Multimodal Wearable Sensing for Sport-Related Activity Recognition Using Deep Learning Networks

Sakorn Mekruksavanich
Department of Computer Engineering, School of Information and Communication Technology, University of Phayao, Phayao, Thailand
Email: sakorn.me@up.ac.th

Anuchit Jitpattanakul
Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
Intelligent and Nonlinear Dynamic Innovations Research Center, Science and Technology Research Institute, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
Email: anuchit.j@sci.kmutnb.ac.th

***Abstract*—Wearable sensors using sensor-based Human Activity Recognition (S-HAR) are generally capable of regular simple actions (walking, sitting, or standing), but are indistinguishable from sophisticated activities, such as sports-related activities. Because these involve a more comprehensive, contextual, and fine-grained classification of complex human activities, simplex activity recognition systems are ineffective for growing real-world applications, for example remote rehabilitation observation and sport performance tracking. So, an S-HAR framework for recognizing sport-related activity utilizing multimodal wearable sensors in numerous body positions is proposed in this study. A public dataset named UCI-DSADS was used to investigate the recognition performance of five deep learning networks. According to the experimental results, the BiGRU recognition model surpasses other deep learning networks with a maximum accuracy of 99.62%.**

***Index Terms*—deep learning, multimodal wearable sensor, human activity recognition, CNN, LSTM**

I. Introduction

Wearable sensor technology is rapidly evolving due to a variety of factors, including lower sensor device costs and significant computational improvements in miniaturized sensors [1], [2]. Wearable sensors are small gadgets that users can take around with them while going about their regular activities. Motion sensors, such as accelerometers, gyroscopes, and magnetometers, could collect a human's bodily movement signals at any time and from any location [3]-[5]. Many mobile applications, including remote healthcare services for monitoring elderly people [6]-[9], abnormal driving monitoring [10], sport performance tracking [11], and the assistance mobile platform for individuals with disabilities [12], [13], make use of the benefits of all these wearable sensors.

Human-centered computing is a relatively new field of research and application that focuses on human behavior and the interaction of people and their social contexts with digital technology [14]. Human Activity Recognition (HAR), which attempted to determine the behavior, characteristics, and objectives of one or more individuals from a temporal series of data provided from one or more sensors [15], [16], is necessary and encompassed by this [17]. In sensor-based HAR, classification models were developed using standard Machine Learning (ML) algorithms such as decision trees, naive Bayes, and Support Vector Machine. Although various machine learning algorithms have demonstrated a high-performance model for HAR, these methods are restricted by the issue of manual feature extraction. Human knowledge and experience limit the precision of manually derived characteristics, resulting in low accuracy. Many researchers have subsequently proposed deep learning solutions to handle the limited concerns [18]-[21]. Deep neural networks have recently been suggested to learn features automatically without any need for handcrafted feature extraction, bypassing the limitations of human expertise and experience [22].

Most recognition techniques are currently still dealing with HAR problems for suitable performance. These findings suggest a gap in HAR research to know the unified model of DL, in terms of accuracy and computational time, to automatically extract characteristics and recognize complex human activities. Therefore, in this study, the multimodal wearable sensor-based HAR of sport-related activities is focused on. With five different deep learning models, we investigate the use of different wearable sensors (accelerometer, gyroscope, and magnetometer) to enhance performance of sport-related activity recognition. These wearable sensors were

Manuscript received October 12, 2021; revised February 10, 2022.

doi: 10.12720/jait.13.2.132-138

placed on the torso, left arm, right arm, left leg, and right leg, among other body positions. BiGRU outperforms other deep learning networks in terms of maximum accuracy, according to testing data.

The remaining part of the paper is divided into the sections following. Section II describes the proposed multimodal wearable sensor-based HAR for sport-related activity identification. Section III provides the research findings. Section IV brings the results to a conclusion.

II. Proposed Methodology

This paper proposes a multimodal wearable sensor-based human activity recognition framework that uses sensor data from wearable sensors to characterize the activity that the individuals have accomplished. The proposed methodology followed in this study to obtain our research goal is demonstrated in Fig. 1.

