Pattern Recognition Letters
journal homepage: www.elsevier.com

# Deep Learning for Sensor-based Activity Recognition: A Survey

Jindong Wang[a,b], Yiqiang Chen[a,b,**], Shuji Hao[c], Xiaohui Peng[a,b], Lisha Hu[a,b]

[a]*Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China*
[b]*University of Chinese Academy of Sciences, Beijing, China*
[c]*Institute of High Performance Computing, A*STAR, Singapore*

ABSTRACT

Sensor-based activity recognition seeks the profound high-level knowledge about human activities from multitudes of low-level sensor readings. Conventional pattern recognition approaches have made tremendous progress in the past years. However, those methods often heavily rely on heuristic hand-crafted feature extraction, which could hinder their generalization performance. Additionally, existing methods are undermined for unsupervised and incremental learning tasks. Recently, the recent advancement of deep learning makes it possible to perform automatic high-level feature extraction thus achieves promising performance in many areas. Since then, deep learning based methods have been widely adopted for the sensor-based activity recognition tasks. This paper surveys the recent advance of deep learning based sensor-based activity recognition. We summarize existing literature from three aspects: sensor modality, deep model, and application. We also present detailed insights on existing work and propose grand challenges for future research.

*Keywords:* Deep learning; activity recognition; pattern recognition; pervasive computing



## 1. Introduction

Human activity recognition (HAR) plays an important role in people's daily life for its competence in learning profound high-level knowledge about human activity from raw sensor inputs. Successful HAR applications include home behavior analysis (Vepakomma et al., 2015), video surveillance (Qin et al., 2016), gait analysis (Hammerla et al., 2016), and gesture recognition (Kim and Toomajian, 2016). There are mainly two types of HAR: *video-based* HAR and *sensor-based* HAR (Cook et al., 2013). Video-based HAR analyzes videos or images containing human motions from the camera, while sensor-based HAR focuses on the motion data from smart sensors such as an accelerometer, gyroscope, Bluetooth, sound sensors and so on. Due to the thriving development of sensor technology and pervasive computing, sensor-based HAR is becoming more popular and widely used with privacy well protected. Therefore, in this paper, our main focus is on sensor-based HAR.

HAR can be treated as a typical pattern recognition (PR) problem. Conventional PR approaches have made tremendous progress on HAR by adopting machine learning algorithms such as decision tree, support vector machine, naive Bayes, and hidden Markov models (Lara and Labrador, 2013). It is no wonder that in some controlled environments where there are only a few labeled data or certain domain knowledge is required (e.g. some disease issues), conventional PR methods are fully capable of achieving satisfying results. However, in most daily HAR tasks, those methods may heavily rely on heuristic hand-crafted feature extraction, which is usually limited by human domain knowledge (Bengio, 2013). Furthermore, only shallow features can be learned by those approaches (Yang et al., 2015), leading to undermined performance for unsupervised and incremental tasks. Due to those limitations, the performances of conventional PR methods are restricted regarding classification accuracy and model generalization.

Recent years have witnessed the fast development and advancement of deep learning, which achieves unparalleled performance in many areas such as visual object recognition, natural language processing, and logic reasoning (LeCun et al., 2015). Different from traditional PR methods, deep learning can largely relieve the effort on designing features and can learn much more high-level and meaningful features by training an end-to-end neural network. In addition, the deep network

**Corresponding author
*e-mail:* yqchen@ict.ac.cn (Yiqiang Chen)

**Table 2. Deep learning models for HAR tasks**

| Model | Description |
|---|---|
| **DNN** | Deep fully-connected network, artificial neural network with deep layers |
| **CNN** | Convolutional neural network, multiple convolution operations for feature extraction |
| **RNN** | Recurrent neural network, network with time correlations and LSTM |
| **DBN / RBM** | Deep belief network and restricted Boltzmann machine |
| **SAE** | Stacked autoencoder, feature learning by decoding-encoding autoencoder |
| **Hybrid** | combination of some deep models |

### 4.4. *Restricted Boltzmann Machine*

Restricted Boltzmann machine (RBM) is a bipartite, fully-connected, undirected graph consisting of a visible layer and a hidden layer (Hinton et al., 2006). The stacked RBM is called deep belief network (DBN) by treating every two consecutive layers as an RBM. DBN/RBM is often followed by fully-connected layers.

