The 6th International Conference on Digital Arts, Media and Technology (DAMT) and
4th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (NCON)

2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering | 978-1-6654-1569-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/ECTIDAMTNCON51128.2021.9425769

A Multichannel CNN-LSTM Network for Daily Activity Recognition using Smartwatch Sensor Data

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*—Recognition of human behavior is recently an active and stimulating study. The HAR can provide valuable information on human movement and the behavior of everyday life activities. In the last decade, a large range of HAR-based applications have been implemented, such as healthcare tracking, biometric user authentication, and so on. Previously, several deep learning approaches have been introduced to focus on the issue of conventional machine learning approaches with handcrafted features. So, a novel deep learning architecture to solve the HAR problem is proposed in this study. The introduced architecture is a hybrid model called a multichannel CNN-LSTM network. The model is evaluated by state-of-the-art evaluation metrics; accuracy, precision, recall and F1-score, with a public dataset of smartwatch's accelerometer data called DHA dataset. The proposed multichannel CNN-LSTM outperforms other deep learning methods in terms of accuracy, with a score of 96.87%.**

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

I. Introduction

Human-centered computation is an emerging field of research and application that helps to identify human actions and to incorporate digital technology with users and the social context. This entails and subsumes human activity recognition (HAR) aimed at identifying one or more people's acts, characteristics, and objectives from a temporal sequence of observations streamed from one or more sensors. In several areas, including health monitoring [1], elderly care, indoor and outdoor sports, virtual coaching, security systems, learning and cooperating robots, human robot interaction, military, etc., automated recognition of behavioral context can be used. For a detailed survey of applications, see [2].

In general, the purpose of HAR systems is to determine the ongoing actions/activities of an individual, a group of individuals, or even a crowd based on sensory observation data, to determine certain personal characteristics such as the identity of individuals in a given room, gender, age, etc., and to know the context in which the activities observed take place. Based on the modality of sensory information used, HAR systems can be categorized, as the type of sensory data greatly affects the kinds of features, algorithms, architectures, and analytical techniques used [3]. The following research streams and advances in HAR systems can generally be identified: HAR systems based on visual data, HAR systems based on motion inertial sensors such as Inertial Measurement Units, and HAR systems based on the signal intensity obtained from commodity routers mounted in the surrounding setting [4]–[6]. The second modality, namely inertial motion time series data, is the subject of the current analysis.

In wearable-based applications, measurements are taken from mobile sensors connected to human body components such as hands, knees, waist, and chest [7], [8]. Many devices can be used to monitor selected features of human body motion, including accelerometers, gyroscopes, compasses, and GPSs [9]–[12]. Temperature, light and humidity sensors, barometers, magnetometers, and handheld cameras are all examples of phenomena that can be used to evaluate phenomena around the user. These sensor types are basically based on embedded MEMS sensors including the accelerometer and gyroscope contained an inertial measurement unit (IMU). Wearable devices are capable of measuring user data anywhere, while standing, sleeping, or even working anywhere. Unlike fixed-sensor based devices, since they are not constrained by a particular location where the sensors are mounted [13], [14]. It's also very simple to concentrate on directly measuring data from individual body parts without a lot of preprocessing, such as in fixed depth cameras. Smart watches and shoes, sensor gloves, hand bands, and garments are wearable examples [15].

In this research, a new hybrid deep learning model called multichannel CNN-LSTM is proposed, formed on the CNN and LSTM models. In order to find the right window sizes for HAR problem, this work has conducted many experiments.

The structure of this work is arranged as follows. A principle of the associated recognition of human behavior is given in Section II. Our design is outlined in Section III. The experiments performed on the DHA dataset are presented in Section IV. Section V summarized the derived results and contributes potential recommendations for the future.

II. Background Concept

A. HAR Design

HAR helps to describe human behaviors that enable computing systems to assist users proactively based on their requirements. Human activities, for example walking, standing,

working, and laying down, can be defined in a given protocol as a series of actions carried out by the user over a time. The design of a HAR system in machine learning is carried out in five basic steps: data collection, data segmentation, extraction of features, model training, and classification, as shown in Fig. 1.

