ICIC Express Letters
Volume **19**, Number **3**, March **2025**

 ISSN 1881-803X
pp. **343–351**

A DEEP NEURAL NETWORK WITH AGGREGATED RESIDUAL TRANSFORMATION FOR SMARTWATCH-BASED HUMAN ACTIVITY RECOGNITION IN REAL WORLD SITUATIONS

Sakorn Mekruksavanich[1], Olarik Surinta[2] and Anuchit Jitpattanakul[3,4,*]

[1]Department of Computer Engineering
School of Information and Communication Technology
University of Phayao
19 Moo 2, Tambon Mae Ka, Amphur Mueang, Phayao 56000, Thailand
sakorn.me@up.ac.th

[2]Multi-agent Intelligent Simulation Laboratory (MISL) Research Unit
Department of Information Technology
Faculty of Informatics
Mahasarakham University
Khamriang Sub-District, Kantarawichai District, Mahasarakham 44150, Thailand
olarik.s@msu.ac.th

[3]Department of Mathematics
Faculty of Applied Science
[4]Intelligent and Nonlinear Dynamic Innovations Research Center
Science and Technology Research Institute
King Mongkut's University of Technology North Bangkok
1518 Pracharat 1 Road, Wongsawang, Bangsue, Bangkok 10800, Thailand
[*]Corresponding author: anuchit.j@sci.kmutnb.ac.th

Received April 2024; accepted July 2024

Abstract. *The field of pervasive computing focuses on using sensors to identify human activities, a practice commonly known as Sensor-based Human Activity Recognition (S-HAR). The objective of S-HAR is to automatically evaluate and understand real-time events and their contextual information by utilizing sensor data. Activity identification has various applications, including surveillance systems, medical monitoring systems, and systems involving wearable intelligent devices like smartwatches. Contemporary HAR algorithms are typically developed and evaluated using controlled conditions, which limits their effectiveness in real-life scenarios where sensor data may be incomplete or corrupted and human actions are spontaneous and unscripted. This study aims to identify human behavior in real-world scenarios. To improve the efficiency of the action comprehension structure, we propose a novel deep neural network architecture called ResNeXt, which incorporates an aggregated residual transformation component. This component enables the framework to categorize different human actions effectively and accurately. We evaluated the proposed network using the publicly available IDLab Real-World dataset for human activity recognition. This dataset was utilized for training and testing the model, employing a 5-fold cross-validation approach. Based on extensive investigations, we found that ResNeXt achieved the highest accuracy rate of 98.32% and an F1-score of 87.90%.*
Keywords: Deep neural network, Aggregated residual transformation module, Human activity recognition, Deep learning, Smartwatch sensor

1. **Introduction.** Sensor-based Human Activity Recognition (S-HAR) uses sensors to identify people's actions and is an important area in ubiquitous computing. The key goal is to detect and analyze human behavior patterns by processing sensor data from smartwatches, smartphones, and wearables. These electronics collect data from diverse

DOI: 10.24507/icicel.19.03.343

individuals, and machine learning can categorize the signals [1]. S-HAR with handheld devices shows promise for healthcare by monitoring patients with various conditions. It can track treatment adherence and prevent problematic behaviors [2]. Beyond health applications, S-HAR has uses in gaming [3], human-robot interaction, automation [4], and sports [5, 6]. Deep learning has gained traction in many domains, including S-HAR [7, 8]. Prior research has used smartphone accelerometers effectively for S-HAR [9]. Modern smartwatches contain diverse sensors to collect movement data, like accelerometers, gyroscopes, magnetometers, Wi-Fi, Bluetooth, microphones, light sensors, and cell broadcast monitors. These provide insights into daily behaviors and movement analysis. Sensors like accelerometers, gyroscopes, magnetometers, heart rate monitors, and GPS facilitate context-aware identification, social communication, and coarse-grained positioning [10]. Motion sensors especially provide significant data to recognize and track physical movements like walking, standing, and jogging. Most current S-HAR studies use sensor networks to capture interaction data, followed by algorithms to classify actions [11]. However, despite promising results, most studies use controlled laboratory datasets. Real-world S-HAR remains challenging for reliable categorization.

