Development of a Hybrid Deep-Learning Model for the Human Activity Recognition based on the Wristband Accelerometer Signals☆

Seungmin Jeong[1]   Dongik Oh[2*]

ABSTRACT

This study aims to develop a human activity recognition (HAR) system as a Deep-Learning (DL) classification model, distinguishing various human activities. We solely rely on the signals from a wristband accelerometer worn by a person for the user's convenience. 3-axis sequential acceleration signal data are gathered within a predefined time-window-slice, and they are used as input to the classification system. We are particularly interested in developing a Deep-Learning model that can outperform conventional machine learning classification performance. A total of 13 activities based on the laboratory experiments' data are used for the initial performance comparison. We have improved classification performance using the Convolutional Neural Network (CNN) combined with an auto-encoder feature reduction and parameter tuning. With various publically available HAR datasets, we could also achieve significant improvement in HAR classification. Our CNN model is also compared against Recurrent-Neural-Network(RNN) with Long Short-Term Memory(LSTM) to demonstrate its superiority. Noticeably, our model could distinguish both general activities and near-identical activities such as sitting down on the chair and floor, with almost perfect classification accuracy.

☞ keyword : Activities of Daily Living, Human Activity Recognition, Smartwatch, Accelerometer, Machine Learning, Activity Classification, Feature Extraction, Feature Reduction

1. Introduction

Human Activity Recognition (HAR), research for automatically predicting human behavior, has been continuously evolving over the past few decades. The main reason is that HAR can be used in various fields necessary for modern people, such as in health monitoring, sports monitoring, gaming, social security, and smart home, etc. [1]. In particular, the healthcare field, which is engaged with aging, shifts from the concept of cure to care; it is becoming essential to managing patients as well as people's daily lives. Accordingly, automatic detection of human behavior is emerging as an important problem [2].

Two main methods of HAR are a method using a sensor

and that of using an image. Various actions can be classified through the action image, but this method has two main drawbacks the computational overhead is high, and human privacy can be invaded [3]. For the methods using sensors, there are two major approaches. One is installing sensors in the surrounding environment and using the signal data acquired from them for behavior recognition. The other is attaching sensors to a person's body and analyzing and utilizing the action's signal. Of course, these two methods can be combined. Installing sensors in the surrounding environment guarantee the patient's convenience, but the problem is that the recognition environments and the recognizable behaviors are limited [4].

On the other hand, attaching a sensor to a person's body has the advantage of acquiring signals from all human actions. Many commonly available devices, such as smartphones, smartwatches(bands), and smartshirts, have various sensors installed. However, a wristband-type is the most convenient and familiar device for everyday life, and HAR research actively uses them [4]. In this study, a 3-axis acceleration sensor provided by almost all wearable devices of today, mounted on a wristband-type device, is used for

1 Department of Software Convergence, Soonchunhyang University, Asan-si, Chungnam-do, 336-745, Korea
2 Department of Medical IT Engineering, Soonchunhyang University, Asan-si, Chungnam-do, 336-745, Korea
* Corresponding author (dohdoh@sch.ac.kr)

sensing. The signal acquired is used to recognize human behavior. In this way, we have attempted to develop a HAR system that minimizes discomfort in human behavior and is affordable to the general public.

In a sensor-based HAR, it is common to classify acquired data using an artificial intelligence model. As an artificial intelligence model, the Conventional Machine Learning (CML) model and the Deep Learning (DL) model are popular. CML models have to go through extracting and condensing signal features with expert knowledge before applying the classification model. In DL, this process is carried out by the model itself. It has been reported that CML is more appropriate when the size of the dataset is small, and DL is more appropriate when the size is big [5].

This study tries to derive better recognition results by applying the hybrid DL model to the data acquired from our previous CML-based study [6]. As a result, better classification results are obtained; 99.8% for 13 self-designated actions and 99.5% for 'pamap2' public dataset actions. As a result of applying this model to other public datasets (UCI-mobile, USC-HAD), the hybrid DL-based HAR system developed in this study generally shows a higher average recognition rate than the single DL-based system (90.2% vs. 99.3%). Especially in the dataset analysis containing high-level behaviors such as *ironing* and *vacuuming* (81.1% vs.98.8%), and in the dataset analysis containing similar behaviors such as *walking forward, left, and right* (87.9% vs. 99.1%), the model proposed in this study shows much better performance than others.

