2023 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON) |
979-8-3503-9654-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ECTIDAMTNCON57770.2023.10139396

Physique-Based Human Activity Recognition Using Deep Learning Approaches and Smartphone Sensors

Sakkayaphop Pravesjit[1], Ponnipa Jantawong[2], Anuchit Jitpattanakul[3] and Sakorn Mekruksavanich[2,*]
[1]*Department of Information Technology, School of Information and Communication Technology*
University of Phayao, Phayao, Thailand
sakkayaphop.pr@up.ac.th
[2]*Department of Computer Engineering, School of Information and Communication Technology*
University of Phayao, Phayao, Thailand
ponnipa.jantawong@gmail.com and sakorn.me@up.ac.th
[3]*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*—Understanding human actions via the analysis of sensor data captured by wearable sensors is the goal of the complex subject of study known as sensor-based human activity recognition (S-HAR). Human participants' characteristics are only periodically included in deep learning (DL) approaches to S-HAR. Recognizing people was challenging for these DL methods because of the variety of physical characteristics people have. To address this challenge, we introduce a physique-based S-HAR architecture that could support deep learning networks to achieve higher identification a ccuracies a nd F 1-scores. T he HARSense dataset, a publicly available benchmark S-HAR dataset that compiles raw sensor data acquired from smartphones, was employed to build and evaluate five D L n etworks. A ccording to the experiments, the five models' detection performance improves dramatically when given access to biological data.**

***Keywords*—human activity recognition, smartphone sensor, physical information, deep learning network, classification**

I. Introduction

The field of human activity recognition (HAR) studies methods to precisely identify common human behaviors like walking straight or climbing stairs [1]. In the past, HAR has been used in healthcare systems, including older persons' locomotion and fall detection [2]–[4] and recreational activity tracking [5], [6] to define the performance of life-related aspects better.

Regarding HAR implementations, camera, and radar-based solutions are used, but they have limitations due to high costs, privacy concerns, and processing needs [7]. Wearable inertial measurement units (IMUs: accelerometer and gyroscope) are an opportunity that allows researchers to efficiently analyze longitudinal movement data in both field and laboratory settings at a cheap cost. High-accuracy HAR is now available from wearable IMUs (often coupled with additional sensing modalities, such as a magnetometer, electrocardiograph, or electromyography) using modern classification frameworks[8].

Conventional approaches to sensor-based HAR (S-HAR) [9] have focused on the difficulty of classifying time series of data from several variables. Classifying different kinds of human behavior has been made easier by standard techniques for machine learning, including Naive Bayes, decision trees, and support vector machines [10]. Contrarily, custom feature extraction requires specialized knowledge or experience. A deep model with convolutional layers [11] has been utilized to do autonomous feature extraction in a deep learning (DL) setting. Early DL-based HAR studies focused on using convolutional neural networks (CNNs) to address sensor-based HAR by automatically extracting abstract properties from sensor input [12]. The temporal properties of wearable sensor data are not captured by CNNs, even though CNNs can identify the spatial domain of sensor information and offer adequate performance for basic activities. It could be challenging to set up several classifiers using deep learning techniques to classify complicated human behaviors efficiently. In order to provide weight to temporal information from wearable sensor data, recurrent neural networks (RNNs) are used in HAR. Conversely, the RNN suffers from a training difficulty because it disappears or expands the gradient issue. Researchers created neural networks with long short-term memory (LSTM) to solve this issue. A large number of recent HAR experiments have made use of LSTMs to enhance performance. Hybrid deep learning models have emerged to overcome the shortcomings of CNNs and RNNs [13].

Prevalent in conventional S-HAR arrangements is the usage of accelerometers and gyroscopes. A gyroscope monitors the amount and direction of rotation in 3 dimensions, whereas an accelerometer detects the force of gravity and the acceleration of the user's body in 3 dimensions. Given the difficulty in directly correlating sensor data (raw or processed normalized data) with a particular action, this is one of the most difficult challenges to solve. Several first activity recognition experiments include participants of varying body shapes and genders. They did not consider things like a human's physique, such as height and weight. Data is gathered from various sources,

including sensors, and then actions are categorized using a variety of machine learning and deep learning techniques. Because they need to consider how each person's physique is built, these techniques are not exact. A skinny individual's strolling speed and intensity will be greater than those of a heavier one. The recommended research will concentrate on human activity recognition based on the physique of various people.

