Smartwatch-based Human Activity Recognition Using Hybrid LSTM 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*—As a result of the rapid development of wearable sensor technology, the use of smartwatch sensors for human activity recognition (HAR) has recently become a popular area of research. Currently, a large number of mobile applications, such as healthcare monitoring, sport performance tracking, etc., are applying the results of major HAR research studies. In this paper, an HAR framework that employs spatial-temporal features that are automatically extracted from data obtained from smartwatch sensors is proposed. The hybrid deep learning approach is used in the framework through the employment of Long Short-Term Memory Networks and the Convolutional Neural Network, eliminating the need for the manual extraction of features. The advantage of tuning the hyperparameters of each of the considered networks by Bayesian optimization is also utilized. It was indicated by the results that the baseline models are outperformed by the proposed hybrid deep learning model, which has an average accuracy of 96.2% and an F-measure of 96.3%.**

***Index Terms*—smartwatch, deep learning, human activity recognition, wearable devices, hybrid LSTM**

I. Introduction

There are numerous reasons that the rapid development of wearable sensor technology has occurred, for example the decreasing cost of sensor devices and the significant improvement of miniaturized sensors' computational capacity [1]. Wearable sensors are defined as tiny devices that people are able to continuously carry during the performance of their everyday activities. Sensors such as accelerometers, barometers, global positioning systems (GPS), gyroscopes, and magnetometers are capable of capturing the signal of the physical movement of a person at any time and at any place. The advantages of wearable sensors have been adopted in a number of beneficial mobile applications, including abnormal driving detection [2], healthcare systems for remotely monitoring elderly people [3]–[5], sport performance tracking [6], and mobile assistance systems for people who are visually impaired [7].

Interest in the topic of applying wearable sensors to human activity recognition (HAR) is currently increasing for many researchers in the pervasive field of computing [8]–[12]. These days, direction and motion sensors, for example accelerometers and gyroscopes that can be employed for the classification of human activities, are installed on smartphones and smartwatches [13]. However, the focus of these efforts has mostly been on smartphone-based activity recognition, including those investigated in [14]–[16]. The application of these common commercially available devices significantly increased the possible uses for activity recognition; nevertheless, limits have resulted from their being placed on the body of a user and the unstable orientation, such as when the smartphone moves around while in a the user's pocket or when the position of the pocket is not suitable for the tracking of hand-based activities. Since women do not usually carry their smartphone in their pocket, they are especially prone to the limitations of smartphone-based activity recognition. The use of smartwatches, which are worn in a constant location, addresses the majority of these drawbacks, as they are ideally positioned for the tracking of activities that are hand-based.

In addition, machine learning algorithms can be employed for the developments that can improve the activity recognition model using smartwatches in order to provide more accurate assessments of a broad variety of activities [17]–[19]. Nevertheless, these approaches using conventional machine learning regularly depend upon heuristic manual feature extraction, and are thus normally limited to the knowledge of the human domain. Because of this limitation, there are restrictions regarding the performance of approaches using conventional machine learning in terms of the accuracy of the classification and other metrics used for evaluation [20]. In this paper, approaches utilizing deep learning (DL) are applied in order to overcome these limitations.

An HAR framework that automatically uses spatial-temporal features extracted from the data obtained by smartwatch sensors by applying a hybrid deep learning model known as a CNN-LSTM network is proposed in this work. In this study, an experimental evaluation is performed in order to conduct a comparison of the proposed CNN-LSTM-based approach with the baseline deep learning models found in a public dataset referred to as the WISDM dataset [21]. In addition, Bayesian optimization is employed for the tuning of the CNN-LSTM models' hyperparameters.

978-1-7281-6801-2/20/$31.00 ©2020 IEEE

II. Background and Related Research

In this section, the background knowledge for this study and the related research concerned with HAR and deep learning approaches are summarized.