For the DL model, three representative DL methods, namely, Recurrent Neural Network (RNN) with Long and Short-Term Memory (LSTM), Deep Neural Network (DNN), and Convolutional Neural Network(CNN), are constructed, and their performances are compared. The best results are acquired from the CNN model. Even though DL models automatically extract and utilize features, we extend the model by pre-extracting deep features using Adversarial Auto-encoder (AAE), constructing a hybrid-DL model. The model produces a near-perfect HAR performance. The process summary of the best-performing method is depicted in the lower part of Figure 1.

We compose the paper as follows; In Section 1, the

background, purpose, and objectives of this study are explained. In Section 2, the typical components and classification processes of the sensor-based ML HAR model and their performance are examined. Section 3 describes the details of the CNN model incorporating the AAE preprocessing developed in this study. In Section 4, we report the performance comparisons of DNN, CNN, and RNN models on various experimental data. Finally, Section 5 concludes the study.

2. Background and Related Work

Various studies on artificial intelligence for classifying HAR using a sensor mounted on a wearable device have been conducted [1]. In the researches, the CML method and the DL method are widely used. Recently, the DL method is more common. This section compares the structure and the process of the HAR system for both CML and DL schemes. We will also elaborate on what cases are appropriate for each scheme and summarize their average performance.

2.1 Conventional Machine Learning based Approach

The outline of the ML-based HAR system is shown in the upper part of Figure 1.

The wearable device acquires data and preprocesses them. At this time, a person defines the features necessary to classify the activity and, if necessary, selects or combines them [4]. Finally, the activity is classified by an ML model based on the features. Representative ML models are k-Nearest Neighbors (kNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Extreme Gradient Boosting (XGB) [7].

According to [7], this classification has been the mainstream of the ML method for the past 30 years. With the wristband-type sensor signal, 89.4% average accuracy is reported with the methods. [7] says that these models are practical to apply when the training data size is relatively small. The model's performance may also be affected by the uncertainty that adequate features are extracted.
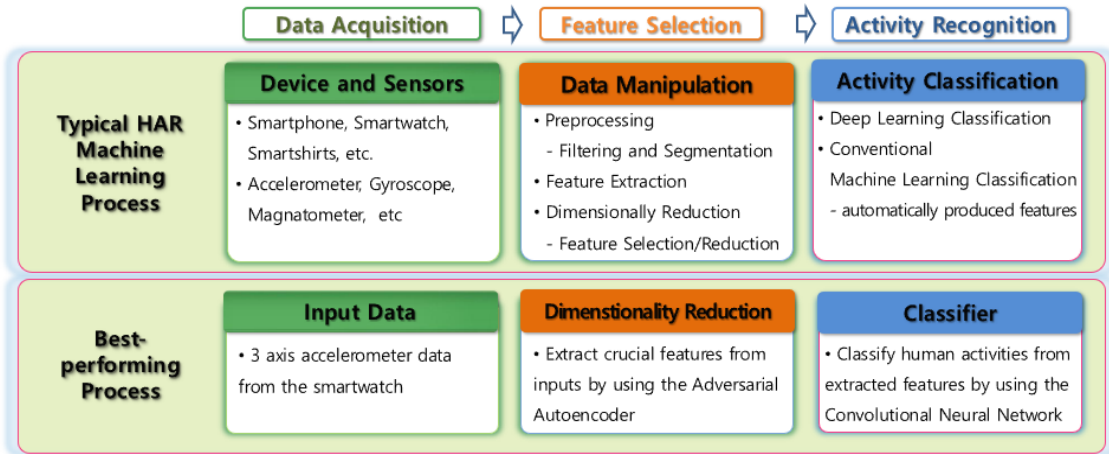
（Figure 1) The outline of Typical HAR Machine Learning Process and our Best-performing Method

2.2 Deep Learning based Approach

In this approach, the same process as in the upper Figure 1 is performed, but the data manipulation, which is in the middle, is not manually performed. Extraction and reduction of the features required in the DL model are automatic [12]. Hidden low-level features that the human cannot manually find can be artificially extracted, and the models use them to derive better classification performance [7]. Representative DL methods used for HAR include DNN, Autoencoder, CNN, RNN with LSTM, and DBN [8]. When the behavior is classified using various sensors, including acceleration sensors, an average of 93.7% based on CNN, 95% for RNN, and 91.5% for LSTM are reported [8].

