

Sport-Related Activity Recognition from Wearable Sensors Using Bidirectional GRU Network

Sakorn Mekruksavanich¹ and Anuchit Jitpattanakul^{2,*}

¹Department of Computer Engineering, School of Information and Communication Technology, University of Phayao, Phayao, 56000, Thailand

²Intelligent and Nonlinear Dynamic Innovations Research Center, Department of Mathematics, Faculty of Applied Science,

King Mongkut's University of Technology North Bangkok, Bangkok, 10800, Thailand

*Corresponding Author: Anuchit Jitpattanakul. Email: anuchit.j@sci.kmutnb.ac.th

Received: 13 January 2022; Accepted: 15 February 2022

Abstract: Numerous learning-based techniques for effective human activity recognition (HAR) have recently been developed. Wearable inertial sensors are critical for HAR studies to characterize sport-related activities. Smart wearables are now ubiquitous and can benefit people of all ages. HAR investigations typically involve sensor-based evaluation. Sport-related activities are unpredictable and have historically been classified as complex, with conventional machine learning (ML) algorithms applied to resolve HAR issues. The efficiency of machine learning techniques in categorizing data is limited by the human-crafted feature extraction procedure. A deep learning model named MBiGRU (multimodal bidirectional gated recurrent unit) neural network was proposed to recognize everyday sport-related actions, with the publicly accessible UCI-DSADS dataset utilized as a benchmark to compare the effectiveness of the proposed deep learning network against other deep learning architectures (CNNs and GRUs). Experiments were performed to quantify four evaluation criteria as accuracy, precision, recall and F1-score. Following a 10-fold cross-validation approach, the experimental findings indicated that the MBiGRU model presented superior accuracy of 99.55% against other benchmark deep learning networks. The available evidence was also evaluated to explore ways to enhance the proposed model and training procedure.

Keywords: Sport-related activity; human activity recognition; deep learning network; bidirectional gated recurrent unit

1 Introduction

Recently, interest in human activity recognition (HAR) has significantly increased [1]. Physiotherapy, elderly healthcare monitoring and exercise performance assessment of patients can all benefit from recent studies on HAR [2,3]. Many applications based on HAR including recommendation systems, gait analysis and rehabilitation monitoring have been developed [4]. Video-based HAR and sensor-based HAR are different [5]. Video-based HAR, as low cost and easy to deploy, uses a video camera to capture



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

human movement. Results are then processed to classify activities. However, video-based HAR is susceptible to background noises such as illumination and camera angles that impact detection accuracy. Accelerometers, gyroscopes and magnetometer sensors capture human movement in sensor-based HAR, with data then split into separate parts and categorized. One significant aspect of this research field involves wearable sensors [6].

Algorithms including decision trees, naïve Bayes and support vector machines (SVM) have all been utilized in sensor-based HAR to develop classification models [7,8]. Efficient HAR models have been built using machine learning (ML) techniques. However, these methods are constrained by the need for manual feature extraction. Manual aspects have been studied in several ways including data averages, variance and deviation. Problems can arise during handcrafted feature extraction because individual expertise and experience may restrict the accuracy of individually derived characteristics. A vast number of characteristics must be gathered to attain high recognition rates. Recently, deep learning algorithms were developed to overcome this constraint [9]. The shortcomings of human hand-designed feature extraction were avoided by using deep neural networks to learn features automatically.

This study focuses on the HAR of sensor-based wearables by employing deep learning (DL), model construction and hyperparameter adjustment [10]. Previous research using deep learning methodologies investigated complex psychological activities including cooking, writing and eating. These approaches had flaws, with effectiveness restricted while handling multimodal HAR. Here, the identification effectiveness of an existing sensor-based HAR model was enhanced and the impacts of different factors on the detection of complicated activities such as sport-related movements were evaluated. The effectiveness of various standard deep learning models was compared with our recommended deep learning models across multiple learning environments. The key contributions of this study are listed below.

- An architecture for a multimodal bidirectional gated recurrent unit (MBiGRU) neural network was proposed for daily sport-related activities.
- Different wearable sensors were studied to improve the performance of the proposed bidirectional gated recurrent unit neural network.
- Performance of the proposed deep learning model of daily sport-related activity recognition was investigated against the baseline performances established in the DSADS dataset.

