Human Activity Recognition Based on Convolutional Neural Network via Smart-phone Sensors

Zesheng Chen
Information and Technology Center, Guangzhou Academy of Fine Arts, Guangzhou, 51006, China
czs917@gzarts.edu.cn

Min Zhou*
Information and Technology Center, Guangzhou Academy of Fine Arts, Guangzhou, 51006, China
zhoumin@gzarts.edu.cn

Lichun Feng
Information and Technology Center, Guangzhou Academy of Fine Arts, Guangzhou, 51006, China
fenglichun@gzarts.edu.cn

Bingnan Li
Information and Technology Center, Guangzhou Academy of Fine Arts, Guangzhou, 51006, China
beyond@gzarts.edu.cn

ABSTRACT

To resolve the problem of insufficient accuracy in human activity recognition based on a single accelerometer sensor and traditional machine learning, this paper collects data on human activities using a smartphone embedded with multisensors, and then develop a framework based on a convolution neural network to classify human activities. When building a four- layer neural network, maximum pooling is utilized for every two layers, and the dropout technique is employed in case of overfitting. A full connection layer is then built using average pooling, and the Softmax approach is used for multi- classification. Experiments reveal that the suggested framework of convolution neural network enhances the accuracy of human activity recognition significantly.

CCS CONCEPTS

• **Computing methodologies**; • **Machine learning**; • **Machine learning approaches**;

KEYWORDS

Human activity recognition, Convolutional neural network, Smartphone sensors, Signal processing

ACM Reference Format:
Zesheng Chen, Min Zhou*, Lichun Feng, and Bingnan Li. 2022. Human Activity Recognition Based on Convolutional Neural Network via Smartphone Sensors. In *2022 5th International Conference on Signal Processing and Machine Learning (SPML 2022), August 04–06, 2022, Dalian, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3556384.3556406

1 INTRODUCTION

Human activity recognition (HAR) has received significant study interest due to its numerous application contexts such as human-computer interface, medical, and sports in recent years [1-3]. According to various activity data, HAR can be basically split into image-based HAR and sensor-based HAR; similarly, HAR technologies can be roughly divided into classical machine learning and neural network in artificial intelligence. The majority of HAR research is divided into three categories: HAR based on machine learning and sensors, HAR based on neural network and sensors, and HAR based on neural network and image.

Early HAR research is based on video image processing technology [4], however this method is impacted by light intensity. With advancements in artificial intelligence, HAR based on deep learning effectively avoids these negative consequences to some extent [5-7]. Deep learning on images, on the other hand, has a high computational complexity, necessitating not just massive computers in general but also posing privacy issues for real-time surveillance. These drawbacks show that, while HAR based on video pictures can be useful in some domains, it cannot cover all of the situations required by society.

Human activity data may be collected using accelerometer sensors as Micro-Electro-Mechanical-System (MEMS) manufacturing technology advances. Sensor-based HAR is getting popular, and the three-axis accelerometer and gravity sensor are generally applied in HAR data collecting [8-10]. Traditional machine learning classification approaches have demonstrated that accuracy is insufficient for HAR. However, deep learning outperforms superior machine learning in classification, and the pace with which deep learning processes acceleration data is significantly faster than that of processing image data. As a result, utilizing deep learning to analyze sensor data not only improves recognition accuracy but also does not affect computational performance.

However, the data collection by a single sensor may contain lacking information to express fully a certain activity class. Therefore, many researchers use multisensors to collect human activities data. But the multisensors method is difficult to be applied in real life. Users are not willing to accept carrying sensors on multiple parts of their body which brings serious inconvenience to life. Researchers and engineers collect data through smartphones because of the universality of smartphones which are embedded with multiple sensors. This way will be more in line with reality because no extra sensing equipment is needed [11]. To sum up, combining neural network and multisensors set in smartphones have become a new research direction. The mainstream deep learning technologies for HAR mainly include Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Restricted Boltzmann Machine