This research focuses on HAR using smartwatch sensors to collect data in real-world scenarios. To achieve our objective, we have introduced a deep neural network called ResNeXt, which incorporates an aggregated residual transformation module to classify human actions accurately. The performance of the proposed network was evaluated using the publicly available IDLab Real-World dataset, designed explicitly for HAR. We conducted model training and testing using a 5-fold cross-validation approach. Experiments showed that the ResNeXt model outperformed other models on critical metrics. It had higher scores predicting correct outcomes and balancing precision versus recall. These results demonstrate its superiority for this application.

The paper is organized as follows to provide a clear structure. Section 2 offers an overview of relevant research. Section 3 delves into the details of the deep learning models utilized in this study. Section 4 presents the experimental findings. Finally, Section 5 concludes the paper by summarizing the findings and proposing potential areas for further research.

2. **Related Works.**

Smartwatches have become integral to our daily activities, offering enhanced computing power, versatile Internet connectivity, and a wide range of mobile applications. Moreover, affordable smartwatches are now equipped with various sensors, including accelerometers, GPS, and gyroscopes. These sensors enable human motion detection, making smartwatches and other intelligent devices valuable tools for tracking and analyzing physical activities [12, 13].

With the widespread availability of cognitive and computational capabilities in modern smartphones, researchers have begun exploring smartwatches as an alternative to wearable sensor technology for HAR [14]. Smartphones offer diverse sensors, including accelerometers and gyroscopes, and wireless connectivity features, making them valuable for activity tracking in smart homes [15]. Additionally, smartwatches possess powerful computing capabilities and are easy to use, making them a practical choice compared to other sensors found in smart home environments. By integrating inertial sensors like gyroscopes and accelerometers, smartwatches can accurately capture motion data for HAR purposes.

Recent studies have shown that the accelerometer data produced by publicly accessible smartwatches is of research-grade quality [16]. This highlights the potential of using smartwatches for data collection without the need for additional devices [17]. For instance, Michelin et al. [18] utilized an Inertia Measurement Unit (IMU) to capture participants' tri-axial accelerometer and gyroscope data. They developed a 1D convolutional neural network to detect facial touching gestures. In addition to their conventional applications,

IMUs have been used by researchers to identify specific facial activities, such as ingestion [19, 20] and tobacco consumption [21, 22].

HAR methods are primarily developed and evaluated using data collected in controlled environments. However, this approach restricts our understanding of the effectiveness of these methods in real-life scenarios, where sensor data can be incomplete and distorted, and human behavior may be unstructured. As a result, exploring and assessing HAR methods in more realistic settings is necessary to understand their efficacy in practical applications comprehensively.

3. **Methodology.** In this study, we utilize an S-HAR process flow consisting of five main procedures: data acquisition, data pre-processing, data segmentation, model architecture, and model fine-tuning. These procedures are illustrated in Figure 1, visually representing the workflow.