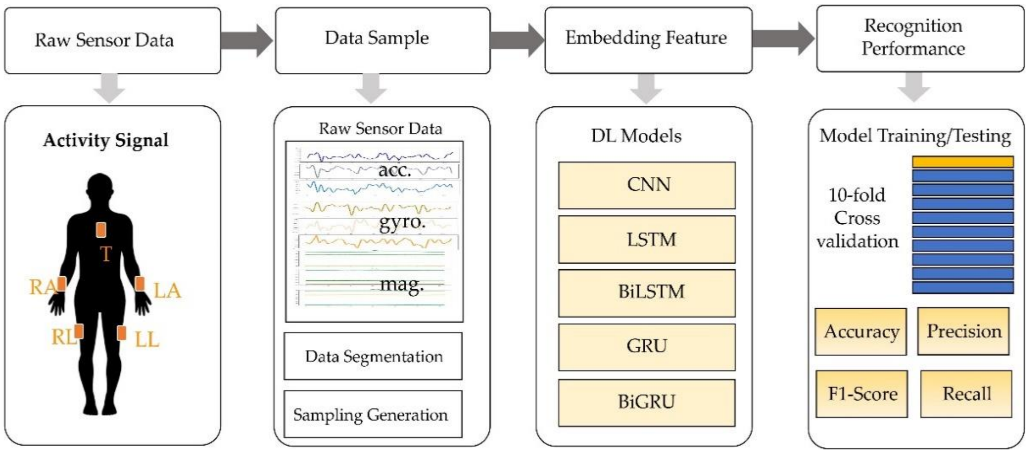

Figure 1. The proposed framework of multimodal wearable sensor-based HAR used in this study.

A. *UCI-DSADS Dataset*

In this study, we used a HAR dataset called "Daily and Sports Activities Dataset" (called UCI-DSADS dataset) [23] that is publicly available at the University of California-data Irvine's repository to investigate the proposed model. The UCI-DSADS dataset used five MTx 2-DOF orientation trackers to interpret activity sensor data from eight participants who were designed to attend 19 activities in 5 different body locations (torso, left arm, right arm, left leg, and right leg) as shown in Table I: 9 activities in everyday life and 10 sports-related activities. Some samples of the activity sensor data are shown in Appendix A.

TABLE I. List of Activities Provided in the UCI-DSADS Dataset

Activity in Daily Life	Sport-related Activity
Sitting	Walking on a treadmill with 4km/h in flat
Standing	Walking on treadmill with 15 inclined pos.
Lying on the back side	Running on a treadmill with 8 km/h
Lying on the right side	Exercising on a stepper
Ascending stairs	Exercising on a cross trainer
Descending stairs	Cycling on an exercise bile in horizontal pos.
Standing still in an elevator	Cycling on an exercise bile in vertical pos.
Moving around in an elevator	Rowing
Walking in a parking lot.	Jumping
	Playing basketball

Each of the foregoing activities is conducted for five minutes by eight volunteer participants (four females and four males, ages 20-30). The eight participants are requested to answer the activities in their own unique style, with no restrictions. Sensor devices are calibrated to collect data at a sampling rate of 25Hz. The five-minute signals are split into five-second portions, from which certain characteristics are obtained. For each activity, this produces 480 signal segments.

B. *Convolution Neural Network*

Convolutional Neural Networks (CNNs) are deep learning models that can entirely work with 2D input like images and videos [24]-[26]. CNNs can extract spatially local information and differentiate objects in the input image using some filters [27]. The convolutional layers are formed by the filter, which are usually followed by some fully-connected layer that perform the classification process. Aside from being better at learning features, CNNs can scale to massive datasets due to various pooling layers. Convolutional (Conv) layer, activation layer, pooling, Fully Connected (FC) layer, and SoftMax layer are the 5 fundamental layers in the CNN model. The convolutional layer is composed of a number of convolutional filters, each of which activates different features from the sensor input. After determining a nonlinear function of the input, the activation layer, also known as a Rectified Linear Unit (ReLU), activates the specific neuron. By reducing the spatial dimension of the input, the pooling layer minimizes the number of parameters. The fully-connected layer, which is similar to hidden layers in conventional neural networks, represents essential composite and aggregated features or information from all convolutional layers that have occurred before it. The SoftMax layer normalizes the predictions and allows the network to provide probabilistic outputs. At a SoftMax layer, cross-entropy loss is also assessed. The structure of a CNN network illustrated in Fig. 2.

