Deep Learning Approaches for HAR of Daily Living Activities Using IMU Sensors in Smart Glasses

Sakorn Mekruksavanich[1], Ponnipa Jantawong[1] and Anuchit Jitpattanakul[2,*]
[1]*Department of Computer Engineering, School of Information and Communication Technology*
University of Phayao, Phayao, Thailand
sakorn.me@up.ac.th and ponnipa.jantawong@gmail.com
[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*—In the latest days, study into the development of intelligent technologies has proven valuable, contributing to attempts to improve the quality of human existence. Smart glass is one of the intelligent wearable devices that can be used for various purposes, including healthcare monitoring, fall detection, sleep tracking, and human activity recognition (HAR). Smartphones and smartwatches are the primary wearables utilized in sensor-based HAR to collect human motions for training recognition models based on physical movement. These wearable tools, nevertheless, are more intrusive than smart glasses. Using IMU sensor data acquired via smart glasses, we investigate deep learning algorithms for detecting people's activities of daily living (ADL). This work proposes a hybrid deep neural network that automatically extracts spatial-temporal information from raw data to enhance identification performance. We performed tests to evaluate deep learning models using a publically available benchmark dataset, UCA-EHAR, which included IMU sensor data from multiple ADL from smart eyewear. The recommended CNN-LSTM model achieved the best effectiveness with the highest F1-score of 93.20%, as determined by experimental findings.**

***Keywords*—smart glasses, deep learning, human activity recognition, IMU sensor**

I. Introduction

The most famous technological solutions include a smartwatch, smart glasses, bright jackets, and innovative gaming gadgets [1], [2]. Wearable intelligent technologies have recently extended into various fields, including the commercial, entertainment, and healthcare industries. This is because the task execution requires active data collecting, processing, and decision-making. Several practical applications have emerged based on intelligent wearables, including fall detection [3]–[5], sleep tracking [6], and human activity monitoring in smart home automation [7].

Human activity recognition (HAR) is a popular area of study motivated by the development of smart wearable tech-nology. HAR is tackled as a machine learning problem that forecasts the everyday tasks conducted by an individual em-ploying sensor data of varying modalities [8].

Vision-based and body-worn sensors are the most common types utilized for human action detection. Vision-based sensing depends on environment-mounted cameras to record a video feed of a person conducting everyday tasks [9]. Inertial measuring units (IMU) such as an accelerometer, gyroscope, and possibly others are used by body-worn sensors to track the individual's movements (magnetometer, barometer, etc.) [10]. There are several data gathering tools, some being more invasive than others. Examples include smartphones [11], wearables [12], and application-specific devices [13]. Wearable sensors on the body are less informative than cameras yet do not need any particular setup, making it easier to implement on self-sufficient products.

Smart glasses are less intrusive than other wearable devices such as specialist IMU devices or even smartphones, particularly for people who wear glasses often. Similar to smartphones, the recently developed smart glasses offer several functionalities. These characteristics primarily favor industrial activities, education programs, and medical assistance. Voice recording, text processing, video recording, location-based services, data transmission, augmented reality (AR), virtual reality (VR), and mixed reality (MR), among other features, are included. Earlier HAR research mainly concentrated on specific activities, such as driving [14]. Consequently, activities of daily living (ADL) detection based on sensor data from smart glasses is still a complex topic in HAR investigation.

Using IMU sensor data from smart glasses, this study examines the effect of HAR on everyday activities. In this study, the HAR is resolved using techniques from deep learning. We propose a CNN-LSTM hybrid deep learning model to achieve our objective that can automatically extract spatial-temporal characteristics from raw sensor data. We evaluate the recognition effectiveness of deep learning models applying UCA-EHAR dataset, a publicly available benchmark dataset, and assess model interpretation employing F1-scores and confusion matrices.

What follows is the rest of the paper's outline: In the second part, we look at several recent publications pertinent to this discussion. The proposed CNN-LSTM model is thoroughly explained in the following Section III. The results of our investigations can be seen in Section IV. The study concludes with a consideration of necessary future studies in Section V.