The remainder of this paper is organized as follows. Section 2 provides results of related HAR studies. Section 3 describes the proposed MBiGRU for daily sport-related activity recognition. Section 4 outlines the experimental settings, dataset descriptions and results. Research findings are discussed in Section 5, with conclusions and future proposed studies presented in Section 6.

2 Background and Related Studies

HAR from wearable sensor data has emerged as a popular study topic in recent years, with diverse applications in professional and personal situations. HAR primarily focuses on recognizing motions or operations in generally unrestricted surroundings using data from sensor-based wearable devices worn by individuals while conducting various activities. These devices generate data on individual physical activities using a mix of on-device processing and cloud servers, enabling users to access multiple context-adaptive services [11,12].

HAR using sensors has demonstrated a high degree of accuracy and utility in specific or semi-controlled environments. However, constructing effective classifiers capable of recognizing numerous actions is a complicated process requiring a substantial quantity of labeled training data specific to the context of interest.

2.1 Sensor-Based HAR

Previously published studies on HAR using inertial measurement units (IMUs) mostly focused on acceleration signals. Accelerometers, as small, affordable and ubiquitous devices are often incorporated into consumer items. Typically, previous studies on IMU-based sensors adopted a multi-step approach involving aggregating and annotating a subset of the sensor signal, summarizing the information in the subset using various signal features and instant classification of the physical activity using machine learning [13,14].

The approaches used in different applications varied and were classified according to their learning methodology (supervised and semi-supervised) and reaction time (real-time and offline) [15–18]. Depending on the feature extraction and summarization methodologies, HAR systems were built using handcrafted or learned features. Handcrafted characteristics were selected independently, and often comprised a wide variety of statistical data and features based on human motion models, referred to as physical features [19]. By contrast, learned features were often based on feature selection procedures as deep learning techniques. Support vector machines, Gaussian mixture models, tree-based classifiers such as random forest, and hidden Markov models are all examples of standard classifiers [8]. Recent developments in end-to-end systems incorporate numerous processes.

2.2 Recognized Activities in Professional Settings

HAR focuses on identifying everyday behaviors in outdoor and indoor situations including jogging, traveling, relaxing and lying down. Considerable emphasis has been placed on sporting situations where actions are combined with precise timing. Up until now, minimal attention has focused on professional activities. Tracking the movements of healthcare professionals such as physicians and nurses [20] and, on a lesser scale, activities such as cooking [21] are now gaining increased attention. A few small-scale studies considered activity recognition in construction activities. Joshua et al. [22] investigated masonry operations in a laboratory using accelerometers. Their investigation demonstrated a categorization accuracy of up to 80% in moderately unrestricted situations, while Akhavian et al. [3] recreated a three-class construction activity using two subjects who engaged in sawing, hammering and twisting a wrench as well as loading and unloading operations. Their three-class model achieved near-90% accuracy, despite the substantial unpredictability of users and activities. Other construction-related research concentrated on instances without a human component, such as monitoring the actions of 191 specific pieces of equipment or machinery. Sensor modality has a significant influence on the effectiveness of an activity detection scheme [5]. Four ways to categorize sensor modalities include wearable, ambient, object and other. These are discussed in detail below.

2.2.1 Wearable Sensors

Wearable sensors are most often employed for individual action detection because they can immediately and effectively collect physical movements [23–25]. They include various gadgets ranging from mobile phones to smartwatches to smart bands [26]. The accelerometer, gyroscope and magnetometer are the three most often utilized innovative technologies in HAR studies.

• Accelerometer: The ratio at which the velocity of a moving object changes is measured using an accelerometer as a motion sensor. Gravity forces (g) or meters per second squared ($m = s^2$) are the units of measurement. Sampling is often in the tens to hundreds of Hz frequency range. Accelerometers can be attached to many body locations including the waist, arm, ankle and wrist. A common accelerometer has three axes that can be used to generate a three-variate time series. Tri-axial accelerometer signal outputs for the activity classes of standing, walking and leaping are shown in Fig. 1a.