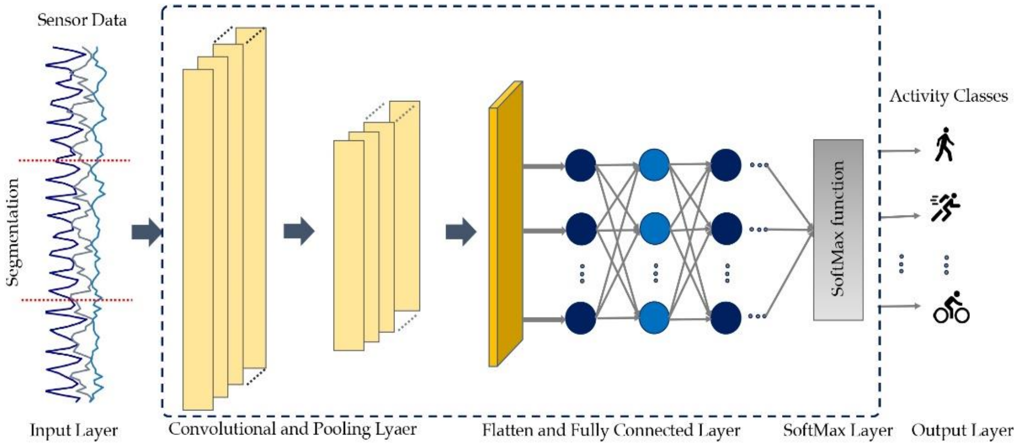

Figure 2. The structure of a CNN network.

C. Recurrent Neural Network

Artificial neural networks with an internal memory are known as Recurrent Neural Networks (RNNs). Because the output computed for the current input is dependent on both the input and the outcomes of previous computations, they are called recurrent neural networks [28]. The current output is really sent back into the network and combined with the current input to generate the next output. RNNs are created for processing sequences, as opposed to CNNs, which are designed specifically for processing a grid of values to extract spatial information. RNNs, unlike normal Feed-Forward Neural Networks (FNNs), preserve a state that can express temporal data from any length of context window [29]. As a result, while a FNN can only map from input to output vectors, an RNN can theoretically map from each input's whole history to each output.

Long Short-Term Memory Networks (LSTMs) are an extension of RNNs that perform substantially better than ordinary RNNs when it concerns to memorizing dependencies for a long period of time [30]. The design of the recurring module in these networks enables this capacity. A layer termed the cell state, as well as three other levels called gates, create these layers. The LSTM memory is the cell state. While LSTM has proven to be a reliable option for avoiding the exploding/vanishing gradient problem, the architecture’s memory cells result in a higher memory required. Cho *et al.* [31] introduced the Gate Recurrent Unit (GRU) network, a unique RNN-based model, in 2014. The GRU is a basic variation of the LSTM that does not include a distinct memory cell in its configuration [32]. In the network of a GRU, there is an update and reset gate that handles the updated degree of each concealed state. It determines which data requirements are to be transferred to the next state and which do not.