SPML 2022, August 04–06, 2022, Dalian, China

ACM ISBN 978-1-4503-9691-2/22/08…$15.00
https://doi.org/10.1145/3556384.3556406

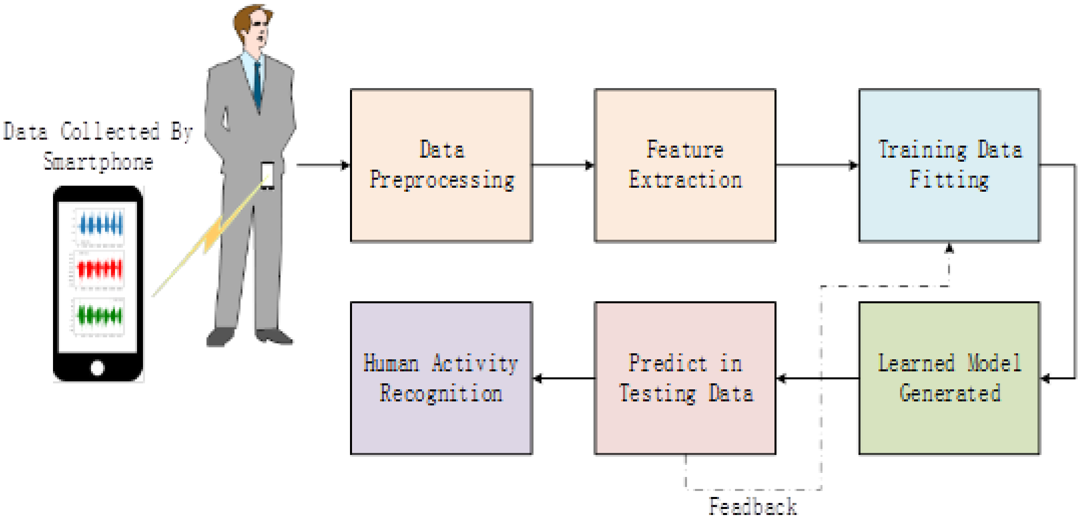

Figure 1: The process of HAR

(RBM). Deep learning methods cannot only effectively improve the recognition accuracy, but also do not need manual feature extraction, so the limits of manual feature extraction can be avoided in experiments, and classification accuracy can be improved.

Ronao et al conducted experiments on the neural network layers, and the number of neurons and learning rate by using CNN method, and gained 95.75% accuracy [12]. Bashar use 5 layers of neural network, the number of neurons in the first layer is 256, and won 95.79% accuracy [13]. However, compared with the deep learning framework proposed in this paper, the complexity of the framework is relatively simple, and the number of neurons is also small. The recognition accuracy achieved 98.5%, which reflects the effectiveness of the CNN framework proposed in this paper.

2 RELATED WORK

The classification algorithm is the focus of academic research in HAR. [14] divides the classification algorithm in HAR into shallow method and depth method. [15] showed that the support vector machine has the highest accuracy of human behavior recognition among the seven classification methods used for 19 different activities. With the improvement of computer performance, HAR based on deep learning has become a research hotspot. The mainstream deep learning technologies for HAR mainly include convolutional neural network (CNN), recurrent neural network (RNN), and restricted Boltzmann machine (RBM). [16] pointed out that higher recognition accuracy can be obtained by using the deep learning method to solve sensor-based HAR compared with the machine learning method. [17] used RBM to integrate multisensors data streams, and experiments verified that this method has better performance than support vector machine (SVM) and ensemble learning. [12] use the data produced by the built-in acceleration sensor of the smartphone to identify six activities. The recognition accuracy of converting triaxle acceleration data into image data is 92.71%. Inadequate recognition accuracy and conversion into image data will inevitably increase the training time of the model. Therefore, a new convolutional neural network model is proposed in this paper. By building a four-layer neural network, maximum pooling is adopted for every two layers, and the dropout method is 2 used in case of overfitting. Construct a full connection layer through average pooling, and finally, the Softmax method is used for multi-classification. The experiments show the proposed convolution neural network effectively improves the accuracy of HAR. Experiments show the effectiveness of the proposed model.

3 FRAMEWORK OF HAR

Sensor-based HAR is studied in terms of the amount, type, wearing position, and other aspects of sensors [17–19], and its overall solution has developed into a mature solution architecture, as shown in Figure 1. The basic processes are data collection, data prepossessing, feature extraction, dimension reduction, and classification. Each process has the potential to affect the final accuracy. As a result, many scholars have conducted research in various processes, yielding rich theoretical achievements in the study of HAR.