In this study, we provide a deep learning–powered S-HAR architecture based on body composition. The provided paradigm considers the respondents' measurable characteristics (such as height and weight) and their everyday behaviors (ADL). This research compares the results of five distinct deep learning networks – CNN, LSTM, bidirectional LSTM (BiLSTM), gated recurrent unit neural network (GRU), and bidirectional GRU (BiGRU) – trained to categorize human activities.

II. Deep Learning for S-HAR

Some recently significant investigations of HAR [14] have shown issues with traditional machine learning algorithms that affect HAR. Since the deployment of handcrafted features depends on the abilities and experience of the individual making the determinations, this constraint applies to them [15]. Even so, deep learning has been deployed in the last several years as a viable alternative strategy that could successfully solve these constraints.

Recently, research has proposed several deep-learning methods that can tackle time-series identification challenges in HAR [16]. Several learning models' identification skills have been studied, with standard activity datasets used as a comparison metric. CNNs and LSTMs are two of the most popular models, and they work well for HAR issues on smartphones because they provide valuable indicators for judging performance. So, this research aims to examine these two models and compare them to one another to see which is more efficient in identifying hand motions from smartwatch input.

In 1997, Schuster and Paliwal presented the BiLSTM to expand the LSTM's capacity for storing information [17]. The BiLSTM communicates with two ostensibly independent hidden layers, and this framework will be able to learn from both current and future sequences simultaneously. The BiLSTM could take future entries without any changes to the setting of the input data. For example, in their S-HAR study, Alawneh et al. [18] compared the efficacy of unidirectional and bidirectional LSTM models using data collected from sensors tracking people's activities. The findings showed that when comparing the BiLSTM to the unidirectional method, the BiLSTM performed better.

The vanishing gradient issue of RNNs can be avoided using LSTM, even though the architecture's memory cells cause a rise in memory utilization. In 2014, Cho et al. [19] unveiled a unique RNN-based model called the GRU network. The GRU is a simplified LSTM implementation that omits the need for dedicated memory cells. Each hidden state in a GRU is updated and reset at a certain point in the network. Specifically, it decides what information is necessary to carry to the next state and what is not. Okai et al. 20 developed an effective GRU-based DL model to supplement data and solve the S-HAR issue. In this investigation, the GRU model beat the LSTM models and showed more resilience.

The result at any given period relies solely on the input sequence's primary data rather than the current input, which is a significant shortcoming of such a network. To produce more accurate forecasts, it could prove helpful in some instances to include not just the previous but also the future. Using a BiGRU model, Alsarhan et al. [21] devised a system for identifying people's actions. Based on the findings, HAR utilizing sensor data employing the BiGRU model was also shown to be a viable option.

III. The Proposed Framework

Now, we detail the approach used in our research to determine the relevance of biological data to S-HAR. The body type S-HAR framework suggested, based on activity recognition chain (ARC) adaptation, is shown in Fig. 1. This methodology outlines the steps we take to identify typical human behavior. We begin preliminary processing on the dataset to clean up the data and normalize the signals. Next, a windowing technique is used to divide the data further. In order to get a set of predicted labels, a deep learning structure is input into the prepared sample data. The resultant set of predictions is then assessed regarding S-HAR benchmarks (i.e., accuracy, precision, recall, and F1-score).

A. HARSense Dataset

This study used the HARSense dataset [22], which may be downloaded for free. Data from this collection was utilized in [23]. Two smartphone sensors (accelerometers and gyroscopes) were employed. The devices were mounted in two distinct places on the individual's body (the front pocket and the waist) to ensure reliable, thorough data gathering. The sensors on both smartphones were used to compile 16 characteristics. The use of an Android app accomplished this. Acceleration caused by gravitational mass, rotation speed, and rotational axes were among the most notable characteristics. Body type similarity was used to group participant data throughout the dataset development process. This method of producing datasets has the drawback of including people of similar body types. This will undoubtedly lead to a decline in the sample size and the quality of the resulting information.