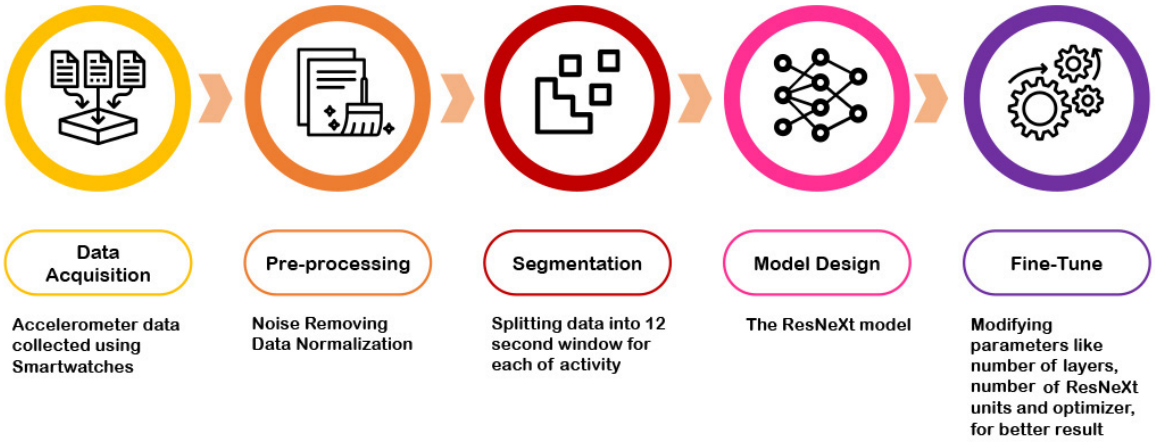

FIGURE 1. The HAR workflow based on smartwatch sensors used in this work

3.1. **Pre-processing and segmentation of data.** The pre-processing of the raw sensor data involved two essential steps: noise reduction and data standardization. To reduce noise in the signal, an average smoothing filter was applied to all three dimensions of the accelerometer sensor. After noise reduction, the sensor data was normalized to ensure all values fell within a comparable range. This standardization helps training models by making the data more consistent and improving the convergence rate of gradient descents. Finally, the normalized data was segmented using fixed-width sliding windows of 12 seconds, with a 50% overlap between consecutive windows. This segmentation process is illustrated in Figure 2.

3.2. **The proposed deep neural network with hyperparameters tuning.** The present study introduces ResNeXt, a deep neural network that incorporates aggregated residual transformations [23]. Unlike InceptionNet [24], which combines kernel feature maps of different sizes, ResNeXt combines them through addition. This approach reduces the number of parameters in the model, making it more suitable for edge and low-latency applications. Figure 3 visually represents the ResNeXt architecture and its components.

The ResNeXt model consists of four distinct components, each utilizing convolutional kernels of different sizes. One of these components is the MultiKernel (MK) component, which incorporates kernels of sizes 1×3, 1×5, and 1×7. Before applying these kernels, 1×1 convolutions are used to reduce the model's complexity and number of parameters. The details of the MK component are depicted in Figure 4.

