., 1996. Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and data Engineering*, 8(6), pp.866-883.
- [2] Shoaib, M., Scholten, H. and Havinga, P.J., 2013, December. Towards physical activity recognition using smartphone sensors. In *2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing*, pp. 80-87.
- [3] Wang, A., Chen, G., Yang, J., Zhao, S. and Chang, C.Y., 2016. A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sensors Journal*, 16(11), pp.4566-4578.
- [4] D'souza, W.T. and Kavitha, R., 2017. Human Activity Recognition Using Accelerometer and Gyroscope Sensors. In *International Journal of Engineering and Technology (IJET)*, Vol 9 No 2 Apr-May.
- [5] Jain, A. and Kanhangad, V., 2018. Human activity classification in smartphones using accelerometer and gyroscope sensors. *IEEE Sensors Journal*, 18(3), pp.1169-1177.
- [6] Wannenburg, J. and Malekian, R., 2017. Physical activity recognition from smartphone accelerometer data for user context awareness sensing. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(12), pp.3142-3149.
- [7] Zhang, C. and Ma, Y. eds., 2012. *Ensemble machine learning: methods and applications*. Springer Science & Business Media.
- [8] Lawi, A., Aziz, F. and Syarif, S., 2017, August. Ensemble GradientBoost for increasing classification accuracy of credit scoring. In *2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT)* (pp. 1-4).
- [9] Rindler, W., 2012. *Essential relativity: special, general, and cosmological*. Springer Science & Business Media.
- [10] Tinker, R.F., 2006. Relativistic flight mechanics and space travel. *Synthesis lectures on engineering*, 1(1), pp.1-140.
- [11] Millette, G. and Stroud, A., 2012. *Professional Android sensor programming*. John Wiley & Sons.
- [12] Kabai, S., 2008. Oldham coupling, The Wolfram Demonstrations Project.
- [13] Hosmer Jr, D.W., Lemeshow, S. and Sturdivant, R.X., 2013. *Applied logistic regression* (Vol. 398). John Wiley & Sons.
- [14] Zhang, C. and Ma, Y. eds., 2012. *Ensemble machine learning: methods and applications*. Springer Science & Business Media.
- [15] Friedman, J.H., 2002. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4), pp.367-378.
- [16] Provost, F. and Kohavi, R., 1998. Guest editors' introduction: On applied research in machine learning. *Machine learning*, 30(2), pp.127-132.

Increasing accuracy of classification physical activity based on smartphone using ensemble logistic regression with boosting method

ORIGINALITY REPORT

16%

SIMILARITY INDEX

7%

INTERNET SOURCES

9%

PUBLICATIONS

13%

STUDENT PAPERS

PRIMARY SOURCES

1	Armin Lawi, Firman Aziz. "Classification of Credit Card Default Clients Using LS-SVM Ensemble", 2018 Third International Conference on Informatics and Computing (ICIC), 2018 Publication	2%
2	earchive.tpu.ru Internet Source	2%
3	Submitted to Northern Consortium UK Student Paper	1%
4	Submitted to ABV-Indian Institute of Information Technology and Management Gwalior Student Paper	1%
5	www.mdpi.com Internet Source	1%
6	teacher.tku.edu.tw Internet Source	1%
7	Submitted to Universitas Pendidikan Indonesia Student Paper	

1%

8

Ankita Jain, Vivek Kanhangad. "Human Activity Classification in Smartphones Using Accelerometer and Gyroscope Sensors", IEEE Sensors Journal, 2018

Publication

1%

9

Submitted to City University

Student Paper

1%

10

Journal of Financial Management of Property and Construction, Volume 19, Issue 2 (2014-09-16)

Publication

1%

11

Submitted to Eastern Mediterranean University

Student Paper

<1%

12

Armin Lawi, Firman Aziz. "Comparison of Classification Algorithms of the Autism Spectrum Disorder Diagnosis", 2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT), 2018

Publication

<1%

13

jcheminf.biomedcentral.com

Internet Source

<1%

14

Submitted to Coventry University

Student Paper

<1%

15	Submitted to STKIP Sumatera Barat Student Paper	<1%
16	repositorio.unesp.br Internet Source	<1%
17	"Computational Science – ICCS 2019", Springer Science and Business Media LLC, 2019 Publication	<1%
18	Submitted to University of Southampton Student Paper	<1%
19	Submitted to University of Malaya Student Paper	<1%
20	Submitted to Indian Institute of Technology- Bhubaneswar Student Paper	<1%
21	Yan Wang, Shuang Cang, Hongnian Yu. "A survey on wearable sensor modality centred human activity recognition in health care", Expert Systems with Applications, 2019 Publication	<1%
22	bdataanalytics.biomedcentral.com Internet Source	<1%
23	Submitted to University of Sussex Student Paper	<1%
24	Submitted to Universiti Teknologi MARA Student Paper	<1%

25

Dong F. Wang, Huan Liu, Xiaodong Li, Yang Li, Weikang Xian, Takeshi Kobayashi, Toshihiro Itoh, Ryutaro Maeda. "Passive MEMS DC Electric Current Sensor: Part II–Experimental Verifications", IEEE Sensors Journal, 2017

Publication

<1%

26

Muhammad Ehatisham-ul-Haq, Muhammad Awais Azam, Usman Naeem, Yasar Amin, Jonathan Loo. "Continuous authentication of smartphone users based on activity pattern recognition using passive mobile sensing", Journal of Network and Computer Applications, 2018

Publication

<1%

Exclude quotes On

Exclude matches < 5 words

Exclude bibliography On