In pre-training, most work applied Gaussian RBM in the first layer while binary RBM for the rest layers (Plötz et al., 2011; Hammerla et al., 2015; Lane et al., 2015). For multi-modal sensors, (Radu et al., 2016) designed a multi-modal RBM where an RBM is constructed for each sensor modality, then the output of all the modalities are unified. (Li et al., 2016a) added pooling after the fully-connected layers to extract the important features. (Fang and Hu, 2014) used a contrastive gradient (CG) method to update the weight in fine-tuning, which helps the network to search and convergence quickly in all directions. (Zhang et al., 2015b) further implemented RBM on a mobile phone for offline training, indicating RBM can be very light-weight. Similar to autoencoder, RBM/DBN can also perform unsupervised feature learning for HAR.

### 4.5. *Recurrent Neural Network*

Recurrent neural network (RNN) is widely used in speech recognition and natural language processing by utilizing the temporal correlations between neurons. LSTM (long-short term memory) cells are often combined with RNN where LSTM is serving as the *memory* units through gradient descent.

Few work used RNN for the HAR tasks (Hammerla et al., 2016; Inoue et al., 2016; Edel and Köppe, 2016; Guan and Ploetz, 2017), where the learning speed and resource consumption are the main concerns for HAR. (Inoue et al., 2016) investigated several model parameters first and then proposed a relatively *good* model which can perform HAR with high throughput. (Edel and Köppe, 2016) proposed a binarized-BLSTM-RNN model, in which the weight parameters, input, and output of all hidden layers are all binary values. The main line of RNN based HAR models is dealing with resource-constrained environments while still achieve good performance.

### 4.6. *Hybrid Model*

Hybrid model is the combination of some deep models.

One emerging hybrid model is the combination of CNN and RNN. (Ordóñez and Roggen, 2016; Yao et al., 2017) provided good examples for how to combine CNN and RNN. It is shown in (Ordóñez and Roggen, 2016) that the performance of 'CNN + recurrent dense layers' is better than 'CNN + dense layers'. Similar results are also shown in (Singh et al., 2017). The reason is that CNN is able to capture the spatial relationship, while RNN can make use of the temporal relationship. Combining CNN and RNN could enhance the ability to recognize different activities that have varied time span and signal distributions. Other work combined CNN with models such as SAE (Zheng et al., 2016) and RBM (Liu et al., 2016). In those work, CNN performs feature extraction, and the generative models can help in speeding up the training process. In the future, we expect there will be more research in this area.

## 5. Applications

HAR is always not the final goal of an application, but it serves as an important step in many applications such as skill assessment and smart home assistant. In this section, we survey deep learning based HAR from the application perspective.

### 5.1. *Featured Applications*

Most of the surveyed work focused on recognizing *activities of daily living* (ADL) and *sports* (Zeng et al., 2014; Chen and Xue, 2015; Ronao and Cho, 2016; Ravì et al., 2017). Those activities of simple movements are easily captured by body-worn sensors. Some research studied people's *lifestyle* such as sleep (Sathyanarayana et al., 2016) and respiration (Khan et al., 2017; Hannink et al., 2017). The detection of such activities often requires some object and ambient sensors such as WiFi and sound, which are rather different from ADL.

It is a developing trend to apply HAR to *health and disease* issues. Some pioneering work has been done with regard to Parkinson's disease (Hammerla et al., 2015), trauma resuscitation (Li et al., 2016a,b) and paroxysmal atrial fibrillation (PAF) (Pourbabaee et al., 2017). Disease issues are always related to the change of certain body movements or functions, so they can be detected using corresponding sensors.

Under those circumstances, the association between disease and activity should be given more consideration. It is important to use the appropriate sensors. For instance, Parkinson's disease is often related to the frozen of gait, which can be reflected by some inertial sensors attached to shoes (Hammerla et al., 2015).