A. Deep Learning applied to HAR

The smartwatch-based HAR focus on human behavior understanding by using sensors of smartwatch or smart phone such as gyroscope, accelerometer, etc. The understanding is processed with sensors while they are performing their activities through recognition. The HAR problem can be systematically mentioned as a a time-series classification [22].

In recent studies, the main focus of research conducted on HAR has been deep learning (DL). Numerous successful outcomes have been reported, which has led researchers to apply a range of approaches that utilize deep learning in order to solve complex problems involving HAR [23]–[26]. Reduction of the effort needed for feature extraction using conventional machine learning is provided by these deep learning approaches. As a result of this advantage, raw sensor data can be used to build DL models, which provides a high level of efficiency when performing recognition.

One type of deep learning networks is known as convolutional neural networks (CNN), which are recommended for the improvement of the performance in solving problems in wearable-based HAR [27]. The CNNs are a type of DL network that can extract the spatial features from raw sensor data automatically. However, there are a number of human activities that require time-series data, which results in the temporal dependencies. In order to solve this temporal dependency issue, the use of Long Short-Term Memory networks (LSTM) has been proposed, and application of LSTMs is currently increasing in the field of HAR [28], [29]. Hybrid LSTMs that provide the advantages of both CNNs and LSTMs by combining several preceding CNN layers that extract the spatial features with LSTM layers that extract the temporal features have been proposed. In the experiments carried out in this study, the implementation of the combined CNN-LSTM was investigated, as described in Section III.

III. Proposed Methods

In the proposed framework for the human activity recognition that is smartwatch-based, the capture of the sensor data from the smartwatch sensor is enabled in order to conduct the classification of the activities that are performed by the smartwatch user. The overall methods applied in this study to achieve the aim of this research are illustrated in Fig. 1.

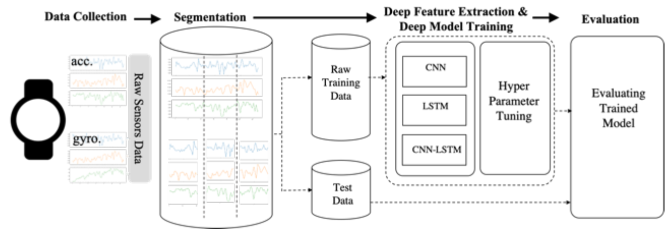

Fig. 1. The proposed framework of smartwatch-based HAR

A. Smartwatch Dataset

The WISDM from the UCI Repository is the publicly available source of the raw sensor data used in this study [21]. Accelerometer and gyroscope data from various smartphones operating with Android 6.0 and a smartwatch (LG G Watch) operating with Android Wear 1.5 are contained in the dataset. The sensor data were collected from the smartwatches worn on the dominant hand of 51 participants during 18 physical activities performed in everyday life. Each of the activities was performed separately for approximately 3 minutes at a rate of 20 Hz.

B. Preprocessing and Segmentation of the Data

In this work, only the smartwatch data included in the WISDM dataset was utilized. An exploratory data analysis of the sensor data was conducted, and it was found that the activity data from seven of the participants did not include all of the pre-determined activities. Therefore, it was necessary that the data from those seven subjects (subjects 1616, 1618, 1637, 1638, 1639, 1640 and 1642) be disregarded. As a result, the smartwatch data was collected from only 44 subjects. For the processing of the time-series data in the HAR problem, a 10-second sliding window with 50% of the overlapping proportion was used to segment the data.

C. Hybrid Long Short-Term Memory Network

In this research, a hybrid LSTM network known as a 2-layer CNN-LSTM is proposed for the improvement of recognition performance. The CNN-LSTM comprises two convolutional layers and a single LSTM layer. The input sensor data's size is 200×6, and the output layer's size is 18×1. Fig. 2 illustrates the CNN-LSTM's structure.