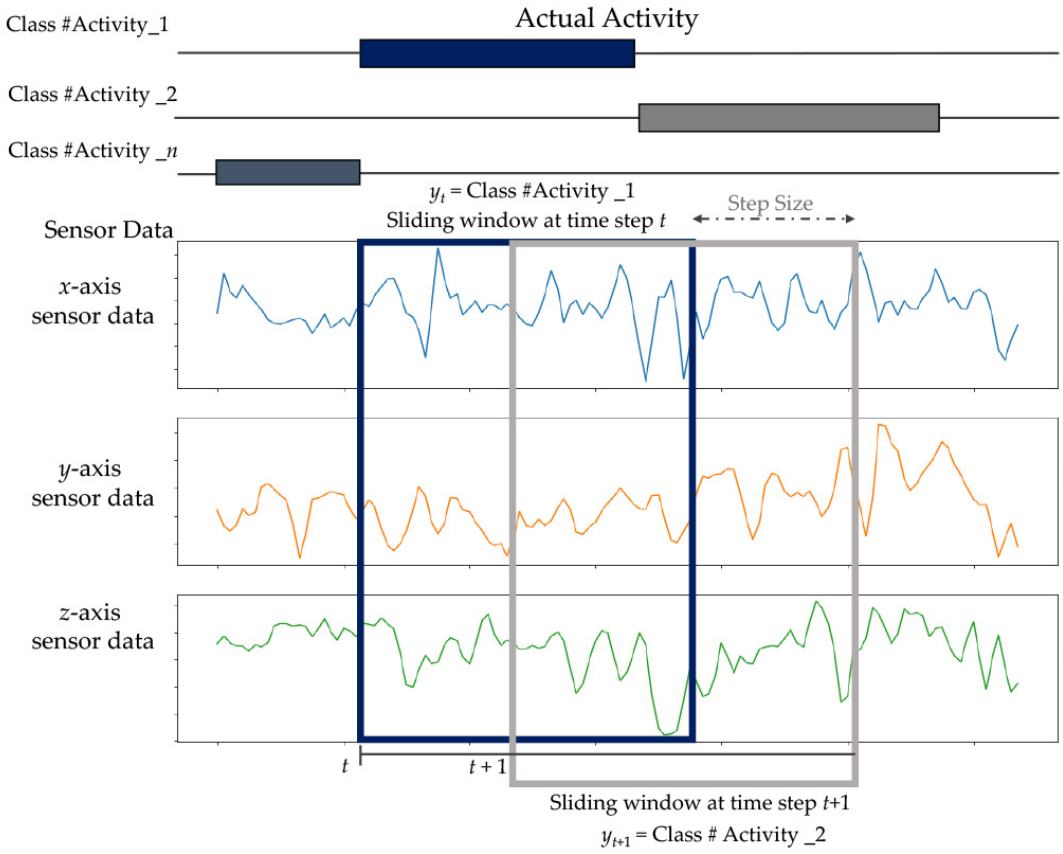

FIGURE 2. Data segmentation using a fixed-width sliding window

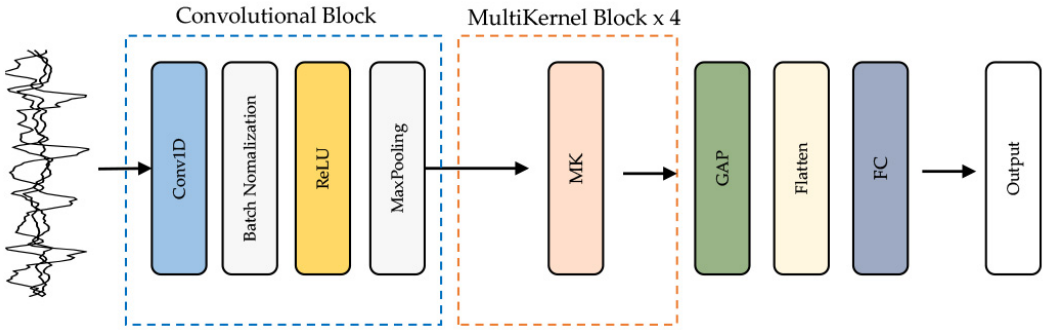

FIGURE 3. The proposed ResNeXt model

The model simplified the feature maps by using Global Average Pooling (GAP). It averaged the values in each map to produce a flattened 1D vector summary. Fully connected layers transformed this into probability estimates via Softmax. These scores reflect confidence in predicted classes. Cross-entropy loss was the error function, as is common in classification models. It quantifies inaccuracy between targets and predictions.

Hyperparameters are used in deep learning to control model training. This model employs several key hyperparameters: epochs, batch size, learning rate, optimization algorithm, and loss function (Table 1). A sample size of 128 and 200 epochs was set to determine suitable hyperparameters. Early stopping after 30 epochs without validation loss improvement concluded training. The initial learning rate was 0.001, reduced by 75% if validation accuracy stalled for seven epochs. The Adam optimizer ($\beta_1 = 0.90$, $\beta_1 = 0.999$, $\epsilon = 1 \times 10^8$) minimized errors. It uses categorical cross-entropy loss, improving this task's mean squared error and classification error.

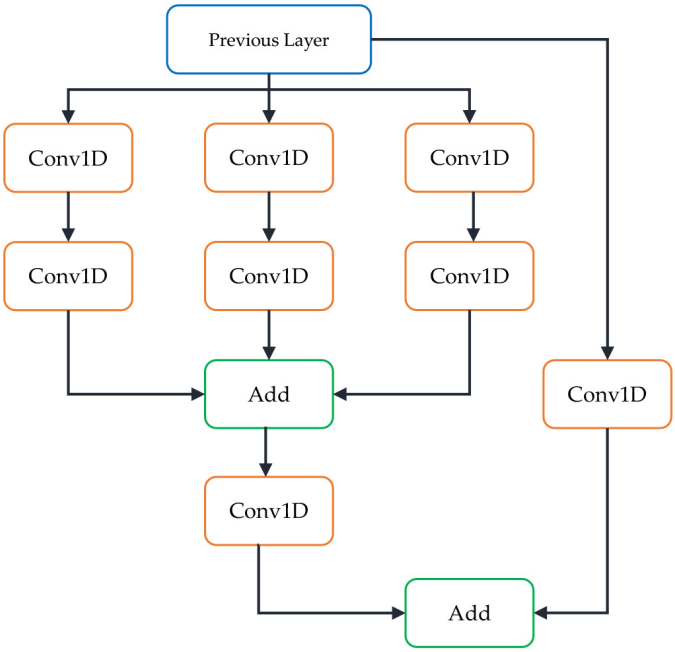

FIGURE 4. The structure of an MK module

TABLE 1. The summary of hyperparameters for the ResNeXt network used in this work