2.3 HAR-related Earlier Research

We have used the values from the 3-axis acceleration sensor mounted on a wristband in our earlier research. As shown in the upper part of Figure 1, we went through the preprocessing process to extract features. After that, dimensionality reduction was performed through Adversarial Auto-Encoder (AAE), and their performance was compared by classifying them using various CML classification models. As a result, we derived near-perfect classification results with models using AAE and XGBoost for 13 self-designated actions (99.6%) and publicly available 'pamap2' actions data

(98.7%) [6].

In this study, focusing on the idea of automatic feature extraction, we try to determine whether the DL model's performance can also be improved by applying AAE preprocessing.

3. HAR System

This section describes the various components of the HAR system we have developed in this study. The system consists of the processes of data acquisition, dimensionality reduction, and activity classification. The description of each part is elaborated in this section.

3.1 Data Acquisition

The data required for this study is acquired using a wristband-type wearable device with a built-in 3-axis acceleration sensor. The wearable device quantifies 3-axis acceleration sensor values that change during the user's behavior and transmits them to the server through Bluetooth communication. The sensor's sampling rate is 40Hz, which is the same as the previous study's optimal sampling rate [6].

3.2 Dimensionality Reduction

As shown in section 2.1, in various HAR studies using wearable sensors, data is refined through different preprocessing, and feature values are extracted to facilitate classification. This process plays a significant role in improving classification accuracy. In this study, this process is conducted through the AAE network. Auto-Encoder is a network that can automatically extract features [9]. It decomposes the input into lower-dimensional data, and the original data can be recovered from them; it learns and finds the appropriate features from the process [10]. However, the extracted features are sometimes fragmented into values that are difficult to classify because there is no fixed rule for feature learning. The adversarial autoencoder is an autoencoder that regularizes an aggregated posterior q(z) to an arbitrary prior p(z). An adversarial network is attached on top of the hidden code vector of the autoencoder. It is the adversarial network that guides q(z) to match p(z) [11].

In this study, a 3D Swiss Roll sample is given to the AAE model, and features are extracted accordingly. Through this process, 128 data per window for the 3-axis (a total of 384 data) are compressed into three crucial feature values.

The use of a classification model after the dimensionality reduction using AAE is beneficial in classifying acceleration-based HAR. Most notably, the original signal can be used as input to the model without creating hand-crafted features. Since the model can find hidden features that are difficult to find in a general DL model, it distinguishes each human activity better. Therefore, recently various attempts have been made for this kind of (hybrid-type) DL methods. [12]

3.3 Activity Classification

We use three representative DL classifier models: DNN, RNN, and CNN for the study.

DNN is a form of a neural network in which several dense layers are stacked wide and deep. This ML model is used for various classifications [13]. In this study, the DNN model is constructed with six dense layers. The Swish Activation function is used in all dense layers except the last one, at which the Sigmoid function is applied. If the dense

layer is deeply stacked, more complex data patterns can be grasped, but vanishing gradient problems may arise in the learning process. It is essential to find an adequate depth to obtain good accuracy while avoiding the issue. We determined to use six dense layers, which produced the best results among 2 to 14 layers comparison experiments.

RNN is an ML method widely used to predict results based on time series data [14]. The RNN structure used in this study consists of one LSTM layer and three dense layers. When the RNN is used alone, it is generally expected to perform better than other DL models. However, if the data with the AAE dimensionality reduction is used as input, the time series characteristics may disappear so that the RNN model's strength fades out.

CNN is an ML method widely used in image processing. However, it can be used as an ML method in various domains if signals can be converted to images [15]. Generally, there are multiple convolutional layers and pooling layers between the input and output layers. Dense layers are placed next to the convolutional layers [16].

In this study, two types of CNN networks are considered. The first is a one-dimensional hybrid CNN that uses AAE. It is consists of one layer of 1D convolution layer and three layers of dense layers. The second is a two-dimensional CNN without AAE. It consists of three layers of 2D convolution layers and three density layers. The hybrid DL model using AAE and 1D CNN shows the highest average accuracy for all datasets used in this study, so it is used as our CNN model. The structure of the hybrid AAE + CNN is shown in Figure 2.