First, the data collection of HAR is mainly realized through the three-axis accelerometer, magnetometer, gyroscope, and gravity acceleration sensor. Sampling frequency is important for data collection. There will not be enough information if the sampling frequency is too low, and data will be redundant if the sampling frequency is too high. According to studies, a sampling frequency of 20 HZ contains enough information about human activities to form a sampling frame. We find that the accelerometer and gyroscope are the two most frequently used categories in HAR.

There are noise and other problems in the raw data after data collection, so it is necessary to preprocess the raw data. The commonly used prepossessing methods include sliding window, filtering, and normalization. The sliding window method can fix the data blocks which is convenient for feature extraction. Filtering can prevent and suppress signal noise, and normalization can compress the data that changes in a larger value range into a smaller value range.

Table 1: The value of feature name and description

The value of feature	Description	The value of feature	Description
mean	Mean value	arCoeff	Autorregresion coefficients with Burg order equal to 4
std	Standard deviation	correlation	Correlation coefficient between two signals
mad	Median absolute deviation	maxInds	Index of the frequency component with largest magnitude
max	Largest value in array	meanFreq	Weighted average of the frequency components to obtain a mean frequency
min	Smallest value in array	skewness	Skewness of the frequency domain signal
sma	Signal magnitude area	kurtosis	Kurtosis of the frequency domain signal
energy	Sum of the squares divided by the number of values	bandsEnergy	Energy of a frequency interval within the 64 bins of the FFT of each window.
iqr	Interquartile range	angle	Angle between to vectors.
entropy	Signal entropy		

In traditional machine learning, dealing directly with the raw data gained by the general effect is not ideal, it always needs to manually conduct feature extraction and dimension reduction in raw data. To better characterize data to identify specific actions, HAR has usually extracted features in the time-domain and frequency-domain. Time-domain characterization is a function of signal change with time, for example, mean, maximum, minimum, and variance. In the frequency-domain according to the frequency and amplitude spectrum. Common features in the frequency-domain include the Fast Fourier Transform (FFT) coefficient, power spectrum, and frequency-domain entropy. However, too many features will lead to dimension disaster. Therefore, after feature extraction, dimension reduction is usually performed, Principal Component Analysis (PCA) and linear discriminant analysis (LDA) are commonly used for dimension reduction.

The data can be used in classification experiments after feature processing. The mainstream human activity recognition and classification methods in academic research include the fixed threshold method, traditional machine learning, and deep learning at present. The algorithm used in classification usually has a great influence on the accuracy of classification. Therefore, many researchers are committed to the study of algorithms. SVM, KNN, DT, integrated learning, and deep learning are commonly used algorithms in human activity recognition. Figure 1 simply describes the framework of HAR.

4 EXPERIMENT

4.1 Dataset Description

We use a dataset called “Smartphone Dataset for Human Activity Recognition (HAR) in Ambient Assisted Living” which was collected by built-in sensors in smartphones in the UCI database, provided by SmartLab Nonlinear Complex Systems in Italy. This dataset was collected from 30 volunteers aged 19-48, who wore a SAMSUNG Galaxy SII on the waist and performed walking, upstairs, downstairs, lying, sitting, and standing. The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap. From each window, a vector of features was obtained by calculating variables from the time and frequency domain. A total of 10411 samples, 561-dimensional features. Due to the excessive number of features, Table 1 only lists the measurement methods of features.

4.2 Experiment Environment

We conducted experiments on Anaconda integrated development environment (with Keras 2.4.3) running on the executing host: 64-bit Windows 10 OS, Intel (R) Core (TM) i7-7700 CPU @ 3.60 GHz, 8GB of RAM.

4.3 Algorithm Design

The sensor collected data is one-dimensional, so we use the Keras library builds a one-dimensional convolutional neural network model, the basic structure of the model is shown in figure 3. The specific model structure is shown in Table 2. We have designed four one-dimensional convolution layers, the meaning of the convolution is a special kind of linear operation, it is shown as formula 1, where $x_{m+i,n+j}$ represents the pixel value of 2D input data at point $(m+i, n+j)$, w_{ij} represents the value of the convolution kernel with size at point (i, j), b represents bias, M and N represent the size of the input image in two dimension respectively, and f represents activation function, and y_{mn} represents the output after convolution at point (m, n).

$$y_{mn} = f\left(\sum_{j=0}^{Q-1}\sum_{i=0}^{p-1} x_{m+i,n+j} w_{ij} + b\right) \quad 0 \le m \le M, \quad 0 \le n \le N \quad (1)$$

We add a maximum pooling layer and a dropout layer between each of the two convolutional layers. The purpose of the maximum pooling layer is to gradually reduce the space size and reduce the amount of computation to prevent over-fitting. Figure 2 shows the maximum pooling principle. Take the maximum value of the four pixels in each region2×2, and set the step size of the region2×2 to 2, and finally obtain the feature graph of $1/4$ of the original data size.

