








Multi-resolution CNN for Lower Limb Movement Recognition Based on Wearable Sensors

Narit Hnoohom¹ , Pitchaya Chotivatunyu¹ , Sakorn Mekruksavanich² ,
and Anuchit Jitpattanakul³  

¹ Image, Information and Intelligence Laboratory, Department of Computer Engineering,
Faculty of Engineering, Mahidol University, Nakorn Pathom, Thailand
narit.hno@mahidol.ac.th, pitchaya.cht@student.mahidol.ac.th

² Department of Computer Engineering, School of Information and Communication
Technology, University of Phayao, Phayao, Thailand
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. Human activity recognition (HAR) remains a difficult challenge in human-computer interaction (HCI). The Internet of Healthcare Things (IoHT) and other technologies are expected to be used primarily in conjunction with HAR to support healthcare and elder care. In HAR research, lower limb movement recognition is a challenging research topic that can be applied to the daily care of the elderly, fragile, and disabled. Due to recent advances in deep learning, high-level autonomous feature extraction has become feasible, which is used to increase HAR efficiency. Deep learning approaches have also been used for sensor-based HAR in various domains. This study presents a novel method that uses convolutional neural networks (CNNs) with different kernel dimensions, referred to as multi-resolution CNNs, to detect high-level features at various resolutions. A publicly available benchmark dataset called HARTH was used to evaluate the recognition performance to collect acceleration data of the lower limb movements of 22 participants. The experimental results show that the proposed approach improves the F1 score and achieves a higher score of 94.76%.

Keywords: Human activity recognition · Deep learning · Wearable sensor · Accelerometer · Multi-Resolution CNN

1 Introduction

Human motion analysis is a topic that receives much attention in robotics and medicine. Research on ambulatory activities is being conducted in rehabilitation science to improve the quality of life and context awareness in designing human-machine interfaces. For example, in [1], an intelligent system for elderly and disabled people is proposed where

the user can communicate with a robot via gesture recognition and recognition of everyday activities. These technologies help monitor the health status of patients and older people. In [2], a multi-sensor system is proposed to allow continuous rehabilitation monitoring. Diagnosing diseases such as multiple sclerosis, Parkinson's disease, and stroke [3] has been performed using human gait analysis.

Moreover, human gait has been utilized to develop indoor pedestrian navigation systems that can lead users to a specific area or track their daily activity level [4]. Multimodal systems have been designed for gait analysis for biometric applications [5]. Upper and lower limb motion analyses are also helpful for the development of prosthetic limbs for amputees [6].

Recognizing lower limb movements is essential for the daily care of the elderly, the weak, and the disabled. It is widely accepted that approaches for identifying lower limb movement can be divided into three types [7]: computer vision-based, ambient device-based, and wearable sensor-based. Computer vision-based types can monitor activities by analyzing video footage captured by cameras with multiple viewpoints placed at the desired location [8]. The implementation of computer vision-based technology is restricted by the space required to install the sensors [9]. The ambient device-based type provides for installing ambient sensors to measure the frequency of vibrations caused by regular activities for motion detection [10].

Nevertheless, activity monitoring can be severely affected by various environmental conditions. Aside from that, privacy concerns may arise with this approach [11]. The wearable sensor-based type uses multiple compact, wireless, and low-cost wearable sensor devices to record lower limb activity information [12]. The wearable sensor is suitable for outdoor use and compatible with the physical environment, and is primarily used for lower limb motion detection [13].

This work was motivated by the desire to develop and propose a method for recognizing lower limb movement that is highly accurate and capable of extracting useful information from inertial signals. To assess the multi-dimensional information included within the inertial signal, the multi-resolution convolutional neural network (M-CNN) was introduced to pull out high-level features and efficiently identify lower limb movements. The proposed model's performance in recognition is assessed with the help of training and testing data taken from a reference dataset known as HARTH, which is open to the public. Finally, the evaluated metrics are compared with three basic deep learning (DL) models.

The following structure can be seen throughout the remainder of this article's content: Sect. 2 presents recent related work on DL approaches for lower limb movement. Section 3 describes in detail the multi-resolution CNN model utilized in this study. Section 4 demonstrates our experimental results using a publicly available benchmark dataset. This section also contrasts the outcomes of the proposed model with those of the fundamental DL models. Section 5 concludes this work and identifies areas for potential future research.

