Improving Inertial Sensor-based Human Activity Recognition using Ensemble Deep Learning

Pornthep Rojanavasu[1], Ponnipa Jantawong[1], Anuchit Jitpattanakul[2] and Sakorn Mekruksavanich[1,*]
[1]*Department of Computer Engineering, School of Information and Communication Technology*
University of Phayao, Phayao, Thailand
pornthep.ro@up.ac.th, ponnipa.jantawong@gmail.com and sakorn.me@up.ac.th
[2]*Intelligent and Nonlinear Dynamic Innovations Research Center, Department of Mathematics*
Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
anuchit.j@sci.kmutnb.ac.th

***Abstract*—Sensor-based human activity recognition (S-HAR) is a famous study focusing on detecting human physiological actions by interpreting various sensors, especially one-dimensional time series information. Typically, S-HAR machine learning methods were developed using handcrafted characteristics. Unfortunately, this is a complicated process that involves feature engineering and a high level of domain knowledge. Due to the development of deep neural networks, classification techniques could efficiently handle relevant characteristics from raw sensor data, resulting in enhanced classification outcomes. In this study, we describe a unique method for S-HAR based on ensemble deep learning with sensor nodes connected to the waist, chest, leg, and arm. Implementing and training three deep learning networks is performed using a publically available dataset, including wearable sensors from eight human actions. The findings demonstrate that the proposed Ens-ResNeXt model provides the maximum accuracy and F1-score, which is superior to existing techniques.**

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

I. Introduction

Human activity recognition (HAR) explores the precise identification of regular tasks, including level strolling and stair climbing [1], [2]. Earlier, HAR has been used in intelligent healthcare, including locomotion and pattern detection in older people [3], [4] and sports and physical activity tracking [5], [6] to better understand life-related consequences [7].

The utilization of camera and radar-based technology in HAR situations is restricted by significant expense, privacy concerns, and computing demands [8], [9]. Alternately, low-cost and compact wearable inertial measuring instruments (IMUs: accelerometer and gyroscopes) allow researchers to analyze longitudinal movement data in supervised or free-living contexts effectively. Wearable IMUs (sometimes coupled with additional sensing modalities, such as magnetometer, electrocardiograph, and electromyography) with automated identification frameworks [10] enable very accurate HAR.

Initially, sensor-based human activity recognition (S-HAR) has been regarded as a multivariate time-series classification task [11]. Conventional machine learning techniques such as Naive Bayes, decision trees, and support vector machines have successfully categorized various human behaviors [12]. In contrast, handcrafted extraction of features requires topic skills or knowledge. Systematic feature extraction in a deep learning (DL) context has been realized using the paradigm of a deep model with convolutional layers [13]. In the initial stages of the DL-based HAR study, convolutional neural networks (CNNs) were employed to address sensor-based HAR by automatically extracting abstract properties from sensor data [14]. CNNs can acquire the spatial dimension of sensor data while offering acceptable interpretation for basic activities, but cannot obtain complicated activities requiring consideration of the temporal properties of wearable sensor data [15]. Developing various classifiers using deep learning techniques to classify complicated human actions with the increase in efficiency is a significant concern. Hence, in HAR [16], recurrent neural networks (RNNs) are used to provide temporal information from wearable sensor data meaning. The RNN, on the other hand, is challenging to train due to its disappearing or expanding gradient issue. Long short-term memory neural networks were designed to address this issue (LSTM). Several recent efforts in HAR have employed LSTMs to increase efficiency and effectiveness [17]. Subsequently, hybrid deep learning models have been created to solve the shortcomings of CNN and RNN neural networks.

Information fusion is an innovative strategy for improving the effectiveness of S-HAR based on constructing an ensemble of classifiers. However, it is unexplained mainly in this area. The ensemble structure is advantageous because it incorporates numerous models' judgments rather than relying on a single model. Many classification problems, including HAR, have lately benefited from employing a collection of distinct models. Mukherjee et al. [18] improved the overall performance of a model by employing different combinations of current DL classifiers, including majority voting, sum rule, and score fusion. Ensem-HAR is a newly proposed ensemble measurement-based DL-based model with four CNN and LSTM-based models for smartphone sensor-based HAR challenges [19].