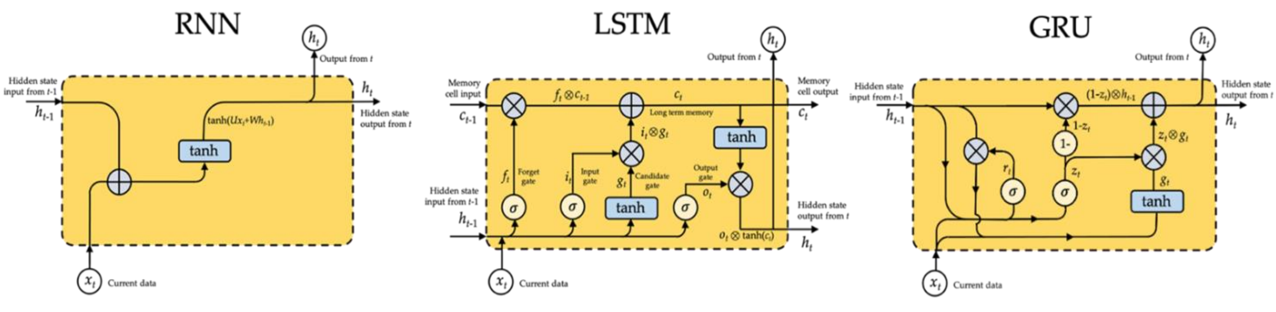

Figure 3. Comparison of (a) RNN, (b) LSTM and (c) GRU structures.

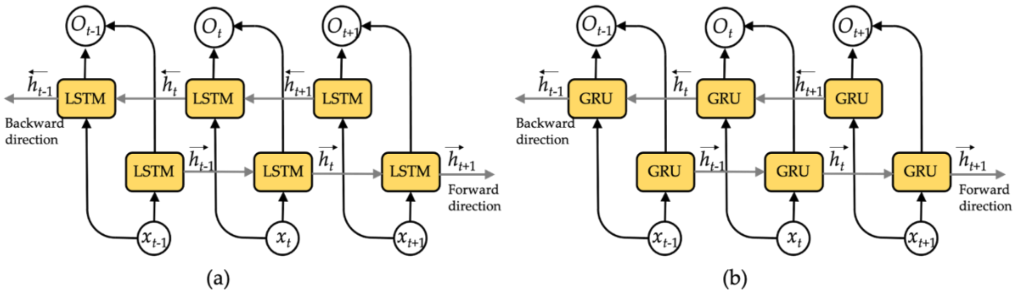

Figure 4. Bidirectional sequence learning models which one hidden layer in the unfold form: (a) Bidirectional LSTM (b) Bidirectional GRU.

Fig. 3 depicts the unit cell of a normal RNN, an LSTM, or a GRU unit cell to summarize the discussion concerning RNN-based models. One significant drawback of such a network is that it is unidirectional. Aside from the current input, the output at any time stage solely depends on the previous data in the input sequence. In some cases, nevertheless, it may be more beneficial to develop predictions based on both the past and the future. This can be performed using a bidirectional network [33], as shown in Fig. 4.

D. *Evaluation Metrics*

To evaluate the five deep learning networks, we used the performance metric from the field of HAR. The accuracy is the standard metric to summarize the overall classification performance for all activity classes:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

where:
TP represents the number of positive instances that were classified as positive,
TN represents the number of negative instances that were classified as negative,
FP represents the number of negative instances that were classified as positive,
FN represents the number of positive instances that were classified as negative.

III. Experiments and Results

The experiment setting, performance measurements, and findings utilized to validate the developed deep learning model for sensor-based HAR are described in this section.

A. *Experiments*

Python's Scikit-learn and Keras were used to develop the deep learning model. The Google Colab Pro platform with GPU Tesla P100-PCIE-16GB was used to run all of the implementations. For the UCI-DSADS dataset, we performed a series of experiments to see which one produced the best results. For the experiments, we employed a 10-fold cross-validation methodology.

B. *Experimental Results*

This study included three different types of sensors (accelerometer, gyroscope, and magnetometer) and five different body positions to investigate the recognition performance of DL networks (torso, left arm, right arm, left leg, and right leg).

Table II shows that the use of sensor data from all body postures, the BiGRU model obtained the maximum accuracy of 99.616%. We separated each type of sensor data to develop the DL models in this experiment, and the study reveals that the BiGRU surpasses the other DL models, which is shown in Fig. 5.