II. Related Works

A. Smart Glasses Dataset in HAR Research

Smart glasses are not yet widely used for human behavior identification. In [15], there have been previous efforts to develop a dataset for intelligent devices, especially smart glasses. This dataset collects data from several sensors using Jins MEME smart glasses, a smartphone, and a wristwatch. The smart glasses deliver information from an inbuilt IMU. Nevertheless, this data collection includes a few notable flaws. First, there was just one participant in the investigation. In addition, there is no clearly defined collection of tasks or methodology, making it impossible to assess or expand. In another dataset supplied by [14], this study used Jins MEME smart glasses to capture driving data from ten professional and ten novice participants. Unfortunately, this activity data cannot be utilized in other contexts, such as everyday tasks.

Efforts to use smart glasses to do action recognition have been made as recently as [16] (Google Glass XE-22). This research looks at how well a Support Vector Machine (SVM) can classify information from smartphones and smart glasses for four activities (biking, jogging, watching movies, and playing a video game). They have developed a system that can infer information from Android phones but not intelligent eyewear.

Furthermore, as stated in the beginning, each dataset will have unique properties based on the equipment used to collect it. The equipment and its location will significantly impact the acceleration angle (both gravitational and linear acceleration) and signal form for specific motions. Moreover, the sensitivity and sampling rate of the sensors themselves might vary. Consequently, utilizing an established dataset for a new device or application will provide unsatisfactory classification performance. For this reason, the researchers of [17] generated our dataset for Ellcie Healthy's smart glasses to utilize in daily activities.

B. Deep Learning Strategies for HAR

Deep learning networks evolved from the artificial neural network (ANN). Conventional ANN often has relatively few shallow hidden layers, while DNN has numerous deep hidden layers. DNN is better equipped to learn from massive data sets with additional layers, and DNN is often the dense layer of other deep learning models. After the convolution layers in a convolution neural network, for instance, dense layers are usually included [18].

A convolutional neural network (CNN) utilizes three essential concepts: sparse interactions, shared parameters, and equivariant representations. Pooling and fully-connected layers often follow convolution and accomplish classification or regression operations. CNN has produced impressive outcomes in image classification, voice recognition, and text analysis and is capable of extracting features from signals. CNN offers two benefits over other models when used in categorizing time series, such as the HAR: local dependence and scale invariance. Local dependence indicates that local signals in HAR are feasible to be linked, while scale invariance denotes that various speeds or frequencies are independent of scale.

Employing temporal connections between neurons, recurrent neural networks (RNN) are used extensively in voice recognition and natural language processing. Gradient descent is frequently used to integrate LSTM (long-short term memory) cells with RNN, where LSTM cells serve as memory blocks. Few works [19] employed RNN for HAR applications, where learning rate and resource usage are the primary considerations.

Multiple deep models have been combined to develop the hybrid model. Developing a hybrid model integrating CNN and LSTM has provided several good case studies of the two methods working together. [20] demonstrates that a convolutional neural network (CNN) with recurrent dense layers outperforms a CNN with dense layers. Similarly, the findings are shown in [21]. Whereas RNN could employ the temporal relationship, CNN can capture the spatial link. CNN and RNN integration may enhance the ability to differentiate activities with variable signal distributions and periods.

III. Sensor-based HAR Approach

Data acquisition, preprocessing, data generation, model training, and evaluation are the four foundational aspects of the sensor-based HAR design employed in this investigation (see Fig. 1).

A. UCA-EHAR Dataset

The UCA-EHAR dataset [17] is unique for recognizing human activities employing smart glasses. The dataset describes the absence of data from smart glasses that may be used for HAR purposes. Twenty volunteers (12 males and 8 females) engaged in eight distinct everyday tasks while wearing smart glasses integrated with an accelerometer, a gyroscope, and a barometer to capture sensor data.

Static actions include standing, sitting, and lying, in which the person remains in the same posture for a set period. Walking, descending stairs, ascending steps, and running are dynamic mobility-related behaviors. The running exercise is more comparable to brisk walking than sprinting. Having a drink is a deliberately specified action since we think dehydration to be a problem for the elders. The drinking action is accomplished by sipping from a glass or bottle.