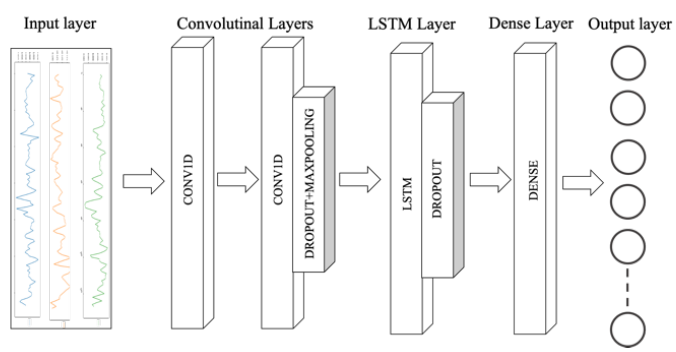

Fig. 2. The structure of the proposed CNN-LSTM network

To evaluate the proposed hybrid LSTM networks, evaluation metrics are used from the field of HAR: accuracy, precision, recall, and F-measure [30].

TABLE I

PERFORMANCE METRICS OF THE DEEP LEARNING MODELS

Scenario	Performance metrics			
	Accuracy	**Precision**	**Recall**	**F-measure**
CNN	93.1%	93.1%	93.1%	93.1%
LSTM	89.6%	89.7%	89.6%	89.6%
Proposed CNN-LSTM	96.2%	96.3%	96.2%	96.3%

TABLE II

SUMMARY HYPERPARAMETERS OF CNN-LSTM NETWORKS IN THE RESEARCH

Stage	Hyperparameters		Values
Architecture	Convolution_1	Kernel Size	3
		Stride	1
		Filters	113
	Convolution_2	Kernel Size	3
		Stride	1
		Filters	123
	Dropout_1		0.06480703
	Maxpooling		2
	LSTM_neuron		128
	Dropout_2		0.21129224
	Dense		458
Training	Optimizer		Adam
	Batch Size		64
	Learning Rate		0.00049271
	Number of Epoches		50

IV. EXPERIMENTS AND RESULTS

In this section, the experimental setting and the results that were employed for the evaluation of the proposed CNN-LSTM networks for the smartwatch-based HAR are described.

A. Experiments

For the comparison of the performance of the proposed CNN-LSTM networks with that of the other DL models, variations were used in the experiments. In the first experiment, a basic CNN network composed of only one convolutional layer working with Dropout and a single Dense layer was used. In the second experiment, a basic LSTM network referred to as Vanilla LSTM composed of LSTM layers working with Dropout and one Dense Layer was used. The third experiment included a proposed CNN-LSTM that is a LSTM network in which a convolution layer is combined with a LSTM layer. The code for the experiments is in Python 3.6.9, using the TensorFlow [31], Keras [32], Sci-kit Learn [33], Numpy [34], and Pandas [35] libraries. The experiments were executed on the Google Colab platform with Tesla K80, and also the hyperparameters of each model were optimized by SigOpt [36].

B. Experimental Results

To evaluate the performance of the smartwatch-based HAR, three experiments on different variations were performed using the WISDM dataset as described in the Section III. The smartwatch sensor data from the WISDM dataset was divided into 70% to be used as training data and the remaining 30% for testing data, which resulted in 41,440 and 17,761 data numbers, respectively. Experiments were conducted in this work to evaluate the recognition performance of the DL networks with a variation of metrics, including accuracy, precision, recall, and F-measure. Table I shows the accuracy and the other metrics obtained from the various DL networks trained on the WISDM dataset.

As seen in Table I, the proposed CNN-LSTM was tuned by Bayesian optimization in order to find a set of hyperparameters that provide the high-performance metrics. The hyperparameters are summarized in Table II.

It can be seen that the proposed CNN-LSTM networks outperform all of the other networks with an accuracy of 96.2% and an F-measure of 96.3%. Therefore, the performance of the CNN-LSTM is better than that of the baseline DL models. The confusion matrix for the CNN-LSTM networks is shown in Fig. 3.