B. DL Architectures

Five deep neural networks were developed and put through their paces to assess the body-based S-HAR provided. Five different kinds of DL networks have been developed, including CNN, LSTM, BiLSTM, GRU, and BiGRU.

A CNN-based feature extraction method was used to capture the local dependency and scale-invariant aspects of the time-series data. The layers that make up a CNN model are the input layer, the convolutional layer, the pooling layer, the flattening layer, the fully connected layer, and the output layer.

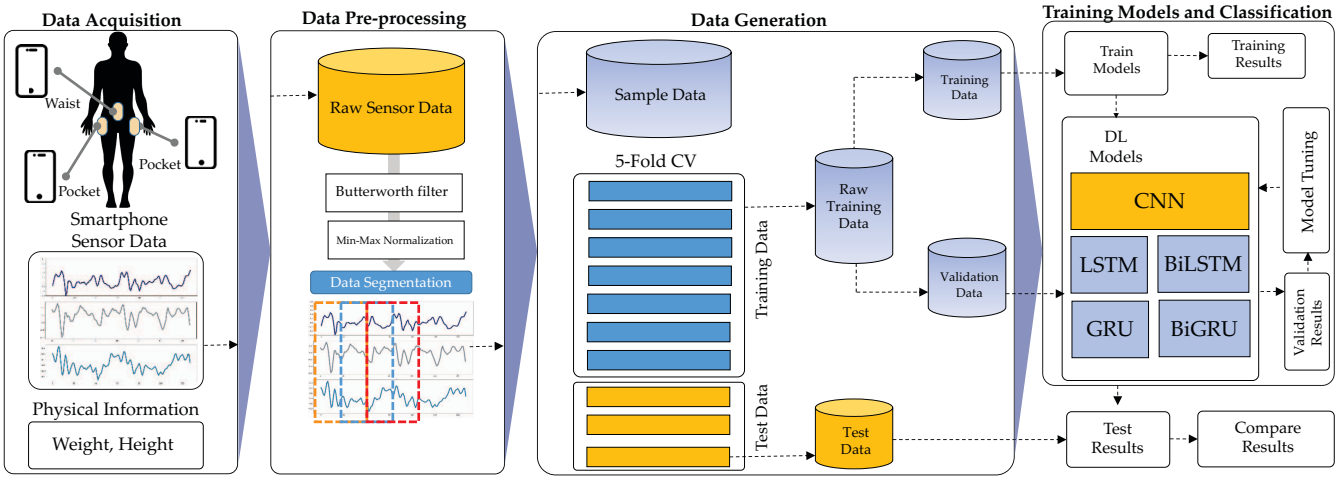

Fig. 1. The proposed physique-based S-HAR framework.

To solve this problem in RNN, an LSTM was developed. It is an improvement over the RNN architecture in that it prevents the gradient from disappearing as quickly and supports more data at each time step. LSTM's unique gate technique allows it to retain and retrieve more data than previous models that express long-time-series data. A large body of research shows that GRU performs better than LSTM in several different scenarios. A BiLSTM or BiGRU could implement LSTM and GRU.

IV. Experimental Results

This work investigated two scenarios to explore the advancement using physical details for S-HAR. Each scenario employed different data for training and testing five deep learning models (CNN, LSTM, BiLSTM, GRU, and BiGRU), as shown in Table I.

TABLE I
RECOGNITION EFFECTIVENESS OF THE DEEP LEARNING MODELS USED IN THIS STUDY

Scenario	Description
I	Only motion sensor data
II	Motion sensor data and physical information

Each experiment's HARSense dataset was operated to train DL models and evaluated by the 5-fold cross-validation procedure. The investigation showed the recognition interpretation of five DL models (CNN, LSTM, BiLSTM, GRU, and BiGRU) on the two scenarios described in Table I.

In Table II, the five deep learning models were trained and tested using only motion sensor data without physical information of subjects. Experimental results showed that the BiGRU achieved the most satisfactory performance with an average accuracy of 89.21% and an average F1-score of 86.16%.

In Table III, the five DL networks were trained and tested employing both sensor data and physical information (weight and height) of participating subjects. Experimental results showed that the BiGRU achieved the most promising interpretation with the highest average accuracy of 93.05% and the highest averaged F1-score of 90.54%.