Other than health and disease, the recognition of *high-level* activities is helpful to learn more resourceful information for HAR. The movement, behavior, environment, emotion, and thought are critical parts in recognizing high-level activities. However, most work only focused on body movements in smart homes (Vepakomma et al., 2015; Fang and Hu, 2014), which is not enough to recognize high-level activities. For instance,

Table 3. Public HAR datasets (A=accelerometer, G=gyroscope, M=magnetometer, O=object sensor, AM=ambient sensor, ECG=electrocardiograph)

| ID | Dataset | Type | #Subject | S. Rate | #Activity | #Sample | Sensor | Reference |
|---|---|---|---|---|---|---|---|---|
| D01 | OPPORTUNITY | ADL | 4 | 32 Hz | 16 | 701,366 | A, G, M, O, AM | (Ordóñez and Roggen, 2016) |
| D02 | Skoda Checkpoint | Factory | 1 | 96 Hz | 10 | 22,000 | A | (Plötz et al., 2011) |
| D03 | UCI Smartphone | ADL | 30 | 50 Hz | 6 | 10,299 | A, G | (Almaslukh et al., 2017) |
| D04 | PAMAP2 | ADL | 9 | 100 Hz | 18 | 2,844,868 | A, G, M | (Zheng et al., 2014) |
| D05 | USC-HAD | ADL | 14 | 100 Hz | 12 | 2,520,000 | A, G | (Jiang and Yin, 2015) |
| D06 | WISDM | ADL | 29 | 20 Hz | 6 | 1,098,207 | A | (Alsheikh et al., 2016) |
| D07 | DSADS | ADL | 8 | 25 Hz | 19 | 1,140,000 | A, G, M | (Zhang et al., 2015c) |
| D08 | Ambient kitchen | Food preparation | 20 | 40 Hz | 2 | 55,000 | O | (Plötz et al., 2011) |
| D09 | Darmstadt Daily Routines | ADL | 1 | 100 Hz | 35 | 24,000 | A | (Plötz et al., 2011) |
| D10 | Actitracker | ADL | 36 | 20 Hz | 6 | 2,980,765 | A | (Zeng et al., 2014) |
| D11 | SHO | ADL | 10 | 50 Hz | 7 | 630,000 | A, G, M | (Jiang and Yin, 2015) |
| D12 | BIDMC | Heart failure | 15 | 125 Hz | 2 | >20,000 | ECG | (Zheng et al., 2014) |
| D13 | MHEALTH | ADL | 10 | 50 Hz | 12 | 16,740 | A, C, G | (Ha and Choi, 2016) |
| D14 | Daphnet Gait | Gait | 10 | 64 Hz | 2 | 1,917,887 | A | (Hammerla et al., 2016) |
| D15 | ActiveMiles | ADL | 10 | 50-200 Hz | 7 | 4,390,726 | A | (Ravì et al., 2017) |
| D16 | HASC | ADL | 1 | 200 Hz | 13 | NA | A | (Hayashi et al., 2015) |
| D17 | PAF | PAF | 48 | 128 Hz | 2 | 230,400 | EEG | (Pourbabaee et al., 2017) |
| D18 | ActRecTut | Gesture | 2 | 32 Hz | 12 | 102,613 | A, G | (Yang et al., 2015) |
| D19 | Heterogeneous | ADL | 9 | 100-200 Hz | 6 | 43,930,257 | A, G | (Yao et al., 2017) |

(Vepakomma et al., 2015) combined activity and environment signal to recognize activities in a smart home, but the activities are constrained to body movements without more information on user emotion and state, which are also important. In the future, we expect there will be more research in this area.

### 5.2. Benchmark Datasets

We extensively explore the benchmark datasets for deep learning based HAR. Basically, there are two types of data acquisition schemes: *self data collection* and *public datasets*.

- *Self data collection:* Some work performed their own data collection (e.g. (Chen and Xue, 2015; Zhang et al., 2015b; Bhattacharya and Lane, 2016; Zhang et al., 2015a)). Very detailed efforts are required for self data collection, and it is rather tedious to process the collected data.
- *Public datasets:* There are already many public HAR datasets that are adopted by most researchers (e.g. (Plötz et al., 2011; Ravi et al., 2016; Hammerla et al., 2016)). By summarizing existing literature, we present several widely used public datasets in Table 3.

## 6. Summary and Discussion

Table 4 presents all the surveyed work in this article. We can make several observations based on the table.