Stage	Hyperparameters		Values
Convolutional Block	Conv1D	Kernel Size	5
		Filters	64
	Activation		ReLU
	Max Pooling		2
Multi-Kernel Block × 3	Branch 1-1		
	Conv1D	Kernel Size	1
		Filters	16
	Conv1D	Kernel Size	3
		Filters	16
	Branch 1-2		
	Conv1D	Kernel Size	1
		Filters	16
	Conv1D	Kernel Size	5
		Filters	16
	Branch 1-3		
	Conv1D	Kernel Size	1
		Filters	16
	Conv1D	Kernel Size	7
		Filters	16
	Branch 1		
	Conv1D	Kernel Size	1
		Stride	1
		Filters	64
	Branch 2		
	Conv1D	Kernel Size	1
		Stride	1
		Filters	64
Classification Block	Global Average Pooling		–
	Flatten		–
	Dense		128
Training	Loss Function		Cross-entropy
	Optimizer		Adam
	Batch Size		128
	Number of Epochs		200

4. **Experimental Setting and Findings.** This section provides an overview of the experimental setup. It presents the results of evaluating five baseline deep learning models and the proposed ResNeXt model for HAR in real-world scenarios. The study considered five fundamental deep learning models: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Gated Recurrent Unit (GRU), and Bidirectional GRU (BiGRU).

4.1. **IDLab Real-World dataset.** The IDLab Real-World dataset [25] is a publicly available standard dataset widely used for HAR research. This dataset consists of sensor data collected from wearable devices, explicitly focusing on acceleration data with a sampling rate of 32 Hz. The data was gathered from a diverse group of eighteen subjects, ranging in age from 22 to 45, including thirteen men and five women. Participants were given the Empatica E4 wristband to wear. The E4 is a device that continuously gathers real-time data as participants live their daily lives. They were also asked to install a program that pairs with the wristband on their mobile phones. This allowed people to follow their routines and activities without restrictions while the wristband collected background data.

During the data-gathering process, participants were instructed to classify everyday actions in their daily lives. These actions included sitting while using a laptop, standing still upright, walking, jogging, and biking. Participants were also free to categorize additional tasks such as cooking, grocery shopping, transportation, or any personal hobbies they wanted to include in their actions.

The present study focused on the first five actions listed for several reasons. These actions were chosen because there was a substantial amount of available data related to them. They have been frequently studied in previous research, and we were confident in our ability to detect them using a single accelerometer worn on the wrist.

4.2. **Experimental setting.** All experiments in this research were conducted using the Google Colab Pro platform with a Tesla V100 GPU. The experiments were implemented in Python, utilizing various libraries, including Python 3.6.9, TensorFlow 2.2.0, Keras 2.3.1, Scikit-Learn, Numpy 1.18.5, and Pandas 1.0.5. The study evaluated the effectiveness of deep learning algorithms in HAR using smartwatch sensors from the IDLab Real-World dataset.

4.3. **Experimental results.** Table 2 presents the evaluation results of deep learning models using wearable sensor data for effectiveness identification. The findings of this study reveal that the suggested ResNeXt model displayed exceptional performance, achieving the highest accuracy among all models when applied to smartwatch sensor data. The proposed approach achieved an impressive accuracy of 98.32% and the highest F1-score of 87.90%.

Table 2. The identification effectiveness of the proposed ResNeXt model and five baseline models using smartwatch sensors for HAR

Model	**Parameter**	**Recognition performance**		
		Accuracy	**Loss**	**F1-score**
CNN	796,133	97.44%(±0.118%)	0.15(±0.004)	76.77%(±0.855%)
LSTM	183,805	96.31%(±1.291%)	0.13(±0.042)	72.23%(±5.873%)
BiLSTM	328,405	97.00%(±0.433%)	0.11(±0.005)	77.04%(±1.447%)
GRU	124,005	97.44%(±0.037%)	0.09(±0.005)	76.21%(±1.337%)
BiGRU	248,005	97.64%(±0.082%)	0.08(±0.001)	79.99%(±2.402%)
ResNeXt	**23,783**	**98.32%**(±0.044%)	**0.07**(±0.003)	**87.90%**(±0.875%)

The findings in Table 2 demonstrate that the suggested ResNeXt model outperforms the baseline model, showcasing its superior performance. Notably, the ResNeXt model achieves this impressive outcome while maintaining a relatively low number of parameters, as depicted in Figure 5.