In this study, we propose a unique method for S-HAR based on ensemble deep learning with sensor nodes placed on the waist, chest, leg, and arm. We examine three distinct deep learning networks (CNN, LSTM, and ResNet) to enhance the effectiveness of S-HAR using ensemble deep learning. The ensemble models were developed and trained using a public dataset, including sensor data for eight human activities.

II. Proposed Framework

In this part, we present the approach employed in our research to assess the significance of the ensemble learning method in S-HAR. Fig. 1 depicts a modified S-HAR procedure from the activity recognition chain (ARC). This process provides a summary of our approach to recognizing human activities. We started by pre-processing the dataset to eliminate noise and standardize data from distinct sources. Then, we segment the data using the windowing technique. A deep learning network is provided with the prepared sample data to generate a set of expected labels. The consequent collection of forecasted labels is then analyzed using conventional HAR measures (i.e., accuracy, precision, recall, and F1-score).

A. Smartphone and Supporting Nodes Dataset

Bernas et al. [20] gathered the Smartphone and Supporting Nodes (SSN) dataset to explore S-HAR utilizing sensor data from various sensor networks associated with various regions throughout the human body. Realme-8 5G smartphones were employed as IoT devices in the SSN dataset. The gadget was chosen as medium-level hardware that symbolizes systems with an expected cost. The smartphone's ARM Dimensity 700 Octa-core CPU and 6 GB of Memory enable it to analyze information and conduct identification rapidly. The gadget is equipped with accelerometer and gyroscope sensors for monitoring the individual's motion. With the hanging strap, the mobiles were fastened to several body areas (waist, chest, arm, and upper leg). Several 2-minute training courses were held in order to collect the sensor data.

The time series were sampled every ten milliseconds and delivered to a database. The action was constantly recorded throughout the specified time frame. The initial and final two percent of the time series were omitted from the analysis to exclude the action's preparation and termination phases. The actions under consideration include walking, running, squatting, leaping, laying, waving, seating, and upright.

B. Ens-ResNeXt Architectures

Relying on ensemble learning, we have proposed a deep residual neural network with aggregated multi-branch transformation to tackle the HAR issue. Ens-ResNeXt is the proposed DL model that automatically extracts features using convolutional layers and residual connections. In this model, the various-sized kernel feature maps are contributed instead of combined, such as InceptionNet [21]. This drastically decreased the number of model parameters, making such interconnections ideal for Edge and low-latency operations.

Ens-ResNeXt consists of four ResNeXt networks with three modules containing convolutional kernels of varying sizes. Each MultiKernel (MK) unit has three kernel sizes: 1×3, 1×5, and 1×7. Moreover, 1×1 convolutions are utilized before deploying these kernels to reduce the system's complexity and the number of parameters. Fig. 2 depicts the multi-kernel component's specifics.

The Ens-ResNeXt design requires 130,016 trainable parameters overall. The complete model consists of four MK units, followed by a 1×1 convolutional procedure that reduces the number of kernels to the appropriate number of classes. The structure of the Ens-ResNeXt prototype employed in this study is shown in Fig. 3.

III. Experiments and Results

This section addresses the laboratory requirements and findings to consider the deep learning models for inertial sensor-based HAR.

A. Experimentations

Every investigation in this study is conducted on the Google Colab-Pro platform with Tesla V100. Python is employed to develop the prototype and DL models with TensorFlow, Keras, Scikit-Learn, Numpy, and Pandas libraries. This research performed four trials to investigate the enhancement of HAR ensemble learning. Table I indicates that each investigation employed distinct categories of deep learning models.