TABLE II
INTERPRETATION INDICATORS OF DL NETWORKS TRAINED AND TESTED WITHOUT PHYSICAL INFORMATION USING SENSOR DATA

Model	Recognition Effectiveness		
	Accuracy	Loss	F1-score
CNN	77.02% (±0.244%)	0.60(±0.006)	71.98%(±0.284%)
LSTM	89.13% (±0.324%)	0.41(±0.007)	85.95%(±0.417%)
BiLSTM	89.20% (±0.250%)	0.40(±0.011)	86.04%(±0.377%)
GRU	89.14% (±0.325%)	0.38(±0.008)	86.00%(±0.437%)
BiGRU	89.21% (±0.260%)	0.39(±0.017)	86.16%(±0.400%)

TABLE III
INTERPRETATION INDICATORS OF DL NETWORKS TRAINED AND TESTED WITH PHYSICAL INFORMATION USING SENSOR DATA

Model	Recognition Effectiveness		
	Accuracy	Loss	F1-score
CNN	93.14%(±0.861%)	0.19(±0.031)	90.89%(±1.055%)
LSTM	92.93%(±0.621%)	0.31(±0.039)	90.49%(±0.693%)
BiLSTM	92.81%(±0.844%)	0.27(±0.047)	90.24%(±1.107%)
GRU	93.03%(±0.583%)	0.29(±0.035)	90.58%(±0.510%)
BiGRU	93.05%(±0.670%)	0.30(±0.060)	90.54%(±0.832%)

To show improvement by utilizing physical information, we described the results of Tables II and III in a bar chart, as shown in Fig. 2. The comparison results show that the physical information combined with motion sensor data enhances the accuracies and F1-scores of the DL networks. Significantly, CNN improved its recognition interpretation up to 16.12% accuracy and 18.91% F1-score.

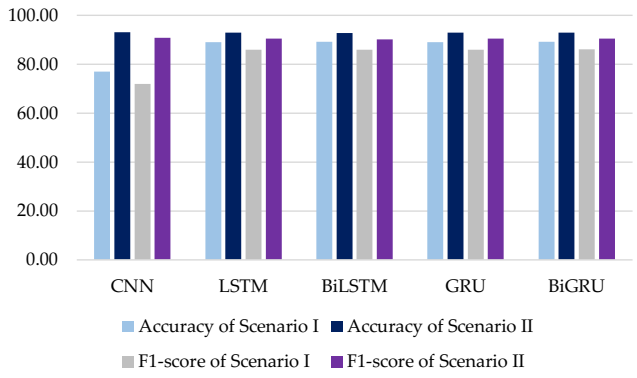

Fig. 2. Comparison results of scenarios I and II.

V. Conclusion and Future Works

This research focuses on body-based S-HAR through the utilization of smartphone sensor data and other physical data. We trained and evaluated five DL models (CNN, LSTM, BiLSTM, GRU, and BiGRU) on two distinct scenarios: with and without physical information. We used a standardized public S-HAR dataset named HARSense dataset. The BiGRU surpasses other DL models with the most excellent average accuracy of 89.21% for the scenario I and 93.05% for scenario II, as determined by experiment results. Nonetheless, CNN is the DL model with the most significant difference in improvement. Based on this research, we can infer that physical information can be supplementary information for S-HAR to increase detection capability.

We will examine further physical facts in the future to determine their influence on S-HAR. Furthermore, we will investigate the physique-based S-HAR for state-of-the-art DL models, including InceptionTime and ResNet.

Acknowledgment

This research project was supported by Thailand Science Research and Innovation Fund; University of Phayao under Grant No. FF66-UoE001; National Science, Research and Innovation (NSRF); and King Mongkut's University of Technology North Bangkok, Contract No. KMUTNB-FF-66-07.

References

[1] 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.

[2] H. Ramirez, S. A. Velastin, I. Meza, E. Fabregas, D. Makris, and G. Farias, "Fall detection and activity recognition using human skeleton features," *IEEE Access*, vol. 9, pp. 33 532–33 542, 2021.