- **Gyroscope:** Gyroscopes are instruments used to determine the speed and direction of rotation. Degrees per second (deg/s) are used to express angular velocity, with tens to hundreds of Hz as the sample rate. These sensors are generally incorporated into one another and attached to the same region of the user's body. There are three axes in a gyroscope. Gyroscope sensing movement outputs are shown in Fig. 1b for standing, walking and leaping.
- **Magnetometer:** Magnetometers are wearable sensors typically used for motion identification by monitoring the magnetic field movement in a specific area. Sampling rates range from tens to hundreds of Hz in units of Tesla (*T*). A magnetometer has three axes. Movement outputs as the sensor's depiction of standing, walking and leaping are shown in Fig. 1c.



Figure 1: Tri-axial signals representing (a) Accelerometer, (b) Gyroscope and (c) Magnetometer

2.2.2 Sensor-Based HAR Framework

The HAR framework is shown in Fig. 2 as a sensor-based architecture. Data captured from wearable sensors, data segmentation, feature extraction, model training and classification are all part of this process.

The initial step in the data collection procedure involves gathering sensor data as part of an activity dataset based on sensors. These sensor datasets are then divided into "windows" with predetermined sample intervals. A tiny window in each display represents the sensor signal. In Fig. 2, the windows are divided into non-overlapping and overlapping categories, respectively. The extraction of features from sensors in time-series data is critical before training a model, with categorization as the last step in the HAR architecture operation.



Figure 2: Sensor-based HAR architecture

2.3 Convolutional Neural Network in HAR

Convolutional neural networks (CNNs) are commonly used for classification, voice recognition and natural language processing tasks [27,28]. Existing studies [29,30] indicated that CNNs accurately identified HAR. A CNN-based feature extraction approach was employed to extract the local dependence and scale-invariant properties of the time-series data. Input, convolutional, pooling layers, flattening and fully connected layers and an output layer comprised the structure of the CNN model, as depicted in Fig. 3.



Figure 3: Structure of the CNN model

The first layer of the CNN model receives sensor data as an input. The convolutional layers are layered with appropriate max-pooling layers and ReLU activation functions. This level provides automatic feature extraction. Flattened and ultimately linked layers contribute to the third segment of the design. When the preceding layer (max-pooling layer) passes the feature maps to the flattening layer, the feature maps are transformed into a single column vector and given to the fully connected layer. The fully linked layer

carries out the categorization procedure. After the fully connected layer, the output layer, i.e., the Softmax layer, receives the outputs and determines the probability distribution of each category.

2.4 Recurrent Neural Network in HAR

For applications requiring the computation of temporal correlations between neurons, such as natural language processing and voice recognition, a recurrent neural network (RNN) varies from a CNN [31]. RNNs are sequentially fed, and previous inputs are less likely to impact the final prediction outcome. Long short-term memory (LSTM) was designed to address this issue in RNNs by tackling the RNN gradient vanishing issue, allowing more time steps of information than the RNN design. LSTM has a distinctive gated method that enables it to store and access more information than other models to represent extended time-series data [32,33]. Numerous studies have demonstrated that GRU outperforms LSTM in several domains. The model has two gate techniques that can be used as multiple storage units including the z_t and z_t gates for the initial LSTM model. The two gates and a storage unit of the GRU at a time step t can be depicted as the following four mathematical formulae:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{1}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{2}$$

$$c_t = \tan h(W_c x_t + U_c R_c h_{t-1} + b_c) \tag{3}$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot b_c \tag{4}$$

where x_t , h_t , and c_t denote the input vector, the hidden state and the storage unit state, respectively. W_t , W_r , W_c , U_z , U_r , and U_c denote the weight matrices, b_z , b_r , and b_c denote the bias vector and σ represents the sigmoid function.