**1) Sensor deployment and preprocessing.** Choosing the suitable sensors is critical for successful HAR. In surveyed literature, body-worn sensors serve as the most common modalities and accelerometer is mostly used. The reasons are two folds. Firstly, a lot of wearable devices such as smartphones or watches are equipped with an accelerometer, which is easy to access. Secondly, the accelerometer is competent to recognize many types of daily activities since most of them are simple body movements. Compared to body-worn sensors, object and ambient sensors are better at recognizing activities related to context and environment such as *having coffee*. Therefore, it is suggested to use body-worn sensors (mostly accelerometer+gyroscope) for ADL and sports activities. If the activities are pertaining to some semantic meaning but more than simple body movements, it is better to combine the object and ambient sensors. In addition, there are few public datasets for object and ambient sensors probably because of privacy issues and deployment difficulty of the data collecting system. We expect there will be more open datasets regarding those sensors.

Sensor placement is also important. Most body-worn sensors are placed on the dominant wrist, waist, and the dominant hip pocket. This placement strategy can help to recognize most common daily activities. However, when it comes to object and ambient sensors, it is critical to deploy them in a non-invasive way. Those sensors are not usually interacting with users directly, so it is critical to collect the data naturally and non-invasively.

Before using deep models, the raw sensor data need to be preprocessed accordingly. There are two important aspects. The first aspect is *sliding window*. The inputs should be cut into individual inputs according to the sampling rate. This procedure is similar to conventional PR approaches. The second one is *channels*. Different sensor modalities can be treated as separate channels, and each axis of a sensor can also be a channel. Using multi-channel could enhance the representation capability of the deep model since it can reflect the hidden knowledge of the sensor inputs.

**2) Model selection.** There are several deep models surveyed in this article. Then, a natural question arises: *which model is the best for HAR?* (Hammerla et al., 2016) did an early work by investigating the performance of DNN, CNN and RNN through 4,000 experiments on some public HAR datasets. We combine their work and our explorations to draw some conclusions: RNN and LSTM are recommended to recognize short activities that have natural order while CNN is better at inferring long-term repetitive activities (Hammerla et al., 2016). The reason is that RNN could make use of the time-order relationship between sensor readings, and CNN is more capable of learning deep features contained in recursive patterns. For multi-modal signals, it is better to use CNN since the features can be integrated through multi-channel convolutions (Zeng et al., 2014;

**Table 4. Summation of existing works based on the three aspects: sensor modality, deep model and application (in literature order)**

