2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC) | 978-1-6654-3553-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ITC-CSCC52171.2021.9501477

Sensor-based Complex Human Activity Recognition from Smartwatch Data using Hybrid Deep Learning Network

Sakorn Mekruksavanich[1] and Anuchit Jitpattanakul[2]
[1]*Department of Computer Engineering, School of Information and Communication Technology*
University of Phayao, Phayao, Thailand
sakorn.me@up.ac.th
[2]*Intelligent and Nonlinear Dynamic Innovations Research Center, Department of Mathematics*
Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
anuchit.j@sci.kmutnb.ac.th

***Abstract*—Sensor-based human activity recognition (HAR) research is being used for tasks like healthcare tracking, fall detection, and misbehavior prevention. Because of the sophistication of hand gesture signals, complex human activity (CHA) recognition is a difficult task in HAR research. Compared to simple human behavior (SHA), the CHA has more input and long sequential information to deal with. To solve the CHA problem, we proposed a hybrid deep learning model that combines a CNN network and an LSTM network in this paper. The proposed model is an end-to-end model that automatically extracts high features. The model is tested on the daily human activity (DHA) dataset, which is a publicly accessible dataset of complex human activities. The results of the experiments show that the proposed hybrid deep learning model outperforms the current state-of-the-art recognition model.**

***Keywords*—smartwatch sensors, accelerometer sensors, activity recognition, deep learning, CNN, LSTM**

I. Introduction

Human Activity Recognition (HAR) is the issue of identifying a person's physical activity based on a trace of movement within a specific environment [1]. Daily physical activities, such as walking, laying, sitting, standing, and ascending stairs, are recognized as normal physical motions and represent person category of action [2], [3]. Since motion sensors are ubiquitous and remote sensing is becoming one of the most important part of human-machine collaboration, HAR using motion sensors is one of the most challenging research topics of the last decade [4]–[7].

Wearable technology are one of the most valuable tools in modern daily lives, and they are becoming more capable of performing market expectations and demands as technology progresses. To make these smart devices more functional and effective, wearable designers add new components and enhancements to the device. Most smartphones present a variety of embedded sensors, providing for the acquisition of vast amounts of information about the person's daily life activities. Sensors serve an important in allowing wearable technology more functional and aware of their surroundings [8]–[10].

Almost all smart-wearable system manufacturers use an accelerometer as a standard sensor. Accelerometers are sensors that track the acceleration of moving objects along reference axes. They are especially useful at tracking behaviors like walking, running, sitting, standing, and climbing because they involve repetitive body movements. The accelerometer's data can be processed to identify sudden changes in motion [11]. Another sensor that has become standard hardware for smartphones is the gyroscope, which uses gravity to determine orientation. The device's position and orientation can be determined using the gyroscope signals [12], [13].

Deep learning is a broad term used to refer to traditional neural networks that retrieve and classify features using multiple layers of nonlinear cognitive processing, with each level modifying the results of the previous level. Deep learning approaches have overcome many conventional machine learning techniques in many research areas, such as image recognition and voice classification. Thus according deep learning techniques, convolutional neural networks (CNNs) are a type of deep neural network that also can perform as feature extractors, stacking multiple convolutional operators to develop a hierarchy of functionally more demanding systems. These systems can learn multiple layers of role hierarchies continuously. Long-short-term memory (LSTMs) neural networks are recurrent networks with a storage for modeling temporal relationship in series data challenges. The combination of CNNs and LSTMs in a single system has indeed delivered state-of-the-art findings in the speech recognition domain [14], whereas modeling temporal information is required. This framework can recognize how features extracted across convolutional operations change with time.

So, this paper proposes a multiscale CNN-LSTM network, a novel hybrid deep learning framework based on CNN and LSTM networks. The aim of this study is to determine the best window sizes for the HAR problem through a series of experiments. The proposed hybrid model outperforms the traditional state-of-the-art recognition system, according to the findings of the experiments.

The remainder of the paper is laid out in the following manner. A survey of sensor-based HAR and deep learning approaches is presented in Section II. The proposed hybrid deep learning approach is explained in Section III. This section also includes an overview of the dataset. The experimental design, as well as experimental findings, are demonstrated in Section IV along with a comparison to the current deep learning model. Section V brings the paper to a conclusion and discusses potential future paths.