V. CONCLUSION AND FUTURE WORKS

In this work, we proposed a CNN-LSTM network to tackle the smartphone-based HAR problem. This hybrid LSTM takes advantage of both the spatial feature extraction of CNN and the temporal feature extraction of LSTM. We also used Bayesian optimization to find the optimized hyperparameters of the model for achieving better performance of HAR. We have evaluated the performance of this CNN-LSTM network by various metrics and a public dataset known as WISDM. The results show that the proposed hybrid LSTM outperforms the other baseline networks by utilizing the automatic spatial-temporal feature extraction from the raw sensor data with an average accuracy of 96.2% and an F-measure of 96.3%. For future research work, we shall develop this model for personalized human activity recognition using a transfer learning approach based on smartwatch sensors.

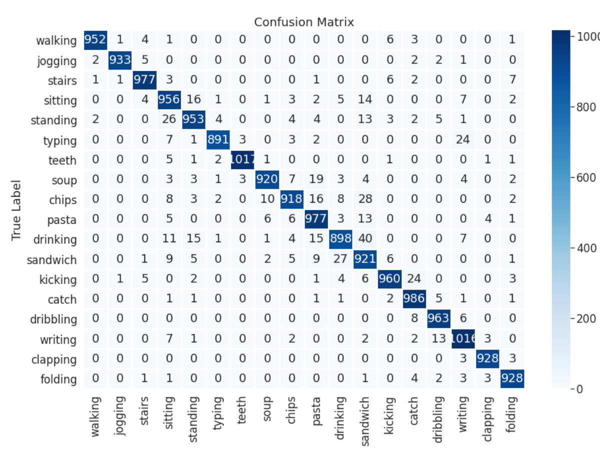

Fig. 3. Confusion Matrix for the proposed CNN-LSTM

ACKNOWLEDGMENT

The authors thank the SigOpt team for the provided optimization services.

REFERENCES

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

[2] L. Liu, C. Karatas, H. Li, S. Tan, M. Gruteser, J. Yang, Y. Chen, and R. P. Martin, "Toward detection of unsafe driving with wearables," in *Proceedings of the 2015 Workshop on Wearable Systems and Applications*, ser. WearSys '15. New York, NY, USA: Association for Computing Machinery, 2015, pp. 27 –32.

[3] 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.

[4] L. M. Schrader, A. V. Toro, S. G. A. Konietzny, S. Rüping, B. Schäpers, M. Steinböck, C. Krewer, F. Mueller, J. Güttler, and T. Bock, "Advanced sensing and human activity recognition in early intervention and rehabilitation of elderly people," *Journal of Population Ageing*, vol. 13, pp. 139–165, 2020.

[5] S. Mekruksavanich, "Medical expert system based ontology for diabetes disease diagnosis," in *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 2016, pp. 383–389.

[6] 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.

[7] L. Porzi, S. Messelodi, C. M. Modena, and E. Ricci, "A smart watch-based gesture recognition system for assisting people with visual impairments," in *Proceedings of the 3rd ACM International Workshop on Interactive Multimedia on Mobile and Portable Devices*, ser. IMMPD '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 19–24.

[8] O. C. Ann and L. B. Theng, "Human activity recognition: A review," in *2014 IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2014)*, 2014, pp. 389–393.

[9] S. Ramasamy Ramamurthy and N. Roy, "Recent trends in machine learning for human activity recognition-a survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, p. e1254, 03 2018.

[10] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical human activity recognition using wearable sensors," *Sensors*, vol. 15, pp. 31 314–31 338, 12 2015.

[11] M. Mario, "Human activity recognition based on single sensor square hv acceleration images and convolutional neural networks," *IEEE Sensors Journal*, vol. 19, no. 4, pp. 1487–1498, 2019.

[12] Y. Zhang, Z. Zhang, Y. Zhang, J. Bao, Y. Zhang, and H. Deng, "Human activity recognition based on motion sensor using u-net," *IEEE Access*, vol. 7, pp. 75 213–75 226, 2019.

[13] U. Alrazzak and B. Alhalabi, "A survey on human activity recognition using accelerometer sensor," in *2019 Joint 8th International Conference on Informatics, Electronics Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision Pattern Recognition (icIVPR)*, 2019, pp. 152–159.