| Literature | Sensor Modality | Deep Model | Application | Dataset |
|---|---|---|---|---|
| (Almaslukh et al., 2017) | Body-worn | SAE | ADL | D03 |
| (Alsheikh et al., 2016) | Body-worn | RBM | ADL, factory, Parkinson | D02, D06, D14 |
| (Bhattacharya and Lane, 2016) | Body-worn, ambiemt | RBM | Gesture, ADL, transportation | Self, D01 |
| (Chen and Xue, 2015) | Body-worn | CNN | ADL | Self |
| (Chen et al., 2016b) | Body-worn | CNN | ADL | D06 |
| (Cheng and Scotland, 2017) | Body-worn | DNN | Parkinson | Self |
| (Edel and Köppe, 2016) | Body-worn | RNN | ADL | D01, D04, Self |
| (Fang and Hu, 2014) | Object, ambient | DBN | ADL | Self |
| (Gjoreski et al., 2016) | Body-worn | CNN | ADL | Self, D01 |
| (Guan and Ploetz, 2017) | Body-worn, object, ambient | RNN | ADL, smart home | D01, D02, D04 |
| (Ha et al., 2015) | Body-worn | CNN | Factory, health | D02, D13 |
| (Ha and Choi, 2016) | Body-worn | CNN | ADL, health | D13 |
| (Hammerla et al., 2015) | Body-worn | RBM | Parkinson | Self |
| (Hammerla et al., 2016) | Body-worn, object, ambient | DNN, CNN, RNN | ADL, smart home, gait | D01, D04, D14 |
| (Hannink et al., 2017) | Body-worn | CNN | Gait | Self |
| (Hayashi et al., 2015) | Body-worn, ambient | RBM | ADL, smart home | D16 |
| (Inoue et al., 2016) | Body-worn | RNN | ADL | D16 |
| (Jiang and Yin, 2015) | Body-worn | CNN | ADL | D03, D05, D11 |
| (Khan et al., 2017) | Ambient | CNN | Respiration | Self |
| (Kim and Toomajian, 2016) | Ambient | CNN | Hand gesture | Self |
| (Kim and Li, 2017) | Body-worn | CNN | ADL | Self |
| (Lane and Georgiev, 2015) | Body-worn, ambient | RBM | ADL, emotion | Self |
| (Lane et al., 2015) | Ambient | RBM | ADL | Self |
| (Lee et al., 2017) | Body-worn | CNN | ADL | Self |
| (Li et al., 2016a) | Object | RBM | Patient resuscitation | Self |
| (Li et al., 2016b) | Object | CNN | Patient resuscitation | Self |
| (Li et al., 2014) | Body-worn | SAE | ADL | D03 |
| (Liu et al., 2016) | Body-worn | CNN, RBM | ADL | Self |
| (Mohammed and Tashev, 2017) | Body-worn | CNN | ADL, gesture | Self |
| (Morales and Roggen, 2016) | Body-worn | CNN | ADL, smart home | D01, D02 |
| (Murad and Pyun, 2017) | Body-worn | RNN | ADL, smart home | D01, D02, D05, D14 |
| (Ordóñez and Roggen, 2016) | Body-worn | CNN, RNN | ADL, gesture, posture, factory | D01, D02 |
| (Panwar et al., 2017) | Body-worn | CNN | ADL | Self |
| (Plötz et al., 2011) | Body-worn, object | RBM | ADL, food preparation, factory | D01, D02, D08, D14 |
| (Pourbabaee et al., 2017) | Body-worn | CNN | PAF disease | D17 |
| (Radu et al., 2016) | Body-worn | RBM | ADL | D19 |
| (Ravi et al., 2016) | Body-worn | CNN | ADL, factory | D02, D06, D14, D15 |
| (Ravì et al., 2017) | Body-worn | CNN | ADL, factory, Parkinson | D02, D06, D14, D15 |
| (Ronao and Cho, 2015a,b, 2016) | Body-worn | CNN | ADL | D03 |
| (Sathyanarayana et al., 2016) | Body-worn | CNN, RNN, DNN | ADL, sleep | Self |
| (Singh et al., 2017) | Ambient | CNN, RNN | Gait | NA |
| (Vepakomma et al., 2015) | Body-worn, object, ambient | DNN | ADL | Self |
| (Walse et al., 2016) | Body-worn | DNN | ADL | D03 |
| (Wang et al., 2016b) | Body-worn, ambient | CNN | ADL, location | Self |
| (Wang et al., 2016a) | Object, ambient | SAE | ADL | NA |
| (Yang et al., 2015) | Body-worn, object, ambient | CNN | ADL, smart home, gesture | D01, D18 |
| (Yao et al., 2017) | Body-worn, object | CNN, RNN | Cartrack, ADL | Self, D19 |
| (Zebin et al., 2016) | Body-worn | CNN | ADL | Self |
| (Zeng et al., 2014) | Body-worn, ambient, object | CNN | ADL, smart home, factory | D01, D02, D10 |
| (Zhang et al., 2015a) | Body-worn | DNN | ADL | Self |
| (Zhang et al., 2015b) | Body-worn | RBM | ADL | Self |
| (Zhang et al., 2015c) | Body-worn | DBN | ADL, smart home | D01, D05, D07 |
| (Zhang et al., 2017b) | Object | CNN | Medical | Self |
| (Zhang et al., 2017a) | Body-worn | DNN | ADL | Self |
| (Zheng et al., 2016) | Body-worn | CNN, SAE | ADL | D04 |
| (Zheng et al., 2014) | Body-worn | CNN | ADL, heart failure | D04, D14 |