A full description of feature representations cannot be represented using the GRU model since the input series can only be considered in one way [34,35]. Thus, the BiGRU architecture is created by modeling the input series in forward and reverse orientations. The mathematics of the BiGRU model can be described as:

$$\overrightarrow{h_t} = \overrightarrow{GRU_t}(x_t, h_{t-1})$$
(5)

$$\overleftarrow{h_t} = \overleftarrow{GRU_t}(x_t, h_{t+1}) \tag{6}$$

The final output of the BiGRU network can be concatenated as follows:

$$Y_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{7}$$

3 Proposed Methodology

Data acquisition, data segmentation, model training and model assessment are all part of the proposed DL-based HAR technique, as shown in Fig. 4. A detailed description of each process is given below.

3.1 UCI-DSADS Dataset Description

The University of California (UC) Irvine machine learning repository introduced a dataset called "Daily and Sports Activities Dataset" (UCI-DSADS dataset) that was utilized here to evaluate the proposed model. The UCI-DSADS dataset used 5 MTx with 2-DOF models of orientation monitors to collect movement sensor data from eight participants as they performed 19 tasks in five distinct body postures (see Fig. 5).



Figure 4: Deep learning-based HAR methodology



Figure 5: Unit positions on the participant's body

Each of the eight participants (four female and four male aged 20 to 30) was allowed five minutes to complete each task. The eight participants were responsible for completing the prescribed duties. Activity data were captured at 25 Hz using the sensor components, with five-second portions of the five-minute signals used to generate features. A total of 480 signal segments were captured for each task. The Machine Learning Repository at the University of California, Irvine as the UCI-DSADS collection contains the sensor data shown in Tab. 1.

The 3D motion monitor was used to capture tri-axial IMU data at a 25 Hz sample rate at five locations on the body of each participant. Eight participants performed 19 different activities during the five minutes allotted (such as sitting, running at varying speeds, bicycling and rowing). A total of 1,140,000 acceleration, gyroscope and magnetometer samples were included in the collection, and the UCI-USADS dataset was relatively balanced, as shown in Fig. 6.

Туре	Detail	Abbreviation
Total subjects	8	S1, S2,, S8
Total activities	19	_
Activities	Sitting	ADL 1
	Standing	ADL 2
	Lying on the back	ADL 3
	Lying on the right side	ADL 4
	Ascending stairs	ADL 5
	Descending stairs	ADL 6
	Standing still in an elevator	ADL 7
	Moving around in an elevator	ADL 8
	Walking in a parking lot	ADL 9
	Walking on a treadmill at 4 km/h on the flat	SP 1
	Walking on a treadmill with 15 inclined pos.	SP 2
	Running on a treadmill at 8 km/h	SP 3
	Exercising on a stepper	SP 4
	Exercising on a cross-trainer	SP 5
	Cycling on an exercise bike in horizontal pos.	SP 6
	Cycling on an exercise bike in vertical pos.	SP 7
	Rowing	SP 8
	Jumping	SP 9
	Playing basketball	SP 10
Body locations	Torso	Т
	Right arm	RA
	Left arm	LA
	Right leg	RL
	Left leg	LL
Motion sensors	Tri-axial accelerometer (m/s^2)	3D-Acc
	Tri-axial gyroscope (deg/s)	3D-Gyro
	Tri-axial gyroscope (deg/s)	3D-Mag

Table 1: Daily sport-related activities in the UCI-DSADS dataset

3.2 Proposed Architecture of MBiGRU

Fig. 7 shows the proposed architecture of the multimodel bidirectional gated recurrent unit deep learning model (MBiGRU), comprising the CNN structure for feature extraction and bidirectional GRU layers for time series prediction across five channels that were eventually coupled. Three layers comprised each track as an input layer, a one-dimensional convolutional neural network (1DCNN) layer and a BiGRU layer. The Conv1D layers offered a direct mapping and abstract representation of sensor inputs utilizing a 64-filter configuration for feature extraction. Convolution operators were used to extract the features from kernels. A Conv1D layer with a kernel size of 5 was used for the appropriate channels. Max-pooling condensed the

learned features into distinct parts without significantly degrading each Conv1D outcome. Conv1D layers *n* and max-pooling layers were fed into a bidirectional GRU layer with 128 units and a dropout ratio of 25%. Due to its forward and backward repetitions, the BiGRU adjusted effectively to the inner state. Flattening and combining the outputs of five channels was performed in the model. An additional layer of connectivity enabled data to be sent in ways interpreted as different activity types. The softmax activation function established the probability distribution throughout the predicted scenarios in the dense layer of the model.