[3] S. Mekruksavanich, P. Jantawong, A. Charoenphol, and A. Jitpattanakul, "Fall detection from smart wearable sensors using deep convolutional neural network with squeeze-and-excitation module," in *2021 25th International Computer Science and Engineering Conference (ICSEC)*, 2021, pp. 448–453.

[4] S. Mekruksavanich and A. Jitpattanakul, "Fallnext: A deep residual model based on multi-branch aggregation for sensor-based fall detection," *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, vol. 16, no. 4, p. 352–364, Sep. 2022.

[5] S. Mekruksavanich and A. Jitpattanakul, "Sport-related activity recognition from wearable sensors using bidirectional gru network," *Intelligent Automation & Soft Computing*, vol. 34, no. 3, pp. 1907–1925, 2022.

[6] S. Mekruksavanich and A. Jitpattanakul, "Multimodal wearable sensing for sport-related activity recognition using deep learning networks," *Journal of Advances in Information Technology*, vol. 13, no. 2, pp. 132–138, April 2022.

[7] F. Demrozi, G. Pravadelli, A. Bihorac, and P. Rashidi, "Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey," *IEEE Access*, vol. 8, pp. 210 816–210 836, 2020.

[8] N. Hnoohom, A. Jitpattanakul, I. You, and S. Mekruksavanich, "Deep learning approach for complex activity recognition using heterogeneous sensors from wearable device," in *2021 Research, Invention, and Innovation Congress: Innovation Electricals and Electronics (RI2C)*, 2021, pp. 60–65.

[9] S. Mekruksavanich, A. Jitpattanakul, K. Sitthithakerngkiet, P. Youplao, and P. Yupapin, "Resnet-se: Channel attention-based deep residual network for complex activity recognition using wrist-worn wearable sensors," *IEEE Access*, vol. 10, pp. 51 142–51 154, 2022.

[10] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "Complex human activity recognition using smartphone and wrist-worn motion sensors," *Sensors*, vol. 16, no. 4, 2016.

[11] Z. Qin, Y. Zhang, S. Meng, Z. Qin, and K.-K. R. Choo, "Imaging and fusing time series for wearable sensor-based human activity recognition," *Information Fusion*, vol. 53, pp. 80–87, 2020.

[12] S. Mekruksavanich and A. Jitpattanakul, "Cnn-based deep learning network for human activity recognition during physical exercise from accelerometer and photoplethysmographic sensors," in *Computer Networks, Big Data and IoT*. Singapore: Springer Nature Singapore, 2022, pp. 531–542.

[13] S. Mekruksavanich and A. Jitpattanakul, "A multichannel cnn-lstm network for daily activity recognition using smartwatch sensor data," in *2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering*, 2021, pp. 277–280.

[14] 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.

[15] 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, deep Learning for Pattern Recognition.

[16] S. Mekruksavanich and A. Jitpattanakul, "Recognition of real-life activities with smartphone sensors using deep learning approaches," in *2021 IEEE 12th International Conference on Software Engineering and Service Science (ICSESS)*, 2021, pp. 243–246.

[17] M. Schuster and K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[18] L. Alawneh, B. Mohsen, M. Al-Zinati, A. Shatnawi, and M. Al-Ayyoub, "A comparison of unidirectional and bidirectional lstm networks for human activity recognition," in *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2020, pp. 1–6.

[19] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," in *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 103–111.

[20] J. Okai, S. Paraschiakos, M. Beekman, A. Knobbe, and C. R. de Sá, "Building robust models for human activity recognition from raw accelerometers data using gated recurrent units and long short term memory neural networks," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 2486–2491.

[21] T. Alsarhan, L. Alawneh, M. Al-Zinati, and M. Al-Ayyoub, "Bidirectional gated recurrent units for human activity recognition using accelerometer data," in *2019 IEEE SENSORS*, 2019, pp. 1–4.

[22] N. A. Choudhury, S. Moulik, and D. S. Roy, "Harsense: Statistical human activity recognition dataset," 2021.

[23] N. A. Choudhury, S. Moulik, and D. S. Roy, "Physique-based human activity recognition using ensemble learning and smartphone sensors," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 16 852–16 860, 2021.