**Table 5. Performance comparison of existing deep models**

| Protocol | Model | Result | Reference |
|---|---|---|---|
| OPP 1 | b-LSTM-S | **92.70** | (Hammerla et al., 2016) |
| | CNN | 85.10 | (Yang et al., 2015) |
| | CNN | 88.30 | (Ordóñez and Roggen, 2016) |
| | DeepConvLSTM | 91.70 | (Ordóñez and Roggen, 2016) |
| OPP 2 | DBN | 73.20 | (Plötz et al., 2011) |
| | CNN | 76.80 | (Zeng et al., 2014) |
| | DBN | **83.30** | (Zhang et al., 2015c) |
| Skoda | CNN | 86.10 | (Zeng et al., 2014) |
| | CNN | 89.30 | (Alsheikh et al., 2016) |
| | DeepConvLSTM | **95.80** | (Ordóñez and Roggen, 2016) |
| UCI smartphone | CNN | 94.61 | (Ronao and Cho, 2016) |
| | CNN | **95.18** | (Jiang and Yin, 2015) |
| | CNN | 94.79 | (Ronao and Cho, 2015a) |
| | CNN | 90.00 | (Ronao and Cho, 2015b) |

Zheng et al., 2014; Ha et al., 2015). While adapting CNN, data-driven approaches are better than model-driven approaches as the inner properties of the activity signal can be exploited better when the input data are transformed into the virtual image. Multiple convolutions and poolings also help CNN perform better. RBM and autoencoders are usually pre-trained before being fine-tuned. Multi-layer RBM or SAE is preferred for more accurate recognition.

Technically there is no model which outperforms all the others in all situations, so it is recommended to choose models based on the scenarios. To better illustrate the performance of some deep models, Table 5 offers some results comparison of existing work on public datasets in Table 3[1]. In Skoda and UCI Smartphone protocols, CNN achieves the best performance. In two OPPORTUNITY protocols, DBN and RNN outperform the others. This confirms that no models can achieve the best in all tasks. Moreover, the hybrid models tend to perform better than single models (DeepConvLSTM in OPPORTUNITY 1 and Skoda). For a single model, CNN with shifted inputs (Fourier transform) generates better results compared to shifted kernels.

## 7. Grand Challenges

Despite the progress in previous work, there are still challenges for deep learning based HAR. In this section, we present those challenges and propose some feasible solutions.

**A. Online and mobile deep activity recognition.** Two critical issues are related to deep HAR: online deployment and mobile application. Although some existing work adopted deep HAR on smartphone (Lane et al., 2015) and watch (Bhattacharya and Lane, 2016), they are still far from online and mobile deployment. Because the model is often trained offline on some remote server and the mobile device only utilizes a trained model. This approach is neither real-time nor friendly to incremental learning. There are two approaches to tackle this problem: *reducing the communication cost between mobile and server*, and *enhancing computing ability of the mobile devices*.

**B. More accurate unsupervised activity recognition.** The performance of deep learning still relies heavily on labeled samples. Acquiring sufficient activity labels is expensive and time-consuming. Thus, *unsupervised* activity recognition is urgent.

- *Take advantage of the crowd.* The latest research indicates that exploiting the knowledge from the crowd will facilitate the task (Prelec et al., 2017). Crowd-sourcing takes advantage of the crowd to annotate the unlabeled activities. Other than acquiring labels passively, researchers could also develop more elaborate, privacy-concerned way to collect useful labels.
- *Deep transfer learning.* Transfer learning performs data annotation by leveraging labeled data from other auxiliary domains (Pan and Yang, 2010; Cook et al., 2013; Wang et al., 2017). There are many factors related to human activity, which can be exploited as auxiliary information using deep transfer learning. Problems such as sharing weights between networks, exploiting knowledge between activity related domains, and how to find more relevant domains are to be resolved.

**C. Flexible models to recognize high-level activities.** More complex high-level activities need to be recognized other than only simple daily activities. It is difficult to determine the hierarchical structure of high-level activities because they contain more semantic and context information. Existing methods often ignore the correlation between signals, thus they cannot obtain good results.

- *Hybrid sensor*. Elaborate information provided by the hybrid sensor is useful for recognizing fine-grained activities (Vepakomma et al., 2015). Special attention should be paid to the recognition of fine-grained activities by exploiting the collaboration of hybrid sensors.
- *Exploit context information*. Context is any information that can be used to characterize the situation of an entity (Abowd et al., 1999). Context information such as Wi-Fi, Bluetooth, and GPS can be used to infer more environmental knowledge about the activity. The exploitation of resourceful context information will greatly help to recognize user state as well as more specific activities.