TABLE II. Performance of the Deep Learning Models Used in This Works Using all Three Sensor Data

Model	Accuracy (%)					
	Torso	Right Arm	Left Arm	Right Leg	Left Leg	All
CNN	97.401	97.105	97.368	98.191	98.246	98.728
LSTM	98.629	97.533	97.982	98.728	98.640	99.583
BiLSTM	98.476	98.158	98.520	98.925	98.980	99.594
GRU	98.520	97.895	98.158	98.805	99.178	99.507
BiGRU	98.969	98.235	98.509	**99.046**	99.002	**99.616**

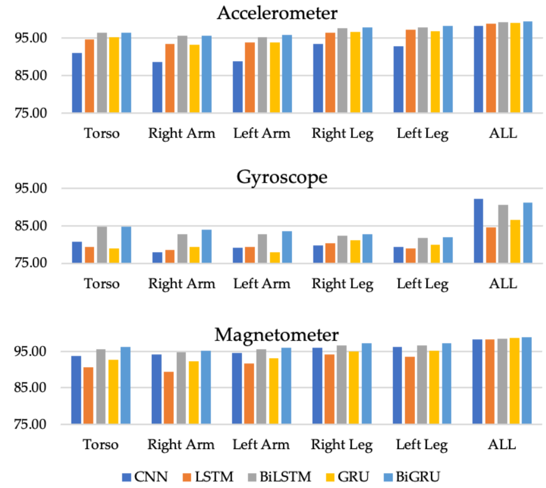

Figure 5. Comparison performance of the five DL models with different sensor data.

Fig. 6 shows the training progress of the proposed BiGRU model, respectively. As shown, we monitored the accuracy and loss trend for up to 200 epochs. In this process, we noticed that the stability of the proposed BiGRU model after ten epochs. Moreover, there is no overfitting for the proposed BiGRU model.

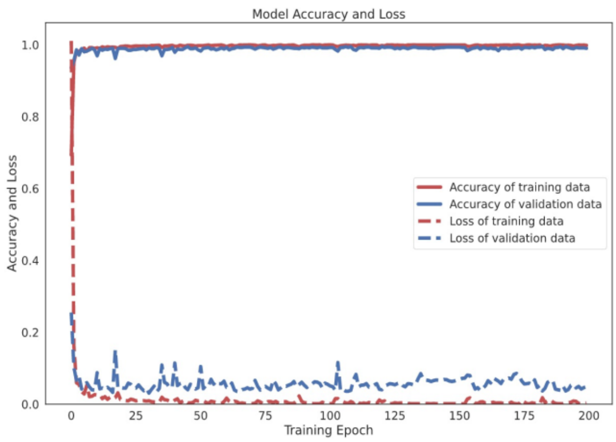

Figure 6. Accuracy and loss trends of the proposed BiGRU model for UCI-DSADS dataset.

Fig. 7 shows a confusion matrix obtained from the experimental results of the proposed BiGRU. The results evidence that the proposed BiGRU well performs for the sport-related activity recognition.

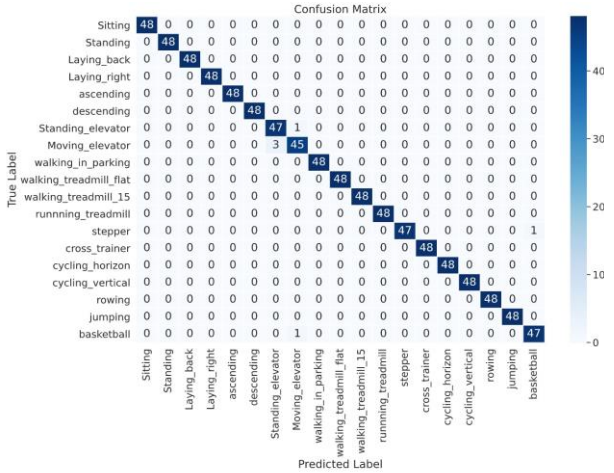

Figure 7. A confusion matrix of the proposed BiGRU model for UCI-DSADS dataset.