II. Research Background

A. Human Activity Recognition: HAR

HAR seeks to comprehend human activity so that computer systems can provide constructive assistance to persons based on their needs. Moving, seating, operating, and traveling are examples of human activities that can be defined as a series of acts carried out by a person over a duration of time in compliance with a particular protocol. Data selection, data segmentation, feature extraction, model training, and classification are the five basic steps in building a HAR system in machine learning, as shown in Fig. 1.

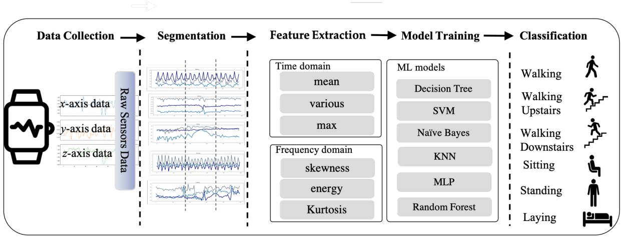

Fig. 1. Sensor-based HAR with conventional machine learning approaches.

The relevant paragraphs provide a more thorough description of each process. Data gathering from wearable devices that can consistently acquire sensor data as participants conduct predefined tasks is the first step in the HAR system. Sensor data must always be partitioned based on the fact that it is usually defined as time-series data. The sensor data is partitioned into equal-length data segments with a set window size and overlapping relation. The most critical step is feature extraction, since it determines the overall efficiency of the classification method. Domain experts systematically extract heuristic or handcrafted characteristics in both of time domains and frequency domains using conventional machine learning approaches. Some time-domain features exist, including maximum, minimum, mean, correlation, SD, and so on. There are also a number of frequency domain characteristics. Handcrafted features, on the other hand, have certain restrictions in both domains. Following that, these characteristics are used to employ a classification method [15]–[17].

Deep learning has the potential to solve the feature extraction challenge that plagues traditional machine learning. Fig. 2 depicts how HAR with deep learning operates for various types of networks. In the deep learning approach, the feature extraction and model training processes are carried out concurrently. Rather than just being physically hand-crafted as in traditional machine learning methods, the features can be acquired dynamically across the network.

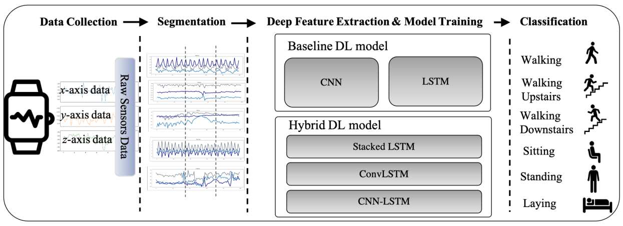

Fig. 2. Sensor-based HAR with deep learning approaches.

B. Convolutional Neural Network: CNN

Deep learning is a scheme of machine learning that the models recognize images, sounds, and other objects explicitly. The CNN has been the most common deep learning algorithm. The CNN knows from the information automatically, classifies behaviors based on trends, and removes the need for hand-operated feature extraction process. The CNN is made up of many convolutional layers and a pooling layer (known as a subsampling layer) in general. At the top, there are one or more fully connected layers. A completely attached layer is a multilayer perceptron with weights w_{ij} connecting all input units i to all output units j. There are no criteria to learn in the subsampling layer, also known as the pooling layer. In other words, the pooling layer contains no weights or intercept units. Weights and intercepts exist in convolutional or completely connected layers. Fig. 3 depicts an arrangement of the one-dimensional CNN.

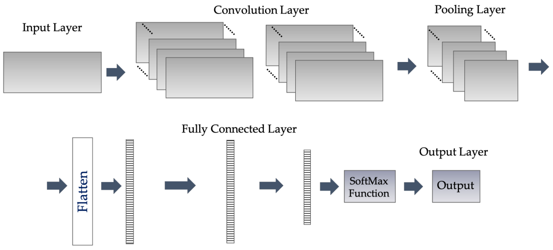

Fig. 3. 1D-CNN architecture.