2 Related Works

2.1 Types of Sensor Modalities

Even though many HAR techniques can be generalized to all sensor modalities, most are specialized and have a limited scope. Modalities can be divided into three categories: body-worn sensors, ambient sensors, and object sensors.

One of the most common HAR modalities is the use of body-worn sensors. Examples of body-worn sensors include gyroscopes, magnetometers, and accelerometers. These devices can collect information about human activity by analyzing angular velocity and acceleration variations. Several studies on DL for lower limb movements have used body-worn sensors; nevertheless, most studies have concentrated on the data gathered from accelerometers. Gyroscopes and magnetometers are commonly used in conjunction with accelerometers to detect lower limb movements [14]. Ambient sensors are often embedded in a user's smart environment and consist of sound sensors, pressure sensors, temperature sensors, and radar. They are commonly used in data collection to study people's interactions and environment. The movement of objects can be measured with various object sensors, while ambient sensors can detect changes in the surrounding environment. Several research papers have investigated ambient sensors for HAR in ADL and hand movements [15]. Some experiments have used accelerometers or sensors in combination with ambient sensors to optimize the HAR accuracy. This shows that adopting hybrid sensors that collect different data sets from other sources can considerably boost research in HAR and encourage applications such as commercial smart home systems [16].

2.2 Deep Learning Approaches

The challenges associated with feature extraction in conventional machine learning (ML) can potentially be solved by DL [17]. Figure 1 demonstrates how DL can improve HAR performance using different network configurations. The features are extracted, and the models are trained simultaneously in DL. The network can learn the features automatically instead of manually hand-crafted as in conventional ML approaches.

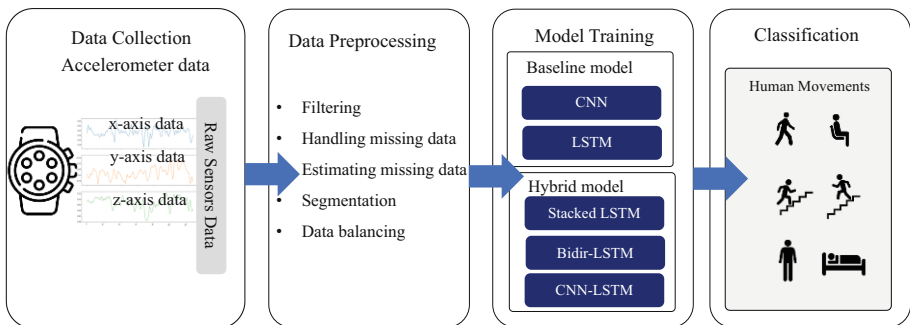


Fig. 1. DL-based-HAR pipeline.

3 The Sensor-Based HAR Framework

The sensor-based HAR framework consists of four main processes: (1) data acquisition, (2) data pre-processing, (3) data generation, and (4) training models and classification, as shown in Fig. 2.

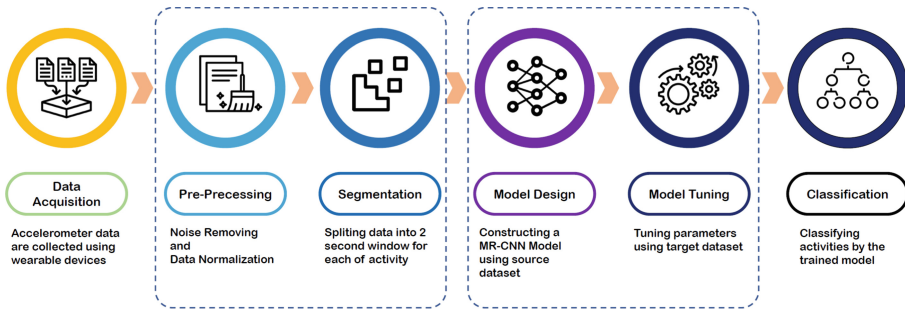


Fig. 2. The framework for HAR developed using sensors was used in this work.

3.1 HARTH Dataset

The human Activity Recognition Trondheim dataset, also known as HARTH, is available as a public dataset [18]. Twenty-two participants were recorded for 90 to 120 min during their regular working hours using two triaxial accelerometers attached to the lower back and thighs and a camera attached to the chest. Experts annotated the data independently using the camera's video signal. They labeled twelve activities. For the HARTH dataset, two triaxial Activity AX3 accelerometers [19] were used to collect data. The AX3 is a compact sensor that weighs only 11 g. Configurable parameters include sampling rate (between 12.5 and 3,200 Hz), measurement range ($\pm 2/4/8/16$ g), and resolution (which can be up to 13 bits).