In this data collection, activity data is captured by the Ellcie Healthy (EH) smart-connected glasses. Ellcie Healthy smart connected glasses include oculography infrared proximity sensors hidden behind the temples. A barometer, temperature sensor, tri-axial accelerometer, and gyroscope are all integrated into the frame's temples. The inertial measurement unit's barometer and temperature sensors are contained in a

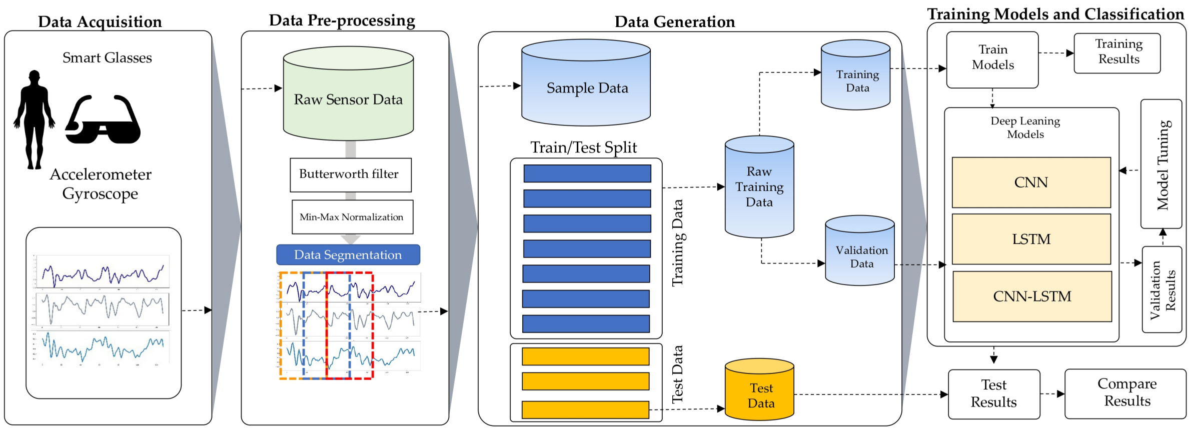

Fig. 1. The HAR approach based on IMU sensors in this study.

different component from the accelerometer and gyroscope. The accelerometer supplies the orthogonal coordinates and each component of the three-dimensional acceleration vector. The accelerometer and gyroscope on the Ellcie Healthy glasses gather data at 26 Hz, while the barometer records data at 6.66 Hz for this dataset.

B. Data Pre-processing

Unprocessed sensor data underwent noise reduction and data normalization during data pre-processing. We used 10-second fixed-width sliding windows with a 50% overlap proportion to partition the analyzed sensor data.

C. The CNN-LSTM model

This research proposes a 2-layer CNN-LSTM hybrid LSTM network to enhance detection capabilities. CNN-LSTM network consists of two convolutional layers and one LSTM layer. 260×6 is the size of sensor input data. The output layer has an 18×1 size. The CNN-LSTM architecture is shown in Fig. 2.

D. Bayesian Optimization for Hyperparameter Tuning

Bayesian optimization [22] was employed in this study to establish the CNN-LSTM network's parameters autonomously. This technique is efficient for solving functions whose extrema determination is operationally demanding. It may be used to solve functions that lack an expression in closed form. It may also be applied to functions whose calculus is costly, whose derivatives are difficult to assess, or whose shape is not convex. The optimization objective is to locate the most significant value at the sampling point for an undetermined function f:

$$x^{+} = \arg\min_{x \in A} f(x) \tag{1}$$

where A denotes the search space of x.

IV. Experimental Results

Utilizing the IMU sensor generated by smart glasses, we analyze the proposed CNN-LSTM model for HAR in this part.

A. Experimental Settings

Each investigation in this research is carried out on the Google Colab-Pro platform with a Tesla-V100 GPU. In addition to related libraries, the Python programming language is also developed using TensorFlow, Keras, Scikit-Learn, and Numpy libraries.