III. The Proposed Framework

This paper proposes the HAR architecture, which is depended on a hybrid deep learning model (called CNN-LSTM network). The proposed DL model combines time-series data from smartwatch sensors with additional location data to characterize complex human behavior.

A. DHA Dataset

School of Electrical Engineering, Kookmin University, Republic of Korea, provided the DHA dataset. The data was collected from two willing participants who wore a smartwatch on their dominant hand for four weeks while participating in 11 protocol activities. It was expected that the user does not perform several tasks at the same time. The smartwatch used for the capture is the Apple Watch Series 2. The activity data was gathered with a sampling rate of 10Hz using the smartwatch's tri-axial accelerometer. Each operation was divided into three sections and completed in three separate locations. In workplaces, five tasks were considered: office work, reading, writing, sleeping, and playing a video game. Writing an email, writing a letter, and coding are all examples of office work. In kitchens, three things were considered: dining, cooking, and dish washing. The remaining three events were deemed suitable for outdoor participation. Table I summarizes the number of samples collected in each activity.

TABLE I
ACTIVITY DETAILS OF DHA DATASET

Activity	Abbreviates	Number of Raw Accelerometer Data
Office working	Ow	62,711
Reading	Re	36,976
Writing	Wr	27,677
Taking a rest	Tr	31,265
Playing a game	Pg	51,906
Eating	Ea	46,155
Cooking	Co	10,563
Dish washing	Wd	10,712
Walking	Wa	25,768
Running	Ru	6,452
Taking a transport	Tt	28,483

Fig. 4 shows graphical plots of accelerometer data from some samples of activities in the DHA dataset.

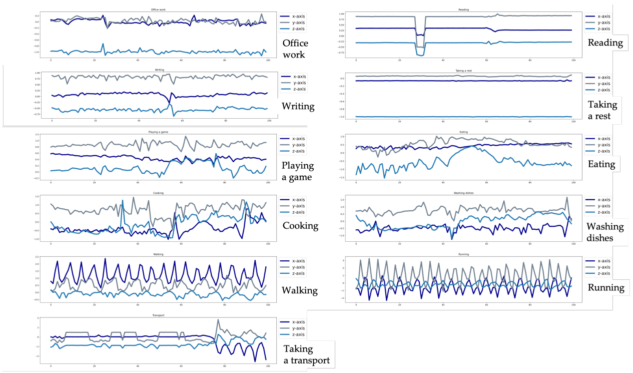

Fig. 4. Graphical plot of the tri-axial accelerometer data for the 11 activities from the DHA dataset.

B. Multiscale CNN-LSTM Architectures

Multiple parallel channels are planned for the multiscale CNN-LSTM to conduct convolutions on different axes separately. A CNN model with different kernel sizes warps in time distributed layers extracted feature maps from input data. After that, the extracted features are flattened and applied to each network's LSTM layers. At the completely connected layer as shown in Fig. 5, the output from each direction is merged, and the labels are expected.

IV. Experimental Results

A. Conducted Experiments

In this work, research experiments are conducted on Google Colaboratory platform with Tesla V100. The libraries Python v.3.6.9, TensorFlow v.2.2.0, Keras v.2.3.1, Scikit-Learn, Numpy v.1.18.5, and Pandas v.1.0.5 are used to implement the Python programming language.

B. Experimental Results

The DHA dataset is used to build deep learning models and evaluated by the 10-fold cross validation method. In the first experiment, two baseline DL models, CNN and LSTM, are used to demonstrate recognition efficiency, as reported in Table II. The results shows that CNN model achieves higher accuracy that LSTM model. Moreover, other evaluation metrics values – precision, recall and F1-score – are all higher than the results from the LSTM model with all values of window sizes.

With the second experiment as illustrated in Table III, the proposed CNN-LSTM is trained with different windows sizes at 5, 10, 20, 30 and 40 seconds. The results shows that the proposed hybrid model can achieved higher accuracy than both of CNN model and LSTM model with all of window sizes. Furthermore, the proposed model achieved highest accuracy of 94.18% at the window size of 10 seconds.