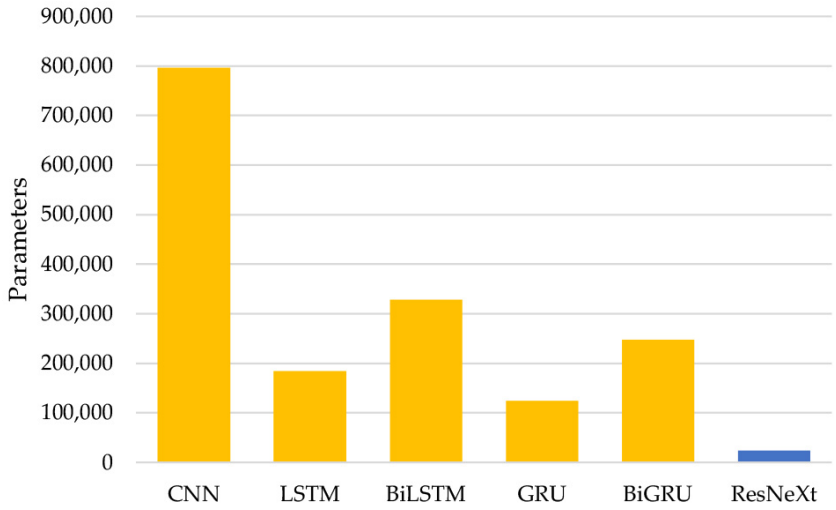

FIGURE 5. Comparative results of model parameters of each DL model used in this work

5. **Conclusion and Future Works.** This study focuses on utilizing deep learning models for HAR using smartwatches in real-world scenarios. We collected acceleration data from smartwatch sensors available in the IDLab Real-World dataset, which provides a wide range of sensor data capturing diverse human behaviors in real-life settings. To achieve our study's objective, we propose a deep neural network, ResNeXt, designed to enhance comprehension effectiveness. We compared ResNeXt with other baseline deep learning models, including CNN, LSTM, BiLSTM, GRU, and BiGRU. The experimental results demonstrate that the ResNeXt model outperforms the other models, achieving the highest accuracies and F1-scores. Specifically, when using acceleration data from smartwatch sensors, ResNeXt achieved an impressive accuracy of 98.32% and an F1-score of 87.90%.

In future endeavors, we plan to explore applying deep learning models, such as ResNet, InceptionTime, and Temporal Transformer, to improve human activity recognition in real-world settings. Additionally, we recognize the potential of data augmentation as a valuable technique to enhance model performance, mainly when dealing with imbalanced datasets. By implementing this methodology, we aim to address the issue above and further improve the accuracy and effectiveness of our models.

Acknowledgment. This project was funded by Thailand Science Research and Innovation Fund; and University of Phayao (Grant No. FF67-UoE-Sakorn).

REFERENCES

[1] S. K. Yadav, K. Tiwari, H. M. Pandey and S. A. Akbar, A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions, *Knowledge-Based Systems*, vol.223, 106970, 2021.

[2] A. Trifan, M. Oliveira and J. Oliveira, Passive sensing of health outcomes through smartphones: A systematic review of current solutions and possible limitations, *JMIR mHealth and uHealth*, vol.7, no.8, 2019.