IV. Conclusion and Future Works

This study introduced an S-HAR framework for recognizing sport-related activity utilizing multimodal wearable sensors on numerous body positions. A public dataset named UCI-DSADS was used to evaluate the recognition performance of five deep learning models. According to the experimental results, the BiGRU network surpasses other deep learning networks with a maximum accuracy of 99.616%.

In the future, we plan to improve the BiGRU model and study them with various hyperparameters such as learning rate, batch size, optimizer, and many others. We also aim to introduce our model to more complicated activities in order to address other DL models.

Appendix A Some Samples of Dataset Used in This Research

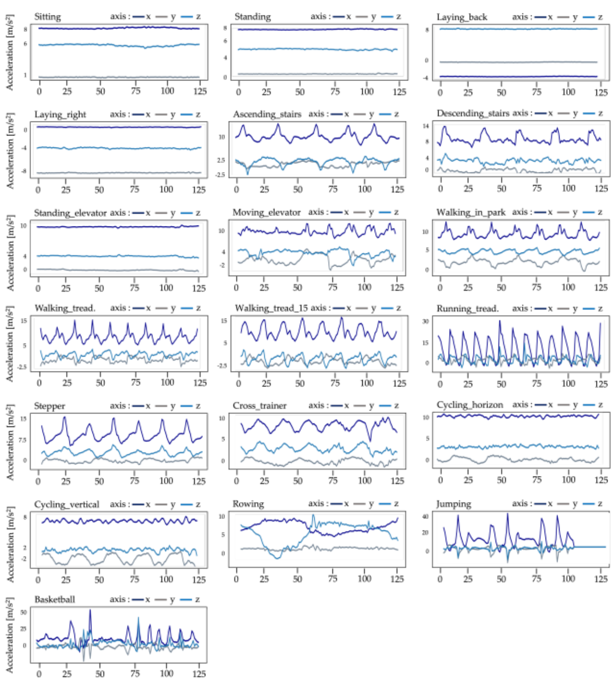

Figure A1. Some samples of accelerometer data from the UCI-DSADS dataset.

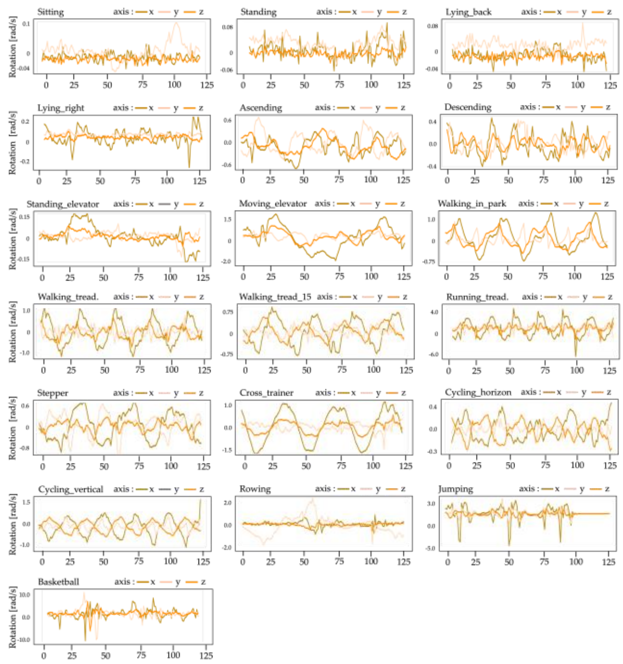

Figure A2. Some samples of gyroscope data from the UCI-DSADS dataset.