**D. Light-weight deep models.** Deep models often require lots of computing resources, which is not available for wearable devices. In addition, the models are often trained off-line which cannot be executed in real-time. However, less complex models such as shallow NN and conventional PR methods could not achieve good performance. Therefore, it is necessary to develop light-weight deep models to perform HAR.

- *Combination of human-crafted and deep features*. Recent work indicated that human-crafted and deep features together could achieve better performance (Plötz et al., 2011). Some pre-knowledge about the activity will greatly contribute to more robust feature learning in deep models (Stewart and Ermon, 2017). Researchers should consider the possibility of applying two kinds of features to HAR with human experience and machine intelligence.
- *Collaboration of deep and shallow models.* Deep mod-

[1] OPP 1, OPP 2, Skoda, and UCI smartphone follow the protocols in (Hammerla et al., 2016), (Plötz et al., 2011), (Zeng et al., 2014), and (Ronao and Cho, 2016), respectively. OPP 1 used weighted f1-score; OPP 2, Skoda, and UCI smartphone used accuracy.

els have powerful learning abilities, while shallow models are more efficient. The collaboration of those two models has the potential to perform both accurate and light-weight HAR. Several issues such as how to share the parameters between deep and shallow models are to be addressed.

**E. Non-invasive activity sensing.** Traditional activity collection strategies need to be updated with more non-invasive approaches. Non-invasive approaches tend to collect information and infer activity without disturbing the subjects and requires more flexible computing resources.

- *Opportunistic activity sensing with deep learning.* Opportunistic sensing could dynamically harness the non-continuous activity signal to accomplish activity inference (Chen et al., 2016a). In this scenario, back propagation of deep models should be well-designed.

**F. Beyond activity recognition: assessment and assistant.** Recognizing activities is often the initial step in many applications. For instance, some professional skill assessment is required in fitness exercises and smart home assistant plays an important role in healthcare services. There is some early work on climbing assessment (Khan et al., 2015). With the advancement of deep learning, more applications should be developed to be beyond just recognition.

## 8. Conclusion

Human activity recognition is an important research topic in pattern recognition and pervasive computing. In this paper, we survey the recent advance in deep learning approaches for sensor-based activity recognition. Compared to traditional pattern recognition methods, deep learning reduces the dependency on human-crafted feature extraction and achieves better performance by automatically learning high-level representations of the sensor data. We highlight the recent progress in three important categories: sensor modality, deep model, and application. Subsequently, we summarize and discuss the surveyed research in detail. Finally, several grand challenges and feasible solutions are presented for future research.

## Acknowledgments

This work is supported in part by National Key R & D Program of China (No.2017YFB1002801), NSFC (No.61572471), and Science and Technology Planning Project of Guangdong Province (No.2015B010105001). Authors thank the reviewers for their valuable comments.

## References

Abowd, G.D., Dey, A.K., Brown, P.J., Davies, N., Smith, M., Steggles, P., 1999. Towards a better understanding of context and context-awareness, in: International Symposium on Handheld and Ubiquitous Computing, Springer. pp. 304–307.

Almaslukh, B., AlMuhtadi, J., Artoli, A., 2017. An effective deep autoencoder approach for online smartphone-based human activity recognition. International Journal of Computer Science and Network Security 17, 160.

Alsheikh, M.A., Selim, A., Niyato, D., Doyle, L., Lin, S., Tan, H.P., 2016. Deep activity recognition models with triaxial accelerometers. AAAI workshop .

Bao, L., Intille, S.S., 2004. Activity recognition from user-annotated acceleration data, in: International Conference on Pervasive Computing, Springer. pp. 1–17.

Bengio, Y., 2013. Deep learning of representations: Looking forward, in: International Conference on Statistical Language and Speech Processing, Springer. pp. 1–37.

Bhattacharya, S., Lane, N.D., 2016. From smart to deep: Robust activity recognition on smartwatches using deep learning, in: 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), IEEE. pp. 1–6.

Bulling, A., Blanke, U., Schiele, B., 2014. A tutorial on human activity recognition using body-worn inertial sensors. ACM Computing Surveys (CSUR) 46, 33.