This model decomposes the input signal x into smaller feature z through the encoder network and learns the features while reconstructing it through the decoder network. Even though z has fewer dimensions than x, it can contain meaningful enough information of the original signal. The Adversarial Network learns to distinguish created $q(z)$ from the target sample $p(z)$, and the encoder network learns to prevent $q(z)$ from being ambiguous. In this way, the model creates a z similar to the $p(z)$, solving the fracture problem, and makes the classification easier.
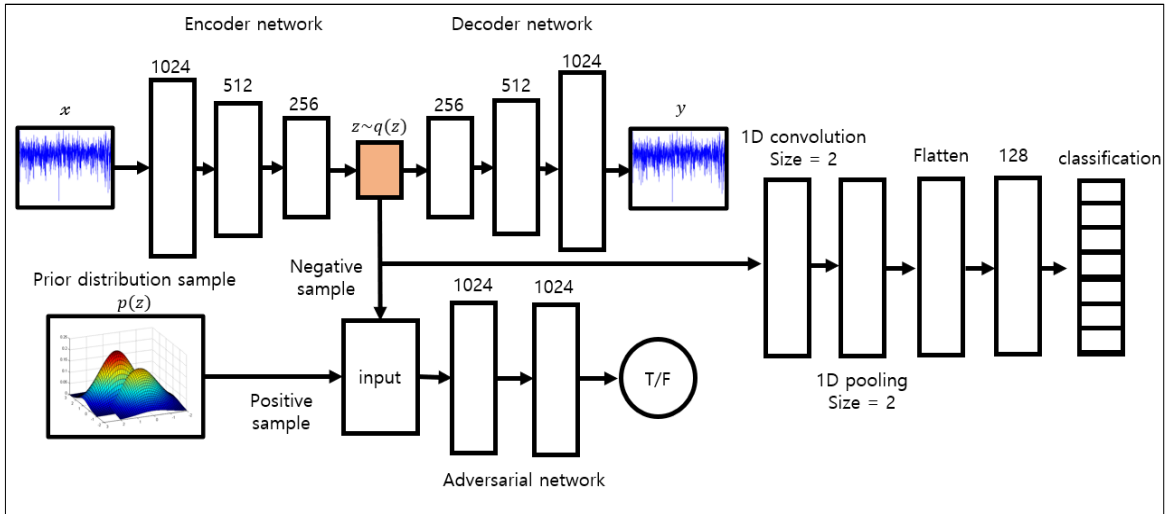
(Figure 2) The structure of the AAE+CNN hybrid Deep Learning model

4. Experimental Results and Analysis

In this study, a total of four datasets are tested; a lab dataset created from an experiment and three public datasets. The public datasets are pamap2 [17], UCI-mobile [18], and USC-HAD [19]. Regardless of the type and number of sensors used, all datasets include the 3-axis acceleration sensor data mounted on the left wrist.

All the data are converted to 3.2-second windows. All data sampling is adjusted to 40Hz to have the same data format as the lab dataset.

Table 1 shows an overview of the datasets used. This study evaluates the DL models' classification performance mentioned in Section 3.3, namely the DNN, CNN, RNN, and their AAE hybrid versions.

4.1 Data Acquisition

This section describes the data and features of each dataset. The summary of each dataset is given in Table 1.

To obtain the lab dataset, six subjects wearing a smartwatch-type sensor with a built-in 3-axis acceleration sensor on their left wrist naturally performed 13 types of activities. The sensor collected 3-axis acceleration data 40

times per second. The data are transmitted to the artificial intelligence server. Through this experiment, a total of 7185 windows of data are collected for the 13 activities, which consist of *stationary, walking, running, stair-up/down, sitting/standing on floor/chair, lying/getting up on floor/bed.*

The pamap2 dataset is collected while subjects wear an IMU (Inertial Measurement Unit) on their wrists, chest, and knees. It contains 12 activity data. They are *lying, sitting, standing, walking, running, cycling, Nordic walking, ascending/descending stairs, vacuum cleaning, ironing, rope jumping.* We use only the data obtained from the wrist-mounted sensor. The dataset consists of a total of 6346 windows.

The UCI-mobile dataset provides data from six activities; *walking, sitting, standing, lying, walking upstairs/downstairs.* Data are collected for a total of 10299 windows. A smartphone mounted on the wrist is used for data collection.

The USC-HAD dataset provides 12 activity data from a 6-axis acceleration and a gyro sensor. Handled activities are *walking forward/left/right/upstairs/downstairs, jumping up, running forward, sitting, standing, sleeping, elevator-up, elevator-down.* The *elevator-up* and *elevator-down* are treated as the same activity; with the acceleration sensor only, it is impossible to detect ascent and descent in an elevator moving at a constant speed. A total of 9057 windows data

(Table 1) Brief introduction of datasets usedin this study.