B. Exploratory Findings

To determine the effectiveness of HAR based on IMU sensors from smart glasses, we assessed the recognition of the conventional CNN, LSTM, and the proposed CNN-LSTM with Bayesian optimization of its hyperparameters using a series of methodologies. Multiple metrics, including accuracy, precision, recall, and F1-score, were used to evaluate the detection capabilities of DL networks.

The proposed CNN-LSTM was optimized by Bayesian optimization through Table I to identify a set of hyperparameters yielding a high-performance measure. This CNN-LSTM architecture consists of two convolutional layers and one LSTM layer. The size of sensor input data is 260×6. The initial convolutional layer has 113 distinct 3×1 filters. The second convolutional layer employs 123 different 3×1 filters with a dropout rate of 0.064, whereas the maximum pooling layer employs 2×1 filters. In Table II, the hyperparameters are described.

Table I demonstrates that the suggested CNN-LSTM networks surpass all existing networks with a 95.11% accuracy and a 93.20% F1-score. The CNN-LSTM model performs better than the standard DL models. The confusion matrix for CNN-LSTM networks is illustrated in Fig. 3.

V. Conclusion and Future Studies

This study developed a CNN-LSTM network to handle sensor-based HAR using IMU sensors from smart glasses. This hybrid model uses both the spatial and temporal feature extraction characteristics of CNN and LSTM. We also utilized Bayesian optimization to establish the optimal model hyperparameters to increase HAR's efficacy. We assessed the effectiveness of our CNN-LSTM network using several indicators

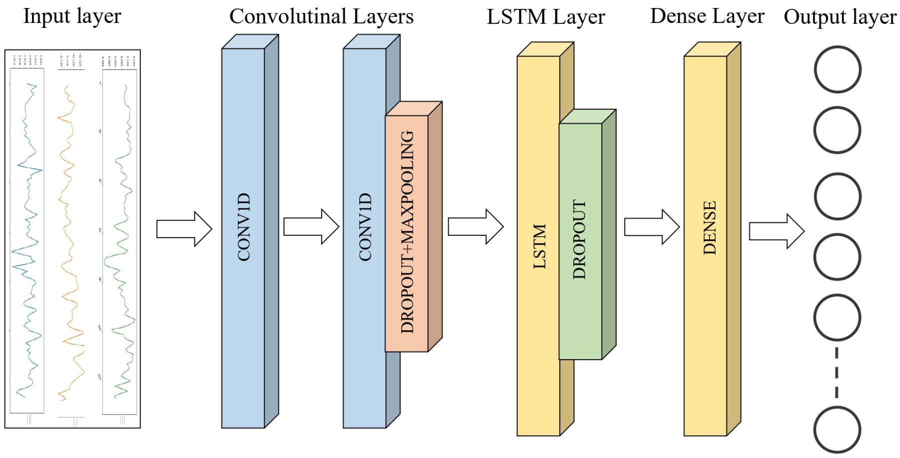

Fig. 2. The CNN-LSTM architecture used in this work.

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

Model	**Recognition Effectiveness**		
	Accuracy	**Loss**	**F1-score**
CNN	85.76% (±0.503%)	2.76% (±0.091%)	82.23% (±0.546%)
LSTM	83.45% (±0.708%)	0.52% (±0.028%)	78.69% (±1.756%)
Proposed CNN-LSTM	95.11% (±0.435%)	0.17% (±0.020%)	93.20% (±0.218%)

TABLE II
HYPERPARAMETER DETAILS OF THE PROPOSED CNN-LSTM NETWORKS IN THIS STUDY.

Stage	**Hyperparameters**		**Values**
Architecture	Convolution-1	Kernel Size	3
		Stride	1
		Filters	113
	Convolution-2	Kernel Size	3
		Stride	1
		Filters	123
	Dropout-1		0.06480703
	Maxpooling		2
	LSTM-neuron		128
	Dropout-2		0.21129224
	Dense		458
Training	Optimizer		Adam
	Batch Size		64
	Learning Rate		0.00049271
	Number of Epoches		50

and the UCA-EHAR public dataset. The results demonstrate that the proposed hybrid deep learning network outperforms conventional networks by automatically extracting spatial-temporal characteristics from raw sensor data with an average accuracy of 95.11% and an F-measure of 93.20%.