A total of twelve different types of physical activities were recorded for the dataset throughout two sessions. In the first session, 15 participants (six women) were asked to perform their daily activities as usual for 1.5 to 2 h while being recorded. They were asked to complete each activity: sitting, standing, lying, walking, and running (including jogging) for a minimum of two to three minutes. For this time, the two sensors collected acceleration data at a sampling rate of 100 Hz (later reduced to 50 Hz) and a measurement range of ± 8 g. At the start of the recordings, each participant conducted three heel drops (i.e., dropped their heels firmly on the ground), which later assisted in synchronizing the acceleration and video signals. The duration of the first recording session was approximately 1,804 min (≈ 30 h). The average recording time was around 120 ± 21.6 min. After the recording was completed, videos were down-sampled to 640×360 pixels at a frame rate of 25 frames per second and annotated frame by frame. In addition to the five activities presented, participants performed other activities, which we labeled as follows: climbing Stairs (up), climbing Stairs (down), shuffling (standing with leg movement), cycling (standing), cycling (sitting), transportation (sitting) (e.g.,

in a car), and transportation (standing) (e.g., in a bus). This resulted in a total of twelve different designations.

3.2 Data Pre-processing

Raw sensor data were processed in the data preprocessing as follows: Removal of noise and normalization of the data. In this work, an average smoothing filter was applied to gyroscope and accelerometer sensors in all three dimensions to remove noise from the signals. Then, the sensor data is normalized, which helps to solve the model learning problem by bringing all data values into a similar range. As a result, the gradient descents can converge faster. Next, the normalized data were segmented using a sliding window with a fixed width of two seconds and a percentage overlap of 50%.

3.3 The Proposed Multi-resolution CNN Model

The multi-resolution technology CNN stands for a convolutional neural network with advanced features. It consists of filters with different kernel sizes, and these filters must be used in each layer to extract relevant information from the convolutional layers successfully. Nafea et al. [16] demonstrated encouraging HAR results with multi-resolution modules based on the inception modules provided by Szegedy et al. [20]. This inspired us to investigate them in more detail. Multiple kernel sizes are used, and the results of these kernel sizes are combined, as opposed to the standard CNN practice of using only a single kernel size in a single layer. The result is that a single layer is used to extract features from various scales. Figure 3 shows the proposed multi-resolution CNN.

3.4 Performance Measurement Criteria

Four standard evaluation metrics, e.g., accuracy, recall, and F1-score, are calculated using 5-fold cross-validation to evaluate the effectiveness of the suggested DL model. The mathematical formulas for the four metrics are given below:

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

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

These four metrics were used to quantify the effectiveness of HAR. The recognition was a true positive (TP) for the class under consideration and a true negative for all other classes (TN). Misclassified sensor data may result in a false positive (FP) recognition for the class under consideration. Sensor data that should belong to another class may be misclassified, resulting in a false negative (FN) recognition of that class.