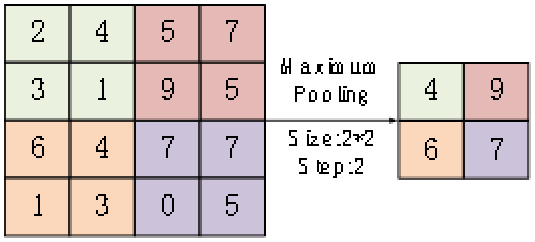

Figure 2: Maximum Pooling

The Dropout layer inhibits overfitting by randomly discarding some neural units. At the end of the model, we added an average pooling layer and a full connection layer, and finally implemented multi-classification through Softmax function. Softmax is used to calculate the probability distribution, and its basic formula is as formula 2, where N is the total number of elements in array X.

$$s_i = \frac{e^{X^i}}{\sum_{n=1}^{N} e^{X^n}} \tag{2}$$

4.4 Experimental Analysis

Table 3: Classification result

Num	Precision	Recall	F1-score
0	0.97	0.97	0.97
1	0.97	0.96	0.96
2	1.00	1.00	1.00
3	1.00	0.99	0.99
4	0.99	0.98	0.98
5	0.98	1.00	0.99

After constructing the CNN model, we started the experiment for analysis, and the experimental data adopted the dataset introduced in section 3.1. The details of classification are represented by a confusion matrix, as shown in Figure 4 In the training process. The training accuracy and loss are shown in Figure 5. The training accuracy and test accuracy are shown in Figure 6. The classification result is shown in Table 3, including Precision, Recall, and F1 score. Precision means the proportion of samples that are predicted to be positive that is positive. Recall represents the proportion of all positive samples that are correctly predicted. F1-score represents the mean of recall and precision.

Experimental results show that the overall recognition accuracy of the CNN model proposed in this paper reaches 98.5%. For dynamic activities such as walking, upstairs, and downstairs, some sample recognition errors occur, but the accuracy rate is above 97%. For static activities (lying, sitting and standing), the recognition accuracy is close to 100%. Experimental results show the effectiveness of the proposed CNN model.

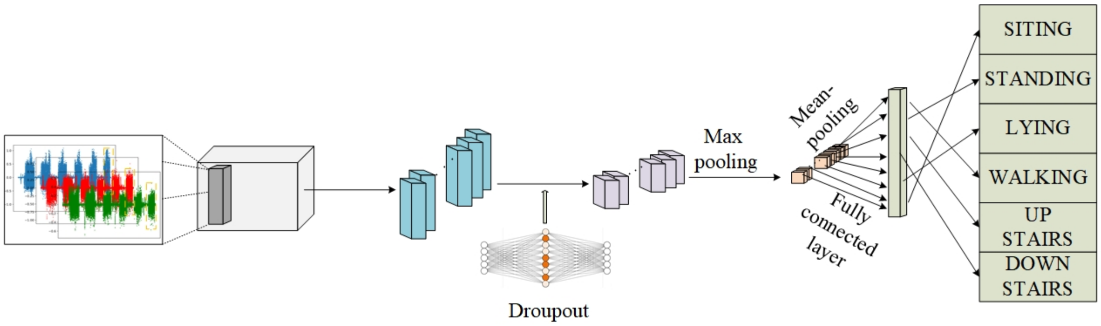

Figure 3: Framework of CNN proposed

Table 2: The specific model structure proposed

Layer(type)	Output Shape	Parameter Number
Convld_1	(None,561,16)	128
Convld_2	(None,561,32)	3616
Max_poolingld_1	(None,187,32)	0
Dropout_1	(None,187,32)	0
Convld_3	(None,187,64)	14400
Convld_4	(None,187,128)	57472
Max_poolingld_2	(None,62,128)	0
Dropout_2	(None,62,128)	0
Global_average_poolingld_1	(None,128)	0
Dense_1	(None,6)	774