In future works, we will employ a transfer learning technique based on sensor data from smart glasses to improve this model for customized human action detection.

ACKNOWLEDGEMENTS

This research is supported by School of Information and Communication Technology, University of Phayao (Grant No. RMICT66005); National Science, Research and Innovation Fund (NSRF); and King Mongkut's University of Technology North Bangkok with Contract no. KMUTNB-FF-66-07.

REFERENCES

[1] M. Chan, D. Estève, J.-Y. Fourniols, C. Escriba, and E. Campo, "Smart wearable systems: Current status and future challenges," *Artificial Intelligence in Medicine*, vol. 56, no. 3, pp. 137–156, 2012.

Fig. 3. A confusion matrix of the proposed CNN-LSTM network.

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

[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] F. Hussain, F. Hussain, M. Ehatisham-ul Haq, and M. A. Azam, “Activity-aware fall detection and recognition based on wearable sensors,” *IEEE Sensors Journal*, vol. 19, no. 12, pp. 4528–4536, 2019.

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

[6] A. Zhao, J. Dong, and H. Zhou, “Self-supervised learning from multi-sensor data for sleep recognition,” *IEEE Access*, vol. 8, pp. 93 907–93 921, 2020.

[7] 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.

[8] 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.

[9] D. R. Beddiar, B. Nini, M. Sabokrou, and A. Hadid, “Vision-based human activity recognition: A survey,” *Multimedia Tools Appl.*, vol. 79, no. 41–42, p. 30509–30555, nov 2020.

[10] 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.

[11] D. Anguita, A. Ghio, L. Oneto, F. Parra, and J. Reyes-Ortiz, “A public domain dataset for human activity recognition using smartphones,” in *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 01 2013, pp. 437–442.

[12] G. M. Weiss, K. Yoneda, and T. Hayajneh, “Smartphone and smartwatch-based biometrics using activities of daily living,” *IEEE Access*, vol. 7, pp. 133 190–133 202, 2019.

[13] A. Reiss and D. Stricker, “Introducing a new benchmarked dataset for activity monitoring,” in *2012 16th International Symposium on Wearable Computers*, 2012, pp. 108–109.

[14] R. J. Doniec, S. Sieciński, K. M. Duraj, N. J. Piaseczna, K. Mocny-Pachońska, and E. J. Tkacz, “Recognition of drivers’ activity based on 1d convolutional neural network,” *Electronics*, vol. 9, no. 12, 2020.

[15] J. Meyer, A. Frank, T. Schlebusch, and E. Kasneci, “A cnn-based human activity recognition system combining a laser feedback interferometry eye movement sensor and an imu for context-aware smart glasses,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 5, no. 4, dec 2022.

[16] J. Ho and C.-M. Wang, “User-centric and real-time activity recognition using smart glasses,” in *Green, Pervasive, and Cloud Computing*, X. Huang, Y. Xiang, and K.-C. Li, Eds. Cham: Springer International Publishing, 2016, pp. 196–210.

[17] P.-E. Novac, A. Pegatoquet, B. Miramond, and C. Caquineau, “Uca-ehar: A dataset for human activity recognition with embedded ai on smart glasses,” *Applied Sciences*, vol. 12, no. 8, 2022.

[18] 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.

[19] S. Mekruksavanich and A. Jitpattanakul, “Deep learning approaches for continuous authentication based on activity patterns using mobile sensing,” *Sensors*, vol. 21, no. 22, 2021.

[20] S. Mekruksavanich, C. Promsakon, and A. Jitpattanakul, “Location-based daily human activity recognition using hybrid deep learning network,” in *2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2021, pp. 1–5.

[21] S. Mekruksavanich and A. Jitpattanakul, “Sensor-based complex human activity recognition from smartwatch data using hybrid deep learning network,” in *2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, 2021, pp. 1–4.

[22] J. Snoek, H. Larochelle, and R. P. Adams, “Practical bayesian optimization of machine learning algorithms,” in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012.