[3] S. Spanogianopoulos, K. Sirlantzis, M. Mentzelopoulos and A. Protopsaltis, Human computer interaction using gestures for mobile devices and serious games: A review, *2014 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL2014)*, pp.310-314, 2014.
[4] A. Anagnostis, L. Benos, D. Tsaopoulos, A. Tagarakis, N. Tsolakis and D. Bochtis, Human activity recognition through recurrent neural networks for human-robot interaction in agriculture, *Applied Sciences*, vol.11, no.5, pp.1-20, 2021.
[5] D. O. Anderez, L. P. Dos Santos, A. Lotfi and S. W. Yahaya, Accelerometer-based hand gesture recognition for human-robot interaction, *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp.1402-1406, 2019.
[6] 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.
[7] Y. LeCun, Y. Bengio and G. Hinton, Deep learning, *Nature*, vol.521, pp.436-444, 2015.
[8] M. H. Ramadhan, T. N. Adi, H. Bae and H. Kim, Obtaining lightweight human activity recognition model through knowledge distillation of deep neural network, *ICIC Express Letters, Part B: Applications*, vol.13, no.5, pp.519-526, 2022.
[9] Z. Gao, H.-Z. Xuan, H. Zhang, S. Wan and K.-K. R. Choo, Adaptive fusion and category-level dictionary learning model for multiview human action recognition, *IEEE Internet of Things Journal*, vol.6, no.6, pp.9280-9293, 2019.
[10] W. Budiharto, Low cost prosthetic hand based on 3-lead muscle/electromyography sensor and 1 channel EEG, *ICIC Express Letters*, vol.13, no.1, pp.77-82, 2019.
[11] C. A. Ronao and S.-B. Cho, Human activity recognition with smartphone sensors using deep learning neural networks, *Expert Systems with Applications*, vol.59, pp.235-244, 2016.
[12] S. Mekruksavanich and A. Jitpattanakul, Recognition of real-life activities with smartphone sensors using deep learning approaches, *2021 IEEE 12th International Conference on Software Engineering and Service Science (ICSESS)*, pp.243-246, 2021.
[13] N. Hnoohom, A. Jitpattanakul and S. Mekruksavanich, Real-life human activity recognition with triaxial accelerometer data from smartphone using hybrid long short-term memory networks, *2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, pp.1-6, 2020.
[14] T. Gönül, O. D. Incel and G. I. Alptekin, Human activity recognition with smart watches using federated learning, in *Intelligent and Fuzzy Systems. INFUS 2022. Lecture Notes in Networks and Systems*, C. Kahraman,, A. C. Tolga, S. C. Onar et al. (eds.), Cham, Springer International Publishing, 2022.
[15] 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.
[16] A. Davoudi, A. A. Wanigatunga, M. Kheirkhahan, D. B. Corbett, T. Mendoza, M. Battula, S. Ranka, R. B. Fillingim, T. M. Manini and P. Rashidi, Accuracy of Samsung Gear S smartwatch for activity recognition: Validation study, *JMIR Mhealth Uhealth*, vol.7, no.2, e11270, 2019.
[17] Y. Vaizman, K. Ellis and G. Lanckriet, Recognizing detailed human context in the wild from smartphones and smartwatches, *IEEE Pervasive Computing*, vol.16, no.4, pp.62-74, 2017.
[18] A. M. Michelin, G. Korres, S. Ba'ara, H. Assadi, H. Alsuradi, R. R. Sayegh, A. Argyros and M. Eid, FaceGuard: A wearable system to avoid face touching, *Frontiers in Robotics and AI*, vol.8, 2021.
[19] X. Ye, G. Chen and Y. Cao, Automatic eating detection using head-mount and wrist-worn accelerometers, *2015 17th International Conference on E-Health Networking, Application & Services (HealthCom)*, pp.578-581, 2015.
[20] Y. Dong, J. Scisco, M. Wilson, E. Muth and A. Hoover, Detecting periods of eating during free-living by tracking wrist motion, *IEEE Journal of Biomedical and Health Informatics*, vol.18, no.4, pp.1253-1260, 2014.
[21] A. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan and E. Kalogerakis, RisQ: Recognizing smoking gestures with inertial sensors on a wristband, *Proc. of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, New York, NY, USA, pp.149-161, 2014.
[22] N. Hnoohom, S. Mekruksavanich and A. Jitpattanakul, An efficient resnetse architecture for smoking activity recognition from smartwatch, *Intelligent Automation & Soft Computing*, vol.35, no.1, pp.1245-1259, 2023.
[23] S. Xie, R. Girshick, P. Dollár, Z. Tu and K. He, Aggregated residual transformations for deep neural networks, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.5987-5995, 2017.

[24] H. Ismail Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P.-A. Muller and F. Petitjean, InceptionTime: Finding AlexNet for time series classification, *Data Min. Knowl. Discov.*, vol.34, no.6, pp.1936-1962, 2020.

[25] M. Stojchevska, M. De Brouwer, M. Courteaux, F. Ongenae and S. Van Hoecke, From lab to real world: Assessing the effectiveness of human activity recognition and optimization through personalization, *Sensors*, vol.23, no.10, 2023.