Chavarriaga, R., Sagha, H., Calatroni, A., Digumarti, S.T., Tröster, G., Millán, J.d.R., Roggen, D., 2013. The opportunity challenge: A benchmark database for on-body sensor-based activity recognition. Pattern Recognition Letters 34, 2033–2042.

Chen, Y., Gu, Y., Jiang, X., Wang, J., 2016a. Ocean: a new opportunistic computing model for wearable activity recognition, in: UbiComp, ACM. pp. 33–36.

Chen, Y., Xue, Y., 2015. A deep learning approach to human activity recognition based on single accelerometer, in: Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on, IEEE. pp. 1488–1492.

Chen, Y., Zhong, K., Zhang, J., Sun, Q., Zhao, X., 2016b. Lstm networks for mobile human activity recognition .

Cheng, W.Y., Scotland, A.e.a., 2017. Human activity recognition from sensor-based large-scale continuous monitoring of parkinsons disease patients, in: Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2017 IEEE/ACM International Conference on, pp. 249–250.

Cook, D., Feuz, K.D., Krishnan, N.C., 2013. Transfer learning for activity recognition: A survey. Knowledge and information systems 36, 537–556.

Edel, M., Köppe, E., 2016. Binarized-blstm-rnn based human activity recognition, in: Indoor Positioning and Indoor Navigation (IPIN), 2016 International Conference on, IEEE. pp. 1–7.

Fang, H., Hu, C., 2014. Recognizing human activity in smart home using deep learning algorithm, in: Chinese Control Conference (CCC), pp. 4716–4720.

Gjoreski, H., Bizjak, J., Gjoreski, M., Gams, M., 2016. Comparing deep and classical machine learning methods for human activity recognition using wrist accelerometer, in: IJCAI-16 workshop on Deep Learning for Artificial Intelligence (DLAI).

Guan, Y., Ploetz, T., 2017. Ensembles of deep lstm learners for activity recognition using wearables. arXiv preprint arXiv:1703.09370 .

Ha, S., Choi, S., 2016. Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors, in: Neural Networks (IJCNN), 2016 International Joint Conference on, IEEE. pp. 381–388.

Ha, S., Yun, J.M., Choi, S., 2015. Multi-modal convolutional neural networks for activity recognition, in: Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on, IEEE. pp. 3017–3022.

Hammerla, N.Y., Fisher, J., Andras, P., Rochester, L., Walker, R., Plötz, T., 2015. Pd disease state assessment in naturalistic environments using deep learning, in: Twenty-Ninth AAAI Conference on Artificial Intelligence.

Hammerla, N.Y., Halloran, S., Ploetz, T., 2016. Deep, convolutional, and recurrent models for human activity recognition using wearables, in: IJCAI.

Hannink, J., Kautz, T., Pasluosta, C.F., Gaßmann, K.G., et al., 2017. Sensor-based gait parameter extraction with deep convolutional neural networks. IEEE journal of biomedical and health informatics 21, 85–93.

Hayashi, T., Nishida, M., Kitaoka, N., Takeda, K., 2015. Daily activity recognition based on dnn using environmental sound and acceleration signals, in: Signal Processing Conference (EUSIPCO), 23rd European, pp. 2306–2310.

Hinton, G.E., Osindero, S., Teh, Y.W., 2006. A fast learning algorithm for deep belief nets. Neural computation 18, 1527–1554.

Hu, L., Chen, Y., Wang, S., Wang, J., Shen, J., Jiang, X., Shen, Z., 2016. Less annotation on personalized activity recognition using context data, in: UIC, pp. 327–332.

Inoue, M., Inoue, S., Nishida, T., 2016. Deep recurrent neural network for mobile human activity recognition with high throughput. arXiv preprint arXiv:1611.03607 .

Jiang, W., Yin, Z., 2015. Human activity recognition using wearable sensors by deep convolutional neural networks, in: MM, ACM. pp. 1307–1310.

Khan, A., Mellor, S., Berlin, E., Thompson, R., McNaney, R., Olivier, P., Plötz, T., 2015. Beyond activity recognition: skill assessment from accelerometer data, in: UbiComp, ACM. pp. 1155–1166.