B. Convolutional Neural Network

The convolutional neural network (CNN) is the most commonly used deep learning algorithm, a form of machine learning in which images, video, text, or sound are directly categorized by models. CNN learns directly from the data, uses patterns to identify events, and removes the need to extract features manually. In general, a CNN consists of multiple layers of convolution and a layer of pooling, also known as a layer of subsampling. One or more completely related layers follow at the top. A fully connected layer is a multilayer perceptron in which all the i input units are connected with w_{ij} weights to all j output units. The one-dimensional CNN structure is shown in Fig. 2.

C. Long Short Term Memory

Long Short Term Memory (LSTM) networks are a form of recurrent neural network that can learn order dependence in series prediction and classification. This is a behavior that is expected in a variety of complex problem domains, including machine translation, speech recognition, and others. LSTM layers usually consist of recurrently linked memory blocks in a memory unit, as shown in Fig. 3. These LSTM cells consist of gates to specify when to forget the memory cell's previous secret states and further update the cells, allowing temporal information to be used by the network [16], [17].

III. The Detailed Framework

The introduced multichannel CNN-LSTM HAR platform allows sensor data collected from smartphone sensors to determine the operation performed by the user of wearable devices.

A. DHA Dataset

We use the daily human activity (DHA) dataset in this paper, which is provided publicly by the School of Electrical Engineering, Kookmin University, Republic of Korea. The dataset was collected from two volunteer participants wearing a smartwatch on their dominant hand during four weeks of 11 protocol activities. Their activity data were captured at a sampling rate of 10Hz by an accelerometer built-in smartwatch. Each operation was situated in the office, kitchen, and outdoors in three separate locations. The sample numbers for each operation are summarized in Table I. The accelerometer 2D-scatter data of activities samples in DHA dataset is illustrated in Fig. 4.

TABLE I
A list of activities in DHA dataset.

Activity	Abbreviates	No. of Raw Accelerometer Data
Office work	Ow	62,411
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
Washing dishes	Wd	10,712
Walking	Wa	25,768
Running	Ru	6,452
Taking a transport	Tt	28,483

B. Mutichannel CNN-LSTM Architectures

For multichannel CNN-LSTM, multiple parallel channels were configured to separately conduct convolutions on source data from different axes. Feature maps were derived by the CNN model, warped in the time distributed layers, from the input data. Then the feature maps were then flatted and projected in each channel to the LSTM layers. Each channel's performance was combined and the labels were predicted at the completely linked layer as shown in Fig. 5.

IV. Experimental Results

In this section, the environment of the experiment and the experimental results used to assess the multichannel CNN-LSTM networks for recognition of human behavior is described.

A. Experiments

In this study, every experiment is run on the Tesla V100 Google Colab platform. Python-3.6.9, TensorFlow-2.2.0, Keras-2.3.1, Scikit-learn, Numpy-1.18.5, and Pandas-1.0.5 libraries are software used in this work. The Python programming language is used for implementation.

B. Results

To develop deep learning models, the DHA dataset was used and assessed by a 10-fold cross validation process. As described in Table II, the first experiment showed the recognition performance of two baseline DL models: CNN and LSTM.

TABLE II
Performance metrics of baseline DL models.