TABLE I
A list of experiments used in this work

Experiment	**Type of Deep Learning Model**
I	Individual DL models
II	Ensemble DL models

B. Results

Each experiment's SSN dataset was employed to train deep learning models and evaluated by the 5-fold cross-validation procedure. As informed in Table II, the investigation revealed the recognition interpretation of the individual DL models and the ensemble DL models, including our proposed Ens-ResNeXt.

In Table II, the individual DL models were trained and tested using sensor data from different body positions. Experimental results showed that the proposed ResNeXt achieved the most satisfactory performance in every scenario.

In Table III, the ensemble deep learning networks were trained and tested operating sensor data gathered in the SSN dataset. From experimental results, it was uncovered that the proposed Ens-ResNeXt reached the most promising performance. Considering the number of trained parameters, the proposed Ens-ResNeXt contains the least number of parameters, with only 130,016 values.

IV. Conclusion and Future Works

This study concentrated on how to apply ensemble deep learning to inertial sensor-based HAR. We introduced the Ens-ResNeXt model to identify actions using inertial sensor data successfully. The SSN dataset was employed to train and assess DL models against the suggested ensemble model.

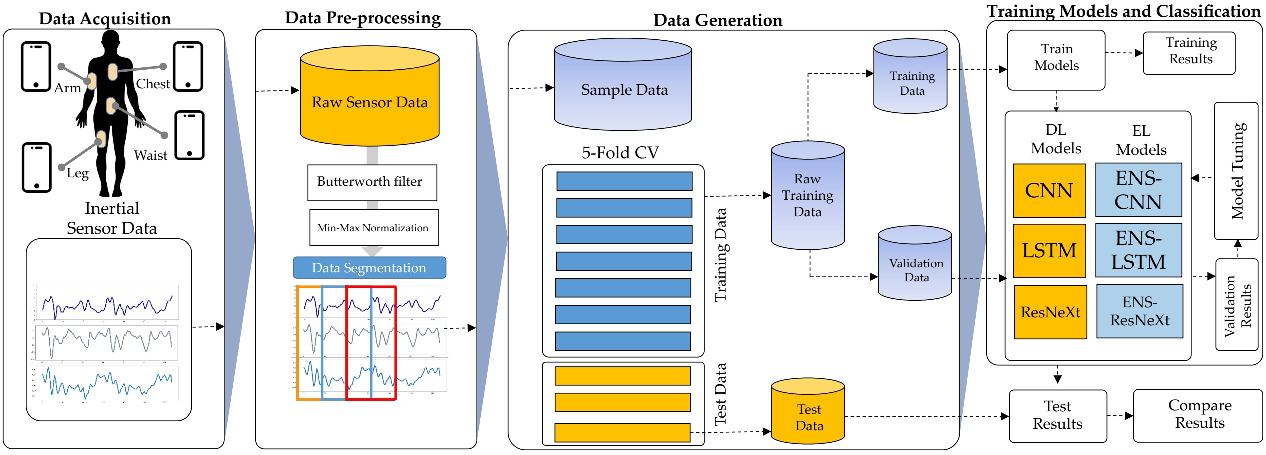

Fig. 1. S-HAR workflow used in this work.

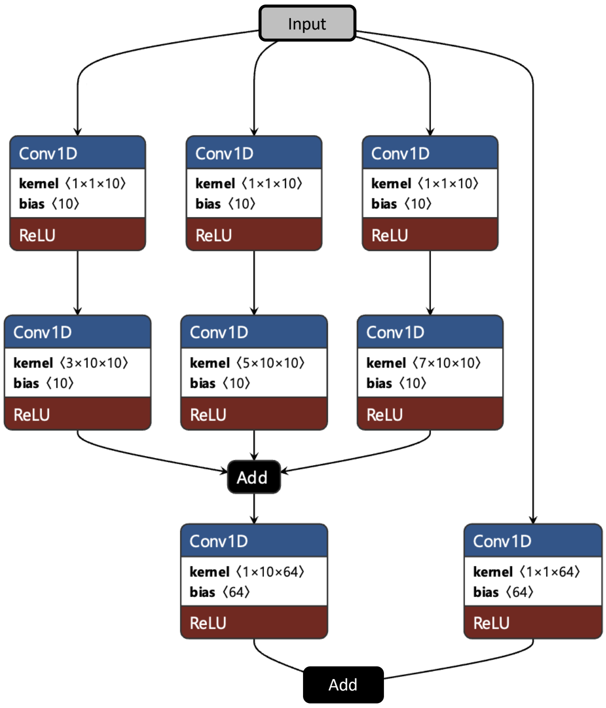

Fig. 2. The structure of a multi-kernel module.