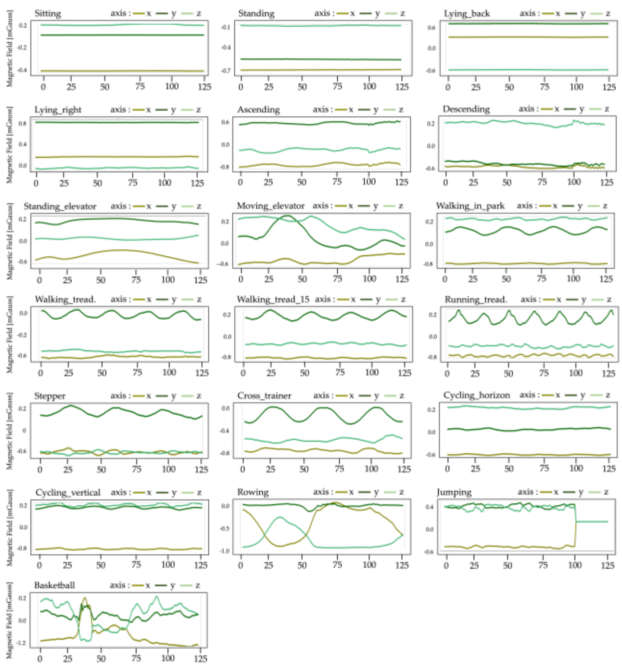

Figure A3. Some samples of magnetometer data from the UCI-DSADS dataset.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization and model analysis, S. Mekruksavanich; resource and data curation, A. Jitpattanakul; methodology and validation, S. Mekruksavanich; data visualization and graphic improvement, A. Jitpattanakul; discussion and final editing, S. Mekruksavanich; writing-review and editing, S. Mekruksavanich; funding acquisition, A. Jitpattanakul and S. Mekruksavanich All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENT

This research was funded by University of Phayao with Grant No. FF64-UoE008 and the Thailand Science Research and Innovation Fund.

REFERENCES

[1] G. Aroganam, N. Manivannan, and D. Harrison, "Review on wearable technology sensors used in consumer sport applications," *Sensors*, vol. 19, no. 9, 2019.

[2] 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.

[3] N. A. Choudhury, S. Moulik, and D. S. Roy, "Physique-based human activity recognition using ensemble learning and smartphone sensors," *IEEE Sensors Journal*, p. 1, 2021.

[4] M. Webber and R. F. Rojas, "Human activity recognition with accelerometer and gyroscope: A data fusion approach," *IEEE Sensors Journal*, p. 1, 2021.

[5] F. A. Garcia, C. M. Ranieri, R. Aparecida, and F. Romero, "Temporal approaches for human activity recognition using inertial sensors," in *Proc. Latin American Robotics Symposium (LARS), Brazilian Symposium on Robotics (SBR) and Workshop on Robotics in Education (WRE)*, 2019, pp. 121-125.

[6] L. Kascak, C. Rebola, R. Braunstein, and J. Sanford, "Mobile application concept development for remote patient monitoring," in *Proc. IEEE International Conference on Healthcare Informatics*, 2013.

[7] S. S. Shah, A. Gvozdanovic, M. Knight, and J. Gagnon, "Mobile app-based remote patient monitoring in acute medical conditions: Prospective feasibility study exploring digital health solutions on clinical workload during the covid crisis," *JMIR Form. Res.*, vol. 5, no. 1, p. e23190, Jan. 2021.

[8] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Comput. Surv.*, vol. 46, no. 3, Jan. 2014.

[9] G. L. Santos, P. T. Endo, K. H. D. C. Monteiro, E. D. S. Rocha, I. Silva, and T. Lynn, "Accelerometer-based human fall detection using convolutional neural networks," *Sensors*, vol. 19, no. 7, 2019.

[10] A. Khandakar, M. E. Chowdhury, R. Ahmed, A. Dhib, M. Mohammed, N. A. M. A. Al-Emadi, and D. Michelson, "Portable system for monitoring and controlling driver behavior and the use of a mobile phone while driving," *Sensors*, vol. 19, no. 7, 2019.

[11] D. Seshadri, R. Li, J. Voos, J. Rowbottom, C. Alfes, C. Zorman, and C. Drummond, "Wearable sensors for monitoring the physiological and biochemical profile of the athlete," *NPJ Digital Medicine*, vol. 2, p. 72, 2019.