DL Network	Evaluation metrics (%mean)			
	Accuracy	Precision	Recall	F1-score
CNN	91.022	89.166	91.021	90.027
LSTM	84.347	84.246	78.232	80.937

In the second experiment, the multichannel CNN-LSTM was trained at 3, 5, 10, 30 and 60 seconds with various window

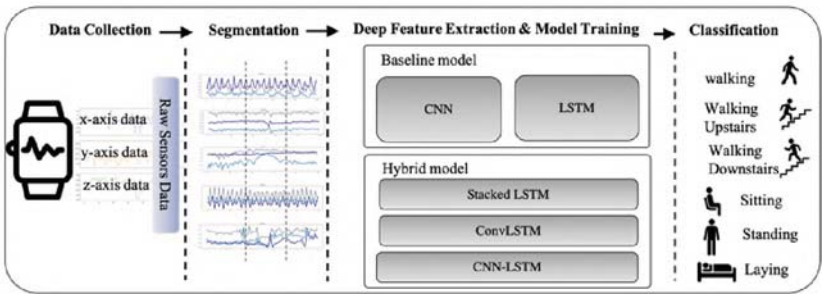

Fig. 1. A common ML approach of sensor-based HAR.

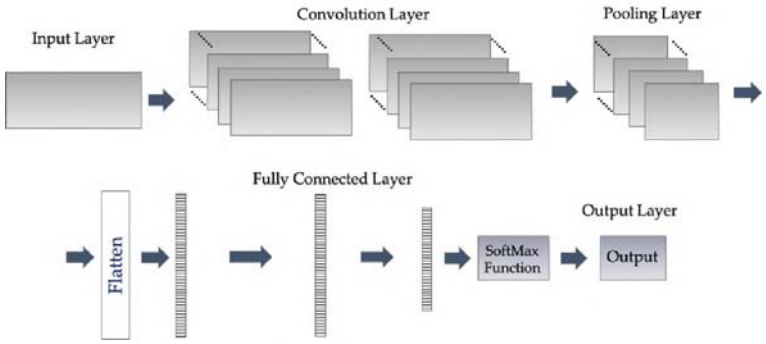

Fig. 2. 1D-CNN architecture.

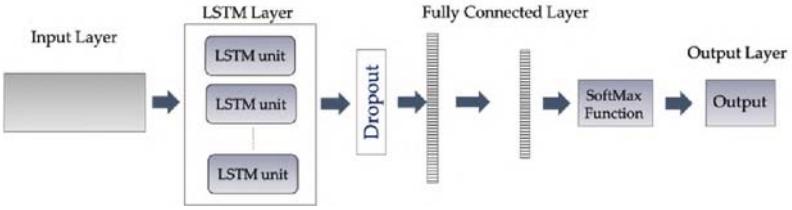

Fig. 3. LSTM architecture.

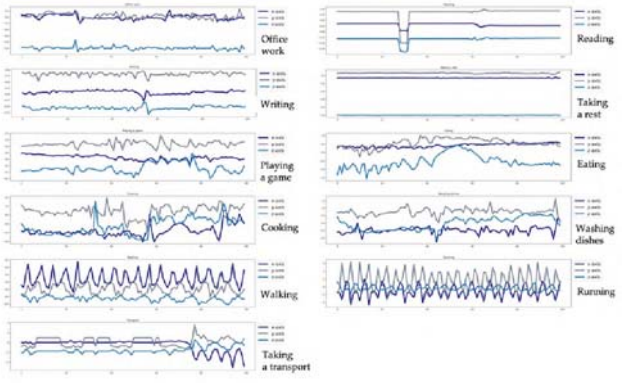

Fig. 4. 2D-scatter plot of samples from DHA dataset.

sizes as illustrated in Table III. The derived results demonstrated that at a window size of 10 seconds, the proposed model achieved the highest precision of 94.55%.

In addition, we included position data in the third experiment to train the multichannel CNN-LSTM model as presented in Table IV. The results showed that with a window size of 10 seconds, the added position data could boost the recognition efficiency with a maximum accuracy of 96.87 percent.