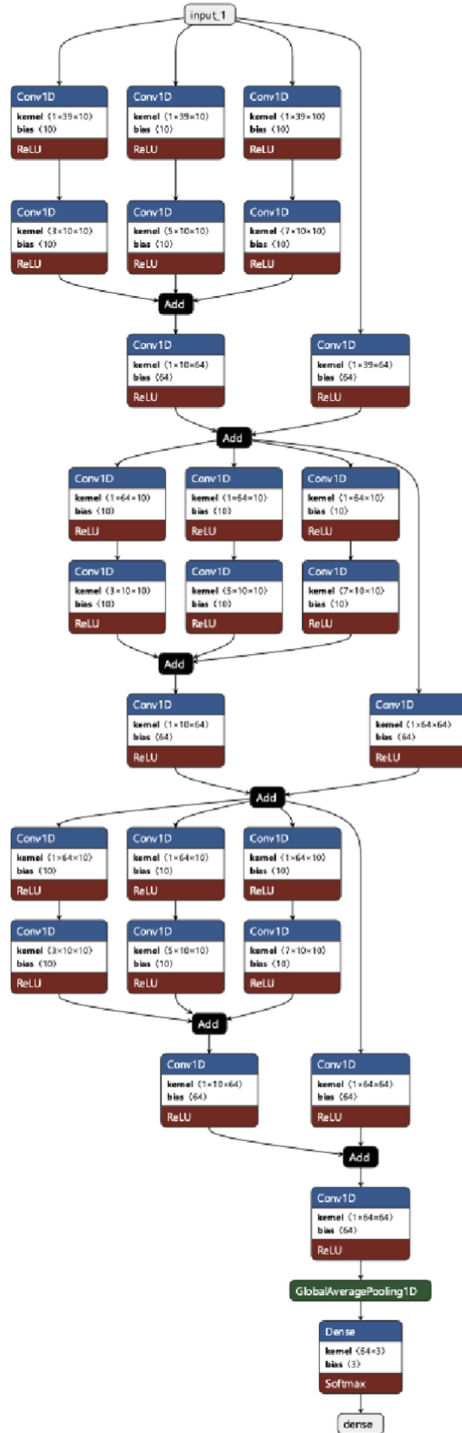


Fig. 3. A detailed description of the proposed multi-resolution CNN.

4 Experiments and Results

We have described the experimental setup and provided the experimental results to evaluate three basic DL models (CNN, LSTM, and CNN-LSTM), including the proposed multi-resolution CNN.

4.1 Experiments

All experiments were conducted on the Google Colab Pro with a Tesla V100. NumPy (NumPy 1.18.5) was used to work with matrices, Pandas (Pandas 1.0.5) was used to work with CSV files, and Scikit-Learn was used to evenly divide examples by class for the training, testing, and validation datasets. The Python programming (Python 3.6.9) and other libraries (Keras 2.3.1 and TensorFlow 2.2.0) were used to perform the experiments.

4.2 Experimental Results

The performance of DL models for recognizing data from wearable sensors is shown in Table 1. According to the experimental results, the proposed MR-CNN model had the highest performance, measured by an F1-score of 94.76%.

Table 1. Performance metrics of DL models using sensor.

Model	Performance		
	Accuracy	Loss	F1-score
CNN	95.98521% ($\pm 0.06931\%$)	0.54300 ($\pm 0.02758\%$)	88.31424% ($\pm 0.82847\%$)
LSTM	97.19258% ($\pm 0.22752\%$)	0.14847 ($\pm 0.02141\%$)	94.03455% ($\pm 0.58847\%$)
CNN-LSTM	97.45165% ($\pm 0.13398\%$)	0.15704 ($\pm 0.01128\%$)	93.99878% ($\pm 0.20887\%$)
MR-CNN	97.59014% ($\pm 0.19828\%$)	0.13416 ($\pm 0.02137\%$)	94.76309% ($\pm 0.55325\%$)

We considered classification results obtained from the MR-CNN as shown in Table 2. Regarding the activities of sitting in the HARTH dataset, the MR-CNN model achieved an F1-score of 1.00, as these activities do not involve movement. In contrast, F1-score values greater than 0.95 identify walking and running activities in the dataset.

Table 2. Performance metrics of DL models using sensor data of lower limb movement.

Activity	Performance			
	CNN	LSTM	CNN-LSTM	MR-CNN
Walking	0.94	0.95	0.95	0.95
Running	0.99	0.99	0.98	0.99
Ascending	0.59	0.89	0.87	0.80
Descending	0.71	0.89	0.84	0.89
Standing	0.92	0.94	0.95	0.94
Sitting	1.00	1.00	1.00	1.00
Lying	0.99	0.98	0.99	0.99
Cycling (sit)	0.96	0.98	0.98	0.98
Cycling (stand)	0.71	0.81	0.83	0.84
Average	0.87	0.94	0.93	0.94

5 Conclusions

This research proposed a new architecture using multiple convolutional layers with different kernel dimensions to achieve feature recognition with different resolutions. The proposed multi-resolution convolutional neural network (MR-CNN) model outperformed previous work in a public HARTH dataset that does not contain hand-crafted features. A comparison of the confusion matrices shows that the MR-CNN model achieved the highest performance of 94.76% in activity differentiation.

In our future work, we intend to use various types of DL networks, including ResNeXt, InceptionTime, Temporal Transformer, etc., in heterogeneous human activity recognition. Moreover, data augmentation is an exciting technique for model improvement in imbalanced datasets. This technique can be used for this problem.

Acknowledgments. The authors gratefully acknowledge the financial support provided by the Thammasat University Research fund under the TSRI, Contract No. TUFF19/2564 and TUFF24/2565, for the project of “AI Ready City Networking in RUN”, based on the RUN Digital Cluster collaboration scheme. This research project was supported by the Thailand Science Research and Innovation fund, the University of Phayao (Grant No. FF65-RIM041), and supported by National Science, Research and Innovation (NSRF), and King Mongkut’s University of Technology North Bangkok, Contract No. KMUTNB-FF-66-07.