Figure 6: Number of sensors in each position of the UCI-DSADS dataset



Figure 7: Proposed MBiGRU model

3.3 Model Assessment Criteria

Various metrics such as the F1-score and confusion matrix were used in the HAR research to evaluate multiple aspects. In most cases, HAR models are assessed according to these criteria. The TP (true positive) and TN (true negative) signify proper classification (true negative). False forecasting occurs when the projected conclusion is "yes", but the true answer is "no" (FP or false positive). A FN (false negative) occurs when the expected outcome is "no", but the actual product is "yes" (false negative). The overall performance of accuracy can be represented as:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$
(8)

Measurements with a wide variety of applications including recall and F1-score were also used, while the precision of Eq. (9) was employed to determine how many attributes of a particular class were properly allocated and how many were incorrectly assigned.

$$Precision = \frac{TP}{TP + FP}$$
(9)

Recall can be defined as the percentage of accurately anticipated positives to the total number of positives observed.

$$Recall = \frac{TP}{TP + FN}$$
(10)

IASC, 2022, vol.34, no.3

Two components for calculating an F1-score are precision (the proportion of adequately predicted positive observations to complete anticipated positive investigations) and recall (the proportion of correctly forecasted positive studies to actual observed positive experiments).

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$
(11)

A confusion matrix, as a square matrix of order n classes, is often used to display the results of multiclass classification to better understand the classification levels of each category. Matrix component c represents the number of times a class i is classed as a class j. The confusion matrix c focuses primarily on errors in classification.

4 Research Experiments and Findings

Aspects of the proposed MBiGRU model evaluation included the experimental setup, performance measures and outcomes.

4.1 Experimental Setting

The model environment was implemented by utilizing Python programming language with various libraries such as Tensorflow, Keras and Scikit-learn to develop the introduced MBiGRU model. A Google Colab-Pro platform with a GPU Tesla P100-PCIE-16GB was employed to execute all algorithm implementations. Several experiments were performed on the UCI-DSADS dataset to evaluate the optimal approach. The experiments employed a 10-fold cross-validation methodology, with three distinct physical activity scenarios in the UCI-DSADS dataset as follows:

- Scenario I: Using only sport-related activities (SPT)
- Scenario II: Using only activities of daily living (ADL)
- Scenario III: Using all sport-related and daily living activities in the UCI-DSADS dataset

Results in Tabs. 2–4 showed that the proposed MBiGRU gave optimal accuracy and F1-score for sportrelated activities in all experimental scenarios. Highest accuracy was 99.92%, with F1-score 99.92%. When comparing the two baseline models, CNN achieved better performance than GRU in scenario I, while the opposite results were found in scenarios II and III.

Model	Recognition performance (Mean ± Std.)			
	Accuracy	Precision	Recall	F1-score
CNN	99.79% (±0.260%)	99.80% (±0.245%)	99.79% (±0.260%)	99.79% (±0.259%)
GRU	99.51% (±0.127%)	99.91% (±0.122%)	99.91% (±0.127%)	99.91% (±0.127%)
MBiGRU	99.92% (±0.126%)	99.93% (±0.124%)	99.92% (±0.126%)	99.92% (±0.126%)

Table 2: Comparison of recognition performance results for scenario I

4.2 Experimental Results

The proposed MBiGRU model was trained on a variety of datasets in each of the three scenarios, and then compared to two other standard deep learning models. Accuracy, precision, recall and F1-score as the four traditional metrics were used to assess each classifier, with results shown in Tabs. 2–4.

Training progress for the baseline CNN and GRU and the proposed MBiGRU are shown in Figs. 8–10. Accuracy and loss trends were tracked for up to 200 epochs. Stability of the proposed MBiGRU model after ten epochs was better than the two benchmark models. Furthermore, the MBiGRU model indicated that overfitting did not occur.