TABLE III
PERFORMANCE METRICS OF THE MULTICHANNEL CNN-LSTM MODELS USING ACCELEROMETER DATA WITHOUT LOCATION DATA

Window sizes	Evaluation metrics (%mean)			
	Accuracy	Precision	Recall	F1-score
3	93.209	92.119	92.306	92.199
5	94.034	91.720	94.099	92.868
10	94.551	93.261	93.756	93.485
30	92.995	95.214	95.709	95.412
60	90.323	93.233	93.183	92.986

V. CONCLUSIONS AND FUTURE WORK

In this study, the recognition system that seeks to work with high output for the HAR issue of the multichannel

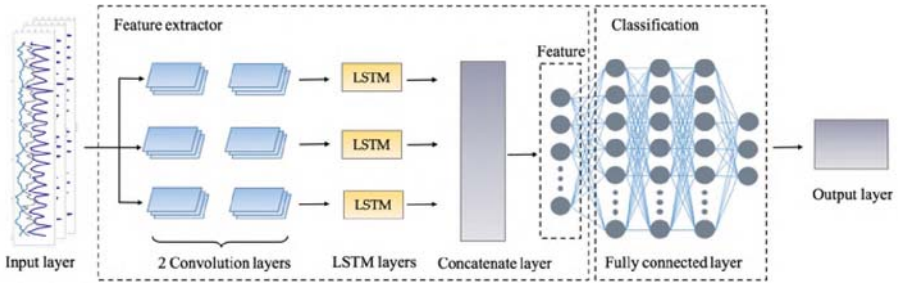

Fig. 5. Architecture of Multichannel CNN-LSTM.

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

Window sizes(s)	Evaluation metrics (%mean)			
	Accuracy	Precision	Recall	F1-score
3	95.880	94.957	94.028	94.473
5	96.123	94.252	95.021	94.607
10	96.869	96.960	94.361	95.625
30	96.099	95.683	96.388	95.946
60	94.054	95.689	96.444	95.894

CNN-LSTM model is presented. To research their recognition efficiency with tri-axial accelerometer sensors, we selected three types of deep learning networks. These deep learning networks are evaluated by a publicly available dataset called DHA with predictive accuracy and other indicators of results. The experimental results demonstrate that the multichannel CNN-LSTM networks developed in this study outperform 96.87 percent of the other baseline deep learning networks.

We will further improve the multichannel CNN-LSTM models for future work and evaluate the model with various hyperparameters, including learning rate, batch size, regularization and others, for future work. In addition, by analyzing it on other public activity datasets, we plan to apply this model to more specific activities to tackle other problems in deep learning and HAR.

ACKNOWLEDGMENT

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

REFERENCES

[1] 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.

[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] 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.

[4] 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.

[5] S. Mekruksavanich and A. Jitpattanakul, "Classification of gait pattern with wearable sensing data," in *2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON)*, 2019, pp. 137–141.

[6] 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.

[7] 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.

[8] 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.

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

[10] S. K. Wong and S. M. Yiu, "Location spoofing attack detection with pre-installed sensors in mobile devices," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, vol. 11, no. 4, pp. 16–30, December 2020.

[11] N. Hnoohom, S. Mekruksavanich, and A. Jitpattanakul, "Human activity recognition using triaxial acceleration data from smartphone and ensemble learning," in *2017 13th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, 2017, pp. 408–412.

[12] 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.

[13] S. Mekruksavanich, N. Hnoohom, and A. Jitpattanakul, "Smartwatch-based sitting detection with human activity recognition for office workers syndrome," in *2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON)*, 2018, pp. 160–164.

[14] S. Mekruksavanich and A. Jitpattanakul, "Exercise activity recognition with surface electromyography sensor using machine learning approach," in *2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT NCON)*, 2020, pp. 75–78.

[15] 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.

[16] S. Ashry, R. Elbasiony, and W. Gomaa, "An lstm-based descriptor for human activities recognition using imu sensors," in *Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics - Volume 1: ICINCO,*, INSTICC. SciTePress, 2018, pp. 494–501.

[17] W. Zhu, C. Lan, J. Xing, W. Zeng, Y. Li, L. Shen, and X. Xie, "Co-occurrence feature learning for skeleton based action recognition using regularized deep lstm networks," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, ser. AAAI'16. AAAI Press, 2016, p. 3697–3703.