Based on experimental findings, the suggested Ens-ResNeXt model surpassed existing DL models with 100% accuracy.

Future studies will investigate how various bio-signals (such as electroencephalogram (EEG) and electrooculogram (EOG)) affect the accuracy of action detection operating this ensemble deep learning technique.

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] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, "Sensing technology for human activity recognition: A comprehensive survey," *IEEE Access*, vol. 8, pp. 83 791–83 820, 2020.

[2] 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.

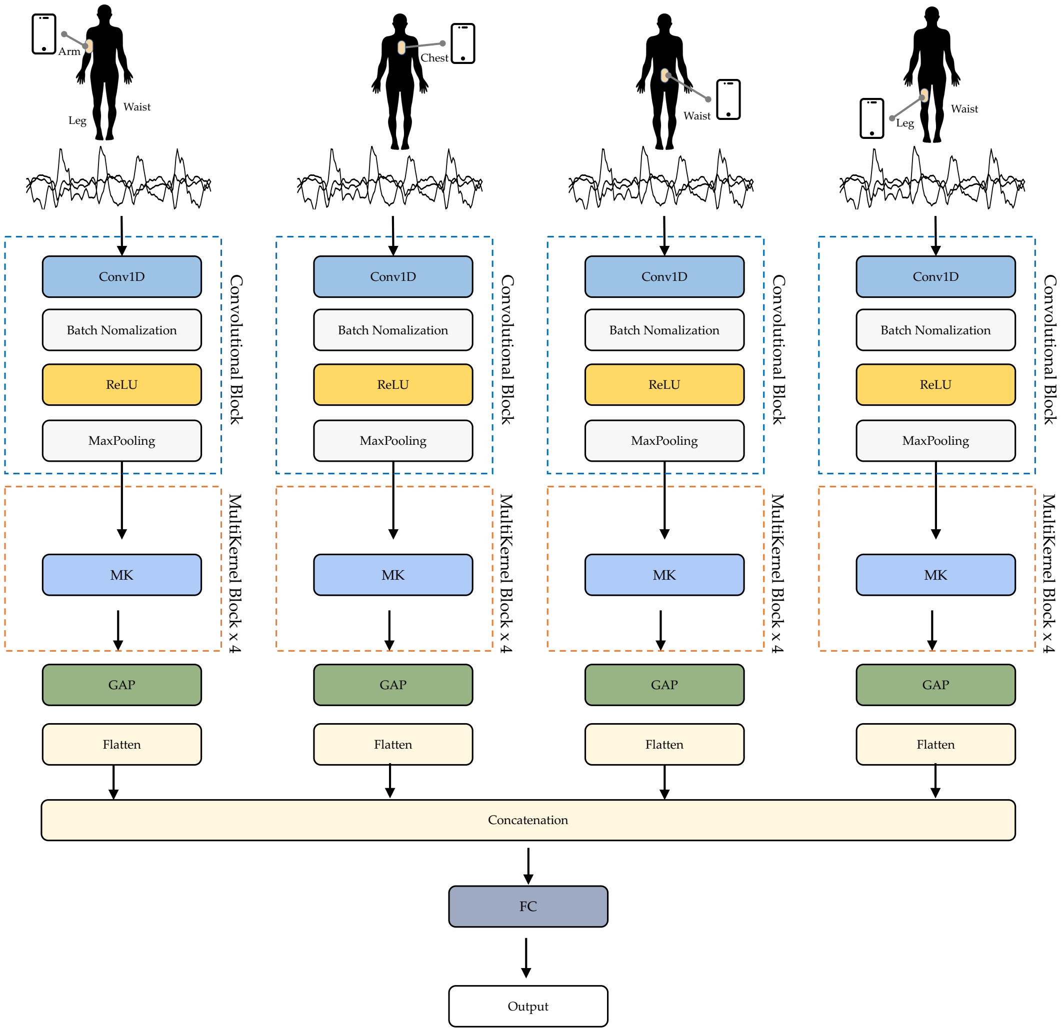

Fig. 3. The Architecture of the Ens-ResNeXt model.

[3] 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.

[4] S. Deep, X. Zheng, C. Karmakar, D. Yu, L. G. C. Hamey, and J. Jin, "A survey on anomalous behavior detection for elderly care using dense-sensing networks," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 352–370, 2020.

[5] 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.

[6] 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.

[7] S. Mekruksavanich, N. Hnoohom, and A. Jitpattanakul, "A hybrid deep residual network for efficient transitional activity recognition based on wearable sensors," *Applied Sciences*, vol. 12, no. 10, 2022.

[8] 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.

[9] S. Mekruksavanich and A. Jitpattanakul, "Deep residual network for smartwatch-based user identification through complex hand movements," *Sensors*, vol. 22, no. 8, 2022.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] A. Murad and J.-Y. Pyun, "Deep recurrent neural networks for human activity recognition," *Sensors*, vol. 17, no. 11, 2017.

TABLE II