Model	Recognition performance (Mean \pm Std.)			
	Accuracy	Precision	Recall	F1-score
CNN	99.00% (±0.835%)	99.03% (±0.827%)	99.00% (±0.835%)	99.00% (±0.835%)
GRU	99.07% (±0.954%)	99.12% (±0.851%)	99.07% (±0.954%)	99.06% (±0.990%)
MBiGRU	99.21% (±0.362%)	99.23% (±0.351%)	99.21% (±0.362%)	99.21% (±0.363%)

Table 3: Comparison of recognition performance results for scenario II

Table 4: Comparison of recognition performance results for scenario III

Model	Recognition performance (Mean ± Std.)			
	Accuracy	Precision	Recall	F1-score
CNN	98.73% (±0.773%)	98.78% (±0.725%)	99.00% (±0.773%)	98.72% (±0.777%)
GRU	99.51% (±0.210%)	99.52% (±0.196%)	99.07% (±0.209%)	99.50% (±0.212%)
MBiGRU	99.55% (±0.279%)	99.56% (±0.277%)	99.21% (±0.279%)	99.55% (±0.281%)



Figure 8: Accuracy and loss trends of the deep learning models using the dataset in scenario I: (a) CNN, (b) GRU and (c) MBiGRU



Figure 9: Accuracy and loss trends of the deep learning models using the dataset in scenario II: (a) CNN, (b) GRU and (c) MBiGRU

5 Discussion

The study findings detailed in Section 4 were examined.

5.1 Impact of Different Types of Activities

Experiments were conducted to further understand how diverse sensory inputs affected recognition by evaluating the UCI-DSADS dataset for the three types of activities as sport-related activities, activities of daily living and both these activities. Mobile sensing data were used to assess sports activities, as shown in Fig. 11. Performance results of the three deep neural networks are shown in Tabs. 2–4 for each scenario. The proposed MBiGRU functioned effectively, with accuracy of 99.92% for sport-related sensor data.

For sport-related activities, the performance of the proposed MBiGRU revealed that the model was efficient, as illustrated in Fig. 12. However, Figs. 12b and 12c show that certain behaviors in actual life, such as standing and traveling in an elevator, confounded the suggested model.



Figure 10: Accuracy and loss trends of the deep learning models using the dataset in scenario III: (a) CNN, (b) GRU and (c) MBiGRU



Figure 11: Results of different activity types in the UCI-DSADS dataset

5.2 Training Process

The CNN, GRU and MBiGRU models were used to analyze the sensor data from the UCI-DSADS. First, raw tri-axial data from the accelerometer, gyroscope and magnetometer were used to classify activities. Fig. 13 shows how deep learning models may be used to learn from raw sensor data. The loss

rate decreased dramatically, while the accuracy rate steadily improved. There was no sign of a problem. In other words, no overfitting issues were found with network learning. The testing set had an average accuracy of 99.92%, 99.21% and 99.55%.





5.3 Limitations of the Proposed Model

This study had limitations because the deep learning models were trained and evaluated using laboratory data. Previous research demonstrated that the effectiveness of classification methods in laboratory circumstances does not always correctly represent their performance in real-world situations [36]. Also, this research did not tackle the difficulty of transitional behaviors (Sit-to-Standing, Sit-to-Lay and so on). This is



a topic for future study. Nevertheless, the recommended HAR framework showed high efficiency as a deep learning model and can be applied for various practical applications in smart homes, including fall detection.

Figure 13: Accuracy and loss trends of the MiGRU model in different scenarios: (a) SPT, (b) ADL and (c) All (SPT and ADL)

5.4 Applications of the Proposed Model

The proposed MBiGRU model was designed to detect complicated psychological actions such as sportrelated activities, eating activities and other tasks involving the use of hands. These human physical behaviors were assessed using movement data from multimodal sensors attached at various body positions. This model used multimodal processes to extract temporal information and concatenate them to categorize complicated actions in parallel computing. Numerous HAR applications, including fall detection and identification of construction worker activity, can be implemented using the proposed approach.