Dataset	Type of Activities	Number of Windows	Including Similar Activities
Lab Data[6]	13 activities including *sitting/standing on floor/chair, lying/getting up on floor/bed*	7185	O
Pamap2[17]	12 activities including *ironing, vacuuming, regular walking, Nordic walking*	6346	X
UCI-mobile[18]	Six activities including *walking, sitting, standing, lying, walking upstairs/downstairs*	10299	X
USC-HAD[19]	12 activities including *walking forward/left/right/upstairs/downstair*	9057	O

are collected.

4.2 Experiment Results

In this experiment, we measured the performance of a total of 6 DL models, including DNN, CNN, RNN, and their AAE hybrid models, against four datasets. For each dataset, train, validation, and test data are divided into the ratio of 6:2:2. We ran each model ten times, and the results are averaged. The results of the experiment are shown in Table 2.

(Table 2) The result of experiments

Dataset Model	Lab Data	Pamap2	UCI Mobile	USC-HAD
DNN only	86.1%	69.5%	97.5%	79.2%
RNN only	92.6%	82.1%	86.1%	86.5%
CNN only	93.6%	81.1%	98.2%	87.9%
AAE + DNN	100 %	92.8%	99.3%	96.2%
AAE + RNN	100%	96.3%	99.3%	98.8%
AAE + CNN	100%	98.8%	99.4%	99.1%

When the DNN, RNN, and CNN models are used alone, the classification accuracy is low in general. Only DNN and CNN show a meaningful classification accuracy of 97.57% and 98.25% on the UCI-mobile dataset. We infer that the poor performance is because the input contains a lot of data with little meaning or relationship that are very difficult to identify.

On the other hand, with the dimensionality reduction of the AAE network, the DL classifier models show higher overall classification performance. The RNN model is better

than the general DNN model, but the best performance is acquired with the CNN model on all datasets. It showed 100% classification accuracy on the lab dataset and 99.4% accuracy on the UCI-mobile dataset for basic activities. However, even with pamap2 and USC-HAD datasets, which include more complex activities that are difficult to distinguish, the hybrid-CNN model showed the highest accuracy at 98.8% and 99.1%, respectively. This result is far better than the 96.8% (the lab dataset) accuracy achieved in the previous CML-based HAR model [6].

5. Conclusion

In this study, we developed an efficient hybrid DL classification model for HAR. In most CML methods, collected data must undergo preprocessing step (e.g., feature selection, feature extraction). However, this study's hybrid-DL model replaces the preprocessing step with AAE, pulling crucial features out from the dimension of the wristband-type acceleration sensor data. Features extracted by AAE have low dimensionality, essential information that can be useful for classification. Hence, we could significantly improve classification accuracy by using these crucial features from the AAE dimensionality reduction model as an input for DL classification models.

We tested three DL methods (DNN, RNN, and CNN) combined with AAE against four datasets made for the human activity classification; one lab dataset and three publically available datasets. Our experiment result shows that the hybrid-DL model outperforms models using only the DL classifier (average accuracy: 98.3% versus 86.6%).

Notably, the hybrid deep learning model using AAE and CNN produces near-perfect classification performance (average accuracy: 99.3%), which is the best performance in the literature as far as we know.

In the future study, we plan to apply our hybrid-DL model to domains other than HAR to demonstrate its applicability and superiority.

References

[1] Guan Yuan, Zhaohui Wang, Fanrong Meng, Qiuyan Yan, Shixiong Xia, ″An overview of human activity recognition based on smartphone,″ Sensor Review, 2018. https://doi.org/10.1108/SR-11-2017-0245

[2] Yassine, Abdulsalam, Shailendra Singh, and Atif Alamri. ″Mining human activity patterns from smart home big data for health care applications,″ *IEEE Access* 5, 13131-13141, 2017. https://doi.org/10.1109/ACCESS.2017.2719921

[3] Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., and Darrell, T., ″Long-term recurrent convolutional networks for visual recognition and description,″ In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2625–2634, 2015. https://doi.org/10.1109/cvpr.2015.7298878

[4] Yan Wang, Shuang Cang, Hongnian Yu, ″A survey on wearable sensor modality centred human activity recognition in health care,″ Journal of Expert Systems With Applications, Vol. 137, pp. 167-190, 2019. https://doi.org/10.1016/j.eswa.2019.04.057