V. Conclusion and Future Works

A system for human activity recognition that uses a multiscale CNN-LSTM network to solve the HAR issue with high performance is introduced in this research work. We looked at three different forms of deep learning networks to see how well they recognized tri-axial accelerometer sensors. We compared the predictive accuracy of these deep learning methods to a dataset that is opened to the public called the DHA dataset, as well as precision, recall, and F1-score are some of the other performance indicators. So, the proposed multiscale CNN-LSTM networks is surpass than the other baseline deep learning networks with a high accuracy of 94.18%, according to the experimental results.

In the future, we expect to further develop multiscale CNN-LSTM models and test them using various hyperparameters such as learning rate, batch size, regularization, and others. Apply this model to more specific activities in order to address other deep learning and HAR problems by testing it on other publicly available activity datasets is also our future planning.

Acknowledgment

This research was funded by UoE for Information Technology of Advanced Data Analysis, University of Phayao with Grant no. FF64-UoE008.

References

[1] C. Jobanputra, J. Bavishi, and N. Doshi, "Human activity recognition: A survey," *Procedia Computer Science*, vol. 155, pp. 698 – 703, 2019.

[2] S. Ranasinghe, F. A. Machot, and H. C. Mayr, "A review on applications of activity recognition systems with regard to performance and evaluation," *International Journal of Distributed Sensor Networks*, vol. 12, no. 8, pp. 1–22, 2016.

[3] Y. A. Shichkina, G. V. Kataeva, Y. A. Irishina, and E. S. Stanevich, "The use of mobile phones to monitor the status of patients with parkinson's disease," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, vol. 11, no. 2, pp. 55–73, June 2020.

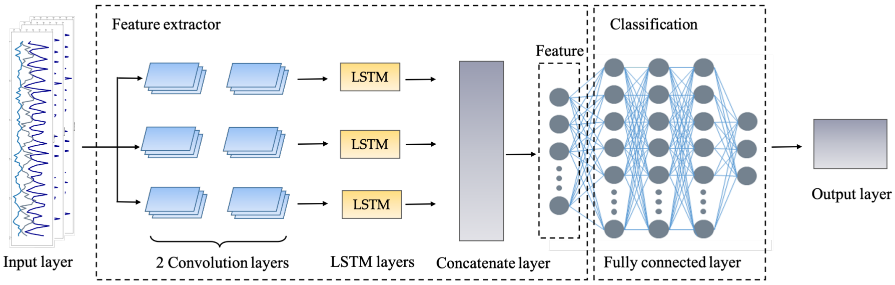

Fig. 5. Architecture of the multiscale CNN-LSTM.

TABLE II
PERFORMANCE METRICS OF BASED DL MODELS

Window sizes(s)	Model	Evaluation metrics (%mean +/-std.)			
		Accuracy	Precision	Recall	F1-score
5	CNN	89.21% (+/-1.29%)	87.25% (+/-3.68%)	88.18% (+/-2.16%)	87.65% (+/-1.84%)
	LSTM	88.78% (+/-0.69%)	86.43% (+/-3.16%)	86.89% (+/-2.43%)	86.59% (+/-1.45%)
10	CNN	87.93% (+/-1.30%)	86.60% (+/-2.68%)	85.41% (+/-5.32%)	85.85% (+/-2.31%)
	LSTM	87.40% (+/-1.63%)	87.98% (+/-4.74%)	85.63% (+/-3.80%)	86.68% (+/-2.90%)
20	CNN	86.37% (+/-1.91%)	86.29% (+/-3.59%)	89.49% (+/-5.32%)	87.77% (+/-3.53%)
	LSTM	79.23% (+/-6.94%)	77.50% (+/-11.15%)	75.13% (+/-15.16%)	75.86% (+/-12.40%)
30	CNN	86.48% (+/-1.62%)	86.29% (+/-5.37%)	89.17% (+/-7.23%)	87.58% (+/-5.60%)
	LSTM	70.39% (+/-8.56%)	76.31% (+/-10.06%)	65.25% (+/-11.11%)	69.79% (+/-9.07%)
40	CNN	85.75% (+/-1.49%)	83.46% (+/-5.28%)	91.61% (+/-5.43%)	87.24% (+/-4.48%)
	LSTM	59.36% (+/-7.11%)	66.09% (+/-16.52%)	49.27% (+/-19.95%)	54.36% (+/-19.67%)