6 Conclusions and Future Studies

Wearable sensors and a deep learning model named MBiGRU were used to identify everyday sporting activities. Standard deep learning algorithms were assessed against our proposed approach using the UCI-DSADS wearable dataset. A 99.55% accuracy rate and a 99.55% F1-score indicated that our proposed MBiGRU model surpassed the baseline deep learning models in the experiments. The proposed model

boosted the effectiveness of sensor-based HAR using the recommended MBiGRU paradigm. The influence of various forms of physical activities was also investigated by classifying them into three categories. Results demonstrated that our proposed model accurately recognized complicated activities. The multi-channel blocks comprised of CNN and BiGRU layers offered primary benefit over existing frameworks for complicated HAR.

In the future, we intend to develop and investigate the MBiGRU model using a variety of hyperparameters including learning rate, batch size and optimizer. We also plan to apply our model to more complex actions to address concerns about other deep learning techniques and sensor-based HAR by evaluating other publicly available complex activity datasets such as UT-Complex and WISDM.

Funding Statement: This research project was supported by the Thailand Science Research and Innovation Fund, the University of Phayao (Grant No. FF65-RIM041), and King Mongkut's University of Technology North Bangkok, Contract No. KMUTNB-65-KNOW-02.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- C. Jobanputra, J. Bavishi and N. Doshi, "Human activity recognition: A survey," *Procedia Computer Science*, vol. 155, pp. 698–703, 2019.
- [2] M. M. Aborokbah, S. Al-Mutairi, A. K. Sangaiah and O. W. Samuel, "Adaptive context aware decision computing paradigm for intensive health care delivery in smart cities-a case analysis," *Sustainable Cities and Society*, vol. 41, pp. 919–924, 2018.
- [3] R. Akhavian and A. H. Behzadan, "Construction activity recognition for simulation input modeling using machine learning classifiers," in *Proc. of the 2014 Winter Simulation Conf.*, Savannah, GA, USA, pp. 3296– 3307, 2014.
- [4] A. R. Javed, U. Sarwar, M. Beg, M. Asim, T. Baker et al., "A collaborative healthcare framework for shared healthcare plan with ambient intelligence," *Human-centric Computing and Information Sciences*, vol. 10, no. 40, pp. 1–21, 2020.
- [5] L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee *et al.*, "Sensor-based and vision based human activity recognition: A comprehensive survey," *Pattern Recognition*, vol. 108, pp. 1–24, 2020.
- [6] T. Zebin, P. J. Scully and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach," in *Proc. of the 2016 IEEE SENSORS*, FL, USA, pp. 1–3, 2016.
- [7] D. Anguita, A. Ghio, L. Oneto, X. Parra and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *Proc. of the 2012 Int. Workshop on Ambient Assisted Living*, Vitoria-Gasteiz, Spain, pp. 216–223, 2012.
- [8] O. Lara and M. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.
- [9] 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.
- [10] I. K. Ihianle, A. O. Nwajana, S. H. Ebenuwa, R. I. Otuka, K. Owa et al., "A deep learning approach for human activities recognition from multimodal sensing devices," *IEEE Access*, vol. 8, pp. 179028–179038, 2020.
- [11] E. Fridriksdottir and A. G. Bonomi, "Accelerometer-based human activity recognition for patient monitoring using a deep neural network," *Sensors*, vol. 20, no. 22, pp. 1–13, 2020.
- [12] X. Zhou, W. Liang, K. I. -K. Wang, H. Wang, L. T. Yang *et al.*, "Deep-learning-enhanced human activity recognition for internet of healthcare things," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6429–6438, 2020.
- [13] 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.