[14] N. Zouba, F. Bremond, and M. Thonnat, "An activity monitoring system for real elderly at home: Validation study," in *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2010, pp. 278–285.

[15] 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.

[16] N. Ahmed, Rafiq, and Islam, "Enhanced human activity recognition based on smartphone sensor data using hybrid feature selection model," *Sensors*, vol. 20, p. 317, 01 2020.

[17] 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.

[18] J. Yang, K. Cheng, J. Chen, B. Zhou, and Q. Li, "Smartphones based online activity recognition for indoor localization using deep convolutional neural network," in *2018 Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS)*, 2018, pp. 1–7.

[19] H. N. Monday, J. Ping Li, G. U. Nneji, M. Folarin Raji, C. C. Ukwuoma, J. Khan, C. J. Ejiyi, A. Ulhaq, S. Nahar, H. S. Abubakar, and C. A. Ijeoma, "A survey on hand-based behavioral activity recognition," in *2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing*, 2019, pp. 119–123.

[20] R. Singh and R. Srivastava, "Some contemporary approaches for human activity recognition: A survey," in *2020 International Conference on Power Electronics IoT Applications in Renewable Energy and its Control (PARC)*, 2020, pp. 544–548.

[21] G. M. Weiss, K. Yoneda, and T. Hayajneh, "Smartphone and smartwatch-based biometrics using activities of daily living," *IEEE Access*, vol. 7, pp. 133 190–133 202, 2019.

[22] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3 – 11, 2019.

[23] A. Baldominos, A. Cervantes, Y. Sáez, and P. Isasi, "A comparison of machine learning and deep learning techniques for activity recognition using mobile devices," *Sensors*, vol. 19, p. 521, 01 2019.

[24] P. P. San, P. Kakar, X.-L. Li, S. Krishnaswamy, J.-B. Yang, and M. N. Nguyen, "Chapter 9 - deep learning for human activity recognition," in *Big Data Analytics for Sensor-Network Collected Intelligence*, ser. Intelligent Data-Centric Systems, H.-H. Hsu, C.-Y. Chang, and C.-H. Hsu, Eds. Academic Press, 2017, pp. 186 – 204.

[25] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep learning models for real-time human activity recognition with smartphones," *Mobile Networks and Applications*, vol. 25, 12 2019.

[26] A. Murad and J.-Y. Pyun, "Deep recurrent neural networks for human activity recognition," *Sensors*, vol. 17, p. 2556, 11 2017.

[27] Song-Mi Lee, Sang Min Yoon, and Heeryon Cho, "Human activity recognition from accelerometer data using convolutional neural network," in *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2017, pp. 131–134.

[28] N. Tüfek and O. Özkaya, "A comparative research on human activity recognition using deep learning," in *2019 27th Signal Processing and Communications Applications Conference (SIU)*, 2019, pp. 1–4.

[29] M. Devanne, P. Papadakis, and S. M. Nguyen, "Recognition of activities of daily living via hierarchical long-short term memory networks," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2019, pp. 3318–3324.

[30] D. Powers and Ailab, "Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation," *J. Mach. Learn. Technol*, vol. 2, pp. 2229–3981, 01 2011.

[31] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 2016, pp. 265–283.

[32] F. Chollet, "keras," https://github.com/fchollet/keras, 2015.

[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, and G. Louppe, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, 01 2012.

[34] S. van der Walt, S. Colbert, and G. Varoquaux, "The numpy array: A structure for efficient numerical computation," *Computing in Science and Engineering*, vol. 13, pp. 22 – 30, 05 2011.

[35] W. Mckinney, "Data structures for statistical computing in python," *Proceedings of the 9th Python in Science Conference*, 01 2010.

[36] I. Dewancker, M. McCourt, S. Clark, P. Hayes, A. Johnson, and G. Ke, "A stratified analysis of bayesian optimization methods," *ArXiv*, vol. abs/1603.09441, 2016.