References

1. Zhu, C., Sheng, W.: Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* **41**(3), 569–657 (2011)

2. González-Villanueva, L., Cagnoni, S., Ascari, L.: Design of a wearable sensing system for human motion monitoring in physical rehabilitation. *Sensors* **13**(6), 7735–7755 (2013). <https://doi.org/10.3390/s130607735>
3. Muro-de-la-Herran, A., Garcia-Zapirain, B., Mendez-Zorrilla, A.: Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors* **14**(2), 3362–3394 (2014). <https://doi.org/10.3390/s140203362>
4. Fourati, H.: heterogeneous data fusion algorithm for pedestrian navigation via foot-mounted inertial measurement unit and complementary filter. *IEEE Trans. Instrum. Meas.* **64**(1), 221–229 (2015)
5. Muaaz, M., Nickel, C.: Influence of different walking speeds and surfaces on accelerometer-based biometric gait recognition. In: 2012 35th International Conference on Telecommunications and Signal Processing (TSP), pp. 508–512. IEEE, Prague, Czech Republic (2012)
6. Gijsberts, A., Caputo, B.: Exploiting accelerometers to improve movement classification for prosthetics. In: 2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR), pp. 1–5. IEEE, Seattle, WA, USA (2013)
7. Mubashir, M., Shao, L., Seed, L.: A survey on fall detection: principles and approaches. *Neurocomputing* **100**, 144–152 (2013)
8. Casilari, E., Lora-Rivera, R., García-Lagos, F.: A study on the application of convolutional neural networks to fall detection evaluated with multiple public datasets. *Sensors* **20**(5), 1466 (2020). <https://doi.org/10.3390/s20051466>
9. Alves, J., Silva, J., Grifo, E., Resende, C., Sousa, I.: Wearable Embedded Intelligence for Detection of Falls Independently of on-Body Location. *Sensors* **19**(11), 2426 (2019). <https://doi.org/10.3390/s19112426>
10. Shah, S.A., Fioranelli, F.: RF sensing technologies for assisted daily living in healthcare: a comprehensive review. *IEEE Aerosp. Electron. Syst. Mag.* **34**(11), 26–44 (2019)
11. Shahzad, A., Kim, K.: FallDroid: an automated smart-phone-based fall detection system using multiple kernel learning. *IEEE Trans. Industr. Inf.* **15**(1), 35–44 (2018)
12. Yang, Y.K., et al.: Performance comparison of gesture recognition system based on different classifiers. *IEEE Trans. Cogn. Dev. Syst.* **13**(1), 141–150 (2021)
13. Xi, X., Tang, M., Miran, S.M., Luo, Z.: Evaluation of feature extraction and recognition for activity monitoring and fall detection based on wearable sEMG sensors. *Sensors* **17**(6), 1229 (2017). <https://doi.org/10.3390/s17061229>
14. Hussain, T., Maqbool, H.F., Iqbal, N., Mukhtaj Khan, N.A., Salman, A.A., Sanij, D.: Computational model for the recognition of lower limb movement using wearable gyroscope sensor. *Int. J. Sens. Netw.* **30**(1), 35 (2019). <https://doi.org/10.1504/IJSNET.2019.099230>
15. Wang, Y., Cang, S., Yu, H.: A survey on wearable sensor modality centred human activity recognition in health care. *Expert Syst. Appl.* **137**, 167–190 (2019)
16. Nafea, O., Abdul, W., Muhammad, G., Alsulaiman, M.: Sensor-based human activity recognition with spatio-temporal deep learning. *Sensors* **21**(6), 2141 (2021). <https://doi.org/10.3390/s21062141>
17. Baldominos, A., Cervantes, A., Saez, Y., Isasi, P.: A comparison of machine learning and deep learning techniques for activity recognition using mobile devices. *Sensors* **19**(3), 521 (2019). <https://doi.org/10.3390/s19030521>
18. Logacjov, A., Bach, K., Kongsvold, A., Bårdstu, H.B., Mork, P.J.: HARTH: a human activity recognition dataset for machine learning. *Sensors* **2021**, **21**, 7853 (2021)
19. Axivity Homepage. <https://axivity.com/lncs>. Last Accessed 8 May 2022
20. Szegedy, C., et al.: Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9 (2015)