- [14] C. Richter, M. O'Reilly and E. Delahunt, "Machine learning in sports science: Challenges and opportunities," Sports Biomechanics, pp. 1–7, 2021.
- [15] J. W. Goodell, S. Kumar, W. M. Lim and D. Pattnaik, "Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis," *Journal of Behavioral and Experimental Finance*, vol. 32, pp. 100577, 2021.
- [16] L. Peng, L. Chen, Z. Ye and Y. Zhang, "Aroma: A deep multi-task learning based simple and complex human activity recognition method using wearable sensors," in *Proc. of the 2018 ACM Interactive, Mobile, Wearable and Ubiquitous Technologies*, NY, USA, pp. 1–16, 2018.
- [17] P. Kaur, R. Gautam and M. Sharma, "Feature selection for Bi-objective stress classification using emerging swarm intelligence metaheuristic techniques," in *Proc. of the Data Analytics and Management*, Singapore, pp. 357–365, 2022.
- [18] S. Sharma, G. Singha and M. Sharma, "A comprehensive review and analysis of supervised-learning and soft computing techniques for stress diagnosis in humans," *Computers in Biology and Medicine*, vol. 134, pp. 140450, 2021.
- [19] 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, pp. 1–24, 2016.
- [20] R. Maskeliunas, R. Damasevicius and S. Segal, "A review of internet of things technologies for ambient assisted living environments," *Future Internet*, vol. 11, no. 12, pp. 1–23, 2019.
- [21] U. Emir, K. Ejub, M. Zakaria, A. Muhammad and B. Vanilson, "Immersing citizens and things into smart cities: A social machine-based and data artifact-driven approach," *Computing*, vol. 102, no. 7, pp. 1567–1586, 2020.
- [22] L. Joshua and K. Varghese, "Construction activity classification using accelerometers," in Proc. of the Construction Research Congress, Alberta, Canada, pp. 61–70, 2010.
- [23] S. Mekruksavanich and A. Jitpattanakul, "LSTM networks using smartphone data for sensor-based human activity recognition in smart homes," *Sensors*, vol. 21, no. 5, pp. 1–25, 2021.
- [24] M. A. Hanif, T. Akram, A. Shahzad, M. A. Khan, U. Tariq et al., "Smart devices based multisensory approach for complex human activity recognition," Computers, Materials and Continua, vol. 70, no. 2, pp. 3221–3234, 2022.
- [25] A. Gumaei, M. Al-Rakhami, H. AlSalman, S. M. M. Rahman and A. Alamri, "DL-HAR: Deep learning-based human activity recognition framework for edge computing," *Computers, Materials and Continua*, vol. 65, no. 2, pp. 1033–1057, 2020.
- [26] L. Peng, L. Chen, M. Wu and G. Chen, "Complex activity recognition using acceleration, vital sign, and location data," *IEEE Transactions on Mobile Computing*, vol. 18, no. 7, pp. 1488–1498, 2019.
- [27] S. Almabdy and L. Elrefaei, "Deep convolutional neural network-based approaches for face recognition," *Applied Sciences*, vol. 9, no. 20, pp. 1–24, 2019.
- [28] H. Polat and H. D. Mehr, "Classification of pulmonary CT images by using hybrid 3d-deepconvolutional neural network architecture," *Applied Sciences*, vol. 9, no. 5, pp. 1–15, 2019.
- [29] W. Gao, L. Zhang, W. Huang, F. Min, J. He *et al.*, "Deep neural networks for sensor-based human activity recognition using selective kernel convolution," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.
- [30] W. Huang, L. Zhang, W. Gao, F. Min and J. He, "Shallow convolutional neural networks for human activity recognition using wearable sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1– 11, 2021.
- [31] A. Murad and J. -Y. Pyun, "Deep recurrent neural networks for human activity recognition," Sensors, vol. 17, no. 11, pp. 1–17, 2017.
- [32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [33] Y. Chen, K. Zhong, J. Zhang, Q. Sun and X. Zhao, "Lstm networks for mobile human activity recognition," in Proc. of the 2016 Int. Conf. on Artificial Intelligence: Technologies and Applications, Bangkok, Thailand, pp. 50–53, 2016.

- [34] K. Cho, B. van Merriënboer, D. Bahdanau and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," in *Proc. of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, Doha, Qatar, pp. 103–111, 2014.
- [35] S. Mekruksavanich and A. Jitpattanakul, "Deep learning approaches for continuous authentication based on activity patterns using mobile sensing," *Sensors*, vol. 21, no. 22, pp. 1–21, 2021.
- [36] I. C. Gyllensten and A. G. Bonomi, "Identifying types of physical activity with a single accelerometer: Evaluating laboratory-trained algorithms in daily life," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2656–2663, 2011.