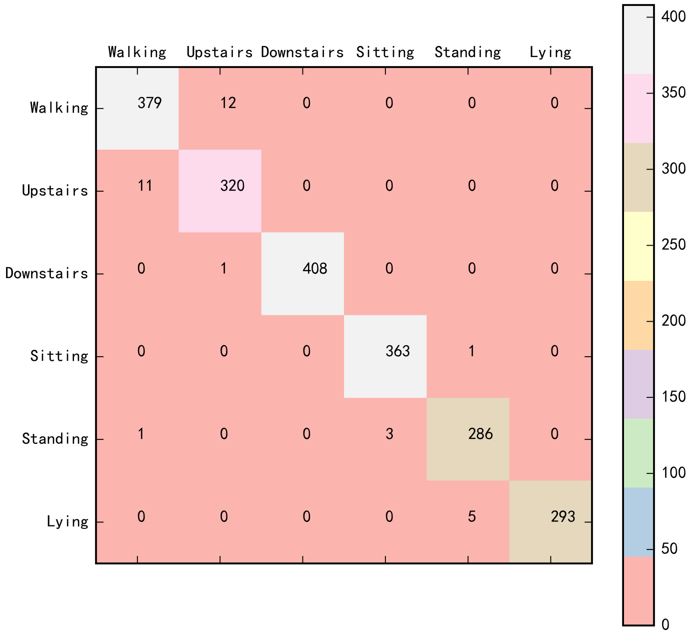

Figure 4: Confusion matrix

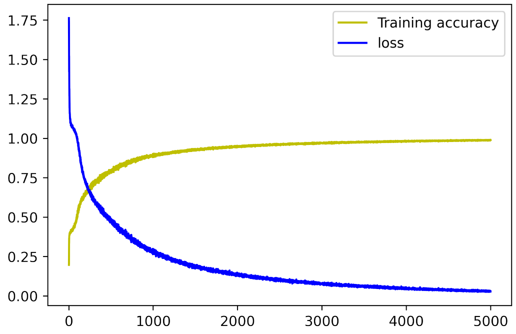

Figure 5: Training accuracy and loss

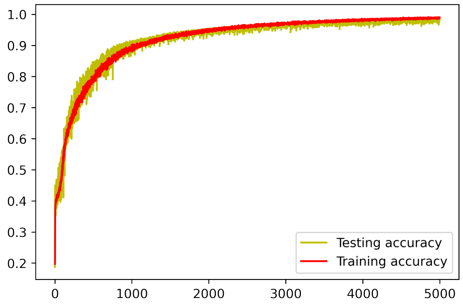

Figure 6: Training and Testing accuracy

5 CONCLUSIONS

The CNN model proposed in this paper's recognition accuracy reaches 98.36%, which fully confirms the effectiveness of the model in HAR. In addition, the six types of activities in the paper are divided into two parts, which are static activities and dynamic activities. Static activities include lying, sitting, and standing, and dynamic activities include walking, upstairs, and downstairs. According to the recognition results of the confusion matrix, we can find that 12 walking samples are recognized as upstairs, 11 upstairs samples are recognized as walking, One downstairs sample was identified as upstairs, One sitting sample was identified as standing, one standing sample was identified as walking, three standing samples were identified as sitting, and five lying samples were identified as standing.

The specific activity in the dynamic activities is incorrectly recognized, but it is still identified as a dynamic activity. The specific activity in the static activities is incorrectly recognized, but it is still identified as a static activity. This phenomenon shows that on the one hand, it can indicate that there is a large difference between dynamic and static features of activities, and on the other hand, it can also indicate the effectiveness of the model proposed in this paper

ACKNOWLEDGMENTS

This work has been supported by the Research on Art Communication Mode under the New Media Environment of Scientific Research Project of Guangzhou Academy of Fine Arts (Grant No.15XJ038), the research team project of Guangzhou Academy of Fine Arts (Grant No. 6043170950).