[5] O. D. Lara and M. A. Labrador, ″A Survey on Human Activity Recognition using Wearable Sensors,″ IEEE Communications Surveys & Tutorials, Vol. 15, no. 3, pp. 1192-1209, Third Quarter 2013. https://doi.org/10.1109/SURV.2012.110112.00192

[6] S. Jeong, C. Choi, D. Oh, ″Development of a Machine-Learning based Human Activity Recognition System including Eastern-Asian Specific Activities,″ Journal of Internet Computing and Services, Vol. 21, No. 4, pp. 127-135, Aug. 2020. https://doi.org/10.1109/ACCESS.2017.2719921

[7] Florenc Demrozi, Graziano Pravadelli, Azra Bihorac, Parisa Rashidi, ″Human Activity Recognition using Inertial, Physiological and Environmental Sensors: A Comprehensive Survey,″ Preprint http://arxiv.org (arXiv:2004.08821v1) Apr. 2020.

[8] Jindong Wang, Yiqiang Chen, Huji Hao, Xiaohui Peng, Lisha Hu, ″Deep learning for sensor-based activity recognition: A survey,″ Pattern Recognition Letters, Volume 119, Pages 3-11, ISSN 0167-8655, 2019. https://doi.org/10.1016/j.patrec.2018.02.010

[9] Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. ″Reducing the dimensionality of data with neural networks,″ Science 28, 504-507, 2006. https://doi.org/10.1126/science.1127647

[10] Wang, Yasi, Hongxun Yao, and Sicheng Zhao. ″Auto-encoder based dimensionality reduction,″ Neurocomputing 184, pp. 232-242, 2016. https://doi.org/10.1016/j.neucom.2015.08.104

[11] Makhzani, Alireza, et al. Adversarial Autoencoders, 2016. https://arxiv.org/abs/1511.05644

[12] Abbaspour, Saedeh, et al. ″A comparative analysis of hybrid deep learning models for human activity recognition,″ *Sensors* 20.19, 5707, 2020. https://doi.org/10.3390/s20195707

[13] Ronao, Charissa Ann, and Sung-Bae Cho. ″Human activity recognition with smartphone sensors using deep learning neural networks,″ Expert systems with applications 59, pp. 235-244, 2016. https://doi.org/10.1016/j.eswa.2016.04.032

[14] Murad, Abdulmajid, and Jae-Young Pyun. ″Deep recurrent neural networks for human activity recognition,″ Sensors 17.11, 2556, 2017. https://doi.org/10.3390/s17112556

[15] O'Shea, Keiron, and Ryan Nash. ″An introduction to convolutional neural networks,″ preprint arXiv: 1511.08458, 2015. https://arxiv.org/abs/1511.08458

[16] Albawi, Saad, Tareq Abed Mohammed, and Saad Al-Zawi. ″Understanding of a convolutional neural network,″ International Conference on Engineering and Technology (ICET). IEEE, 2017. https://doi.org/10.1109/ICEngTechnol.2017.8308186

[17] Reiss, Attila, and Didier Stricker. "Introducing a new benchmarked dataset for activity monitoring," 2012 16th International Symposium on Wearable Computers. IEEE, 2012. https://doi.org/10.1109/ISWC.2012.13

[18] Anguita, Davide, et al. "A public domain dataset for human activity recognition using smartphones," Esann Vol. 3. 2013.

https://doi.org/10.1109/ICEngTechnol.2017.8308186

[19] Zhang, Mi, and Alexander A. Sawchuk. "USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors," Proceedings of the 2012 ACM Conference on Ubiquitous Computing. 2012. https://doi.org/10.1145/2370216.2370438

◑ 저 자 소 개 ◐

정 승 민(Seungmin Jeong)
2021년~현재 순천향대학교 소프트웨어융합과 석사 재학
2021년 순천향대학교 의료IT공학과 학사
관심분야 : u-Healthcare 시스템, 사물인터넷, Machine-Learning, 패턴인식
E-mail : sontneptm2@gmail.com

오 동 익(Dongik Oh)
1985년 뉴욕시립대학교 전산학과 (이학사)
1989년 플로리다주립대학교 대학원 전산학과 석사 (이학석사)
1997년 플로리다주립대학교 대학원 전산학과 박사 (이학박사)
1997년~2007년 순천향대학교 컴퓨터학부 교수
2007년~현재 순천향대학교 의료IT공학과 교수
관심분야 : u-Healthcare, Machine-Learning, 임베디드 시스템, 운영체제, 프로그래밍 언어
E-mail : dohdoh@sch.ac.kr