TABLE III
PERFORMANCE METRICS OF THE MULTICHANNEL CNN-LSTM MODELS USING ACCELEROMETER DATA WITH LOCATION DATA

Window sizes(s)	Evaluation metrics (%mean +/-std.)			
	Accuracy	Precision	Recall	F1-score
5	93.63% (+/-0.634%)	92.24% (+/-3.058%)	91.27% (+/-3.088%)	91.69% (+/-2.095%)
10	94.18% (+/-0.697%)	91.21% (+/-3.787%)	93.15% (+/-2.574%)	92.13% (+/-2.681%)
20	92.52% (+/-2.410%)	92.93% (+/-3.078%)	92.88% (+/-4.636%)	92.86% (+/-3.363%)
30	92.78% (+/-1.269%)	94.03% (+/-2.874%)	95.10% (+/-5.092%)	94.43% (+/-2.227%)
40	92.26% (+/-2.117%)	93.01% (+/-5.205%)	94.78% (+/-5.711%)	93.71% (+/-3.953%)

[4] O. T. Ibrahim, W. Gomaa, and M. Youssef, “Zero-calibration device-free localization for the iot based on participatory sensing,” in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–7.

[5] S. Mekruksavanich, A. Jitpattanakul, P. Youplao, and P. P. Yupapin, “Enhanced hand-oriented activity recognition based on smartwatch sensor data using lstms,” *Symmetry*, vol. 12, no. 9, p. 1570, 2020.

[6] O. T. Ibrahim, W. Gomaa, and M. Youssef, “Crosscount: A deep learning system for device-free human counting using wifi,” *IEEE Sensors Journal*, vol. 19, no. 21, pp. 9921–9928, 2019.

[7] S. Mekruksavanich and A. Jitpattanakul, “Convolutional neural network and data augmentation for behavioral-based biometric user identification,” in *ICT Systems and Sustainability*, M. Tuba, S. Akashe, and A. Joshi, Eds. Singapore: Springer Singapore, 2021, pp. 753–761.

[8] S. J. M. Bamberg, A. Y. Benbasat, D. M. Scarborough, D. E. Krebs, and J. A. Paradiso, “Gait analysis using a shoe-integrated wireless sensor system,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 4, pp. 413–423, 2008.

[9] S. Mekruksavanich and A. Jitpattanakul, “Lstm networks using smartphone data for sensor-based human activity recognition in smart homes,” *Sensors*, vol. 21, no. 5, 2021.

[10] P. Casale, O. Pujol, and P. Radeva, “Human activity recognition from accelerometer data using a wearable device,” in *Pattern Recognition and Image Analysis*, J. Vitrià, J. M. Sanches, and M. Hernández, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 289–296.

[11] M. Ullah, H. Ullah, S. D. Khan, and F. A. Cheikh, “Stacked lstm network for human activity recognition using smartphone data,” in *2019 8th European Workshop on Visual Information Processing (EUVIP)*, 2019, pp. 175–180.

[12] S. Mekruksavanich and A. Jitpattanakul, “Smartwatch-based human activity recognition using hybrid lstm network,” in *2020 IEEE SENSORS*, 2020, pp. 1–4.

[13] R. Mutegeki and D. S. Han, “A cnn-lstm approach to human activity recognition,” in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020, pp. 362–366.

[14] S. Mekruksavanich and A. Jitpattanakul, “Biometric user identification based on human activity recognition using wearable sensors: An experiment using deep learning models,” *Electronics*, vol. 10, no. 3, 2021.

[15] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. Reyes-Ortiz, “Energy efficient smartphone-based activity recognition using fixed-point arithmetic,” *Journal of Universal Computer Science*, vol. 19, pp. 1295–1314, 01 2013.

[16] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” *SIGKDD Explor. Newsl.*, vol. 12, no. 2, p. 74–82, Mar. 2011.

[17] L. Hu, Y. Chen, S. Wang, J. Wang, J. Shen, X. Jiang, and Z. Shen, “Less annotation on personalized activity recognition using context data,” in *2016 Intl IEEE Conferences on Ubiquitous Intelligence Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, 2016, pp. 327–332.