REFERENCES

[1] Tong Z, Wang J, Liang X, *et al.* Fall Detection by Wearable Sensor and One-Class SVM Algorithm. J. lecture notes in control & information sciences, (2006), 858–863. https://doi.org/10.1007/978-3-540-37258-5_104

[2] Al-Janabi and A. H. Salman. Sensitive integration of multilevel optimization model in human activity recognition for smartphone and smartwatch applications. Big Data Mining and Analytics. 4, 2 (June 2021), 124-138. https://doi.org/10.26599/BDMA.2020.9020022

[3] San Buenaventura, C. V, Tiglao, N. M. C, & Atienza, R. O. 2019. Deep Learning for Smartphone-Based Human Activity Recognition Using Multi-sensor Fusion. Wireless Internet, 65–75. https://doi.org/10.1007/978-3-030-06158-6_7

[4] Kushwaha A, Khare A, Khare M. Human Activity Recognition Algorithm in Video Sequences Based on Integration of Magnitude and Orientation Information of Optical Flow. J. International Journal of Image and Graphics. 22, 01 (April 2021), 2250009. https://doi.org/10.1142/S0219467822500097

[5] Singh R, Sonawane A, Srivastava R. Recent evolution of modern datasets for human activity recognition: a deep survey. J. Multimedia Systems, 26, 2 (2020), 83-106. https://doi.org/10.1007/s00530-019-00635-7

[6] Plotz T, Guan Y. Deep Learning for Human Activity Recognition in Mobile Computing. J. Computer, 51, 5 (May 2018), 50-59. https://doi.org/10.1109/MC.2018.2381112.

[7] Ding W, Guo X, and Wang G. Radar-based Human Activity Recognition Using Hybrid Neural Network Model with Multi-domain Fusion. J. IEEE Transactions on Aerospace and Electronic Systems. 57, 5 (October 2021), 2889-2898. https://doi.org/10.1109/TAES.2021.3068436

[8] Chen Z, Cai C, Zheng T, *et al.* RF-Based Human Activity Recognition Using Signal Adapted Convolutional Neural Network. J. IEEE Transactions on Mobile Computing. (2021). https://doi.org/10.1109/TMC.2021.3073969

[9] Capela NA, Lemaire ED, Baddour N. Feature Selection for Wearable Smartphone-Based Human Activity Recognition with Able bodied, Elderly, and Stroke Patients. J. PLoS ONE. 10(4): (2015), e0124414. https://doi.org/10.1371/journal.pone.0124414

[10] Khan AM, Siddiqi MH, Lee S-W. Exploratory Data Analysis of Acceleration Signals to Select Light-Weight and Accurate Features for Real-Time Activity Recognition on Smartphones. J. Sensors. 13 10 (2013) 13099-13122. https://doi.org/10.3390/s131013099

[11] 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. J. IEEE Sensors Journal. 16 11 (June 2016), 4566-4578. https://doi.org/10.1109/JSEN.2016.2545708

[12] Bayat A, Pomplun M, Tran D A. A Study on Human Activity Recognition Using Accelerometer Data from Smartphones. J. Procedia Computer Science, 34 (2014), 450-457. https://doi.org/10.1016/j.procs.2014.07.009

[13] Bashar S K, Fahim A A, Chon K H. Smartphone Based Human Activity Recognition with Feature Selection and Dense Neural Network. 2020. IEEE 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in conjunction with the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society. CA, 5888-5891. https://doi:10.1109/EMBC44109.2020.9176239

[14] Lima W S, Souto E, El-Khatib K, *et al.* Human Activity Recognition Using Inertial Sensors in a Smartphone: An Overview. J. Sensors. 19 14 (2019), 3213. https://doi.org/ 10.3390/s19143213

[15] Altun K, Barshan B. Human activity recognition using inertial/magnetic sensor units. 2010. Proceedings of the 1st International Conference on Human Behavior Understanding, CA, 38-51. https://doi.org/10.1007/978-3-642-14715-9_5

[16] Zhang L, Wu X, Luo D. Recognizing Human Activities from Raw Accelerometer Data Using Deep Neural Networks. 2015. IEEE 14th International Conference on Machine Learning and Applications (ICMLA). CA, 865-870. https://doi.org/10.1109/icmla.2015.48

[17] Radu V, Lane N D, Bhattacharya S, *et al.* Towards multimodal deep learning for activity recognition on mobile devices. 2016. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. CA, 185-188. https://doi.org/10.1145/2968219.2971461