PERFORMANCE METRICS OF THE INDIVIDUAL DL NETWORKS TRAINED AND EVALUATED USING SENSOR DATA COLLECTED FROM DIFFERENT BODY POSITION

Individual Model	**Parameters**	**Identification Effectiveness**		
		Accuracy	**Loss**	**F1-Score**
Arm				
CNN	235,208	95.59%	0.19	95.58%
LSTM	88,200	97.38%	0.09	97.38%
ResNeXt	24,542	96.81%	0.10	96.80%
Chest				
CNN	235,208	95.82%	0.17	95.79%
LSTM	88,200	96.65%	0.10	96.65%
ResNeXt	24,542	98.28%	0.05	98.28%
Waist				
CNN	235,208	95.00%	0.23	94.95%
LSTM	88,200	97.56%	0.10	97.55%
ResNeXt	24,542	98.18%	0.07	98.18%
Leg				
CNN	235,208	99.12%	0.03	99.12%
LSTM	88,200	99.18%	0.03	99.18%
ResNeXt	24,542	99.51%	0.02	99.51%

TABLE III
PERFORMANCE METRICS OF THE ENSEMBLE DL NETWORKS TRAINED AND EVALUATED USING SENSOR DATA

Ensemble Model	**Parameters**	**Identification Effectiveness**		
		Accuracy	**Loss**	**F1-Score**
Ens-CNN	937,352	100.00%	0.00	100.00%
Ens-LSTM	349,320	99.99%	0.00	99.99%
Ens-ResNeXt	130,016	100.00%	0.00	100.00%

[17] 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.

[18] D. Mukherjee, R. Mondal, P. K. Singh, R. Sarkar, and D. Bhattacharjee, "Ensemconvnet: A deep learning approach for human activity recognition using smartphone sensors for healthcare applications," *Multimedia Tools Appl.*, vol. 79, no. 41–42, p. 31663–31690, nov 2020.

[19] D. Bhattacharya, D. Sharma, W. Kim, M. F. Ijaz, and P. K. Singh, "Ensem-har: An ensemble deep learning model for smartphone sensor-based human activity recognition for measurement of elderly health monitoring," *Biosensors*, vol. 12, no. 6, 2022.

[20] M. Bernaś, B. Płaczek, and M. Lewandowski, "Ensemble of rnn classifiers for activity detection using a smartphone and supporting nodes," *Sensors*, vol. 22, no. 23, 2022.

[21] 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, p. 1936–1962, nov 2020.