[12] D. Quisi-Peralta, V. Robles-Bykbaev, V. Saquicela-Galarza, E. BernalMerchan, P. Suquilanda-Cuesta, and M. Lopez-Nores, "Data repository of mobile applications for people with disabilities in the area of communication and language using data mining techniques," in *Proc. International Conference on Electronics, Communications and Computers*, 2018, pp. 225-231.

[13] J. Dekelver, M. Kultsova, O. Shabalina, J. Borblik, A. Pidoprigora, and R. Romanenko, "Design of mobile applications for people with intellectual disabilities," in *Proc. 1st Conference on Creativity in Intelligent Technologies and Data Science*, 2015.

[14] 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.

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

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

[17] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, "Sensing technology for human activity recognition: A comprehensive survey," *IEEE Access*, vol. 8, pp. 83791-83820, 2020.

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

[19] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach," in *Proc. IEEE SENSORS*, 2016, pp. 1-3.

[20] 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.

[21] K. Wang, J. He, and L. Zhang, "Attention-based convolutional neural network for weakly labeled human activities' recognition with wearable sensors," *IEEE Sensors Journal*, vol. 19, no. 17, pp. 7598-7604, 2019.

[22] M. Munoz-Organero, "Outlier detection in wearable sensor data for human activity recognition (har) based on drnns," *IEEE Access*, vol. 7, pp. 74422-74436, 2019.

[23] B. Barshan and M. C. Yuksek, "Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units," *The Computer Journal*, vol. 57, no. 11, pp. 1649-1667, 2014.

[24] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. International Conference on Engineering and Technology*, 2017, pp. 1-6.

[25] T. Kim, J. Lee, and J. Nam, "Comparison and analysis of samplecnn architectures for audio classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 2, pp. 285-297, 2019.

[26] S. Almabdy and L. Elrefaei, "Deep convolutional neural network-based approaches for face recognition," *Applied Sciences*, vol. 9, no. 20, 2019.

[27] D. Arora, M. Garg, and M. Gupta, "Diving deep in deep convolutional neural network," in *Proc. 2nd International Conference on Advances in Computing, Communication Control and Networking*, 2020, pp. 749-751.

[28] F. A. Uckun, H. Ozer, E. Nurbas, and E. Onat, "Direction finding using convolutional neural networks and convolutional recurrent neural networks," in *Proc. 28th Signal Processing and Communications Applications Conference*, 2020, pp. 1-4.

[29] N. Chowdhury and M. A. Kashem, "A comparative analysis of feedforward neural network recurrent neural network to detect intrusion," in *Proc. International Conference on Electrical and Computer Engineering*, 2008, pp. 488-492.

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.

[31] K. Cho, B. V. Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," in *Proc. Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, Oct. 2014, pp. 103-111.

[32] R. Dey and F. M. Salem, "Gate-variants of Gated Recurrent Unit (GRU) neural networks," in *Proc. IEEE 60th International Midwest Symposium on Circuits and Systems*, 2017, pp. 1597-1600.

[33] H. M. Lynn, S. B. Pan, and P. Kim, "A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks," *IEEE Access*, vol. 7, pp. 145395-145405, 2019.

Sakorn Mekruksavanich received the Ph.D. degree in Computer Engineering from the Chulalongkorn University in 2012. He also holds a M.S. in Computer Science from King Mongkut's Institute of Technology Ladkrabang in 2004 and a B.Eng. in Computer Engineering from Chiang Mai University in 1999. He is currently an assistant professor in the Department of Computer Engineering, School of Information and Communication

Technology at University of Phayao, Phayao, Thailand. His current research interests are deep and machine learning, applying ML techniques in software engineering field, human activity recognition, and wearable sensors.

Anuchit Jitpattanakul received the B.Sc. degree in applied mathematics from King Mongkut's Institute of Technology North Bangkok, the M.Sc. degree in computational science and Ph.D. degree in computer engineering from Chulalongkorn University. He joined the Intelligent and Nonlinear Dynamic Innovations (INDI) Research Center, KMUTNB. He is currently a Faculty Member of the Department of Mathematics, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand. His current research interests include deep learning approaches applied to human activity recognition, wearable sensors, and healthcare applications.