Sensor Fusion-Based Deep Learning Models for Human Activity Classification

Parshuram N. Aarotale[1] and Ajita Rattani[2]
[1] Wichita State University, Kansas, USA
[2] University of North Texas at Denton, Texas, USA

Abstract

Wearable sensors have been widely deployed for human activity recognition (HAR) in various sectors, including health monitoring, medical treatment, and motion analysis. Deep-learning-based HAR models have obtained enhanced accuracy rates over traditional machine learning, but there is still a gap in the acceptable recognition accuracy for HAR models. In this regard, this paper aims to propose sensor-fusion-based deep learning models for human activity classification for enhanced accuracy. Experimental validations are conducted on wearable sensor datasets such as ScientISST MOVE, containing various bio-signals, collected during daily human activities using chest, forearm, and wrist sensors. Experimental results show that our proposed sensor fusion-based deep learning model that systematically fuses signals from different sensors at one of the intermediate layers of the model, obtained enhanced performance in HAR. On average, our proposed sensor fusion-based models obtained an increment in HAR accuracy of about 9.14%*, Precision of* 7.30%*, Recall of* 8.70%*, and F1 score of* 9.0% *for all deep learning models based on single sensor data (such as wrist, chest, and forearm individually) for HAR.*

1. Introduction

Human activity recognition (HAR) is commonly used in applications including human-computer interaction [1, 2], smart-homes [3], health monitoring [4], environment-assisted living [5], and many more.HAR is typically a pattern recognition problem, and the complete pipeline consists of four major steps, ie, data collection, data preprocessing (filtering and segmentation), feature extraction, and recognition of human activity [6]. This HAR is classified into two types i.e., video-based and sensor-based. Video-based methods attempt to deduce human activity using videos captured from the deployed surveillance system in real time [7]. Recently, sensor devices have been increasingly used in HAR research due to their portability, low power consumption, and lower risk of privacy leakage. Sensors are typically placed on the arms, chest, wrist, and legs [8].

There has been significant research on sensor-based HAR using multi-modal time-series data via traditional machine-learning and deep-learning models [9, 10]. The former used hand-made features along with shallow machine learning algorithms such as K-nearest neighbor (KNN) [11], support vector machines (SVM) [12], and decision trees [13] to classify various human activities. The latter uses deep-learning models ([14–16]), such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have been used for HAR. Figure 1 shows the overview of the human activity recognition (HAR) pipeline involving biosignal extraction using various wearable sensors, and data (signal) pre-processing followed by feeding the processed signals to the deep neural network for HAR. Although most of the HAR datasets (such as PPG-Dalia [17]) capture multi-modal time series data from multiple wearable sensors, they have been processed individually or combined before being fed to deep learning models [14]. Thus, the true potential of complementary information from multimodal time series data captured from multiple sensors has not been utilized in existing HAR models.To advance the state-of-the-art in human activity recognition (HAR) using wearable sensors, this paper **aims** to propose sensor-fusion-based deep-learning models for human activity recognition.

Our Contributions: The contributions of this paper are as follows:

- Proposal of sensor fusion-based deep-learning models for human activity classification.
- Cross-comparison with the deep-learning-based HAR models based on single sensor data.
- Experimental investigation on the publicly available dataset namely, ScientISST MOVE [18], consisting of multi-modality time series data from wrist, chest, and forearm sensors.

2. Related work

Traditional HAR systems have collected features from sensor inputs and mapped them to various human activities. For the recognition of activities, supervised machine learning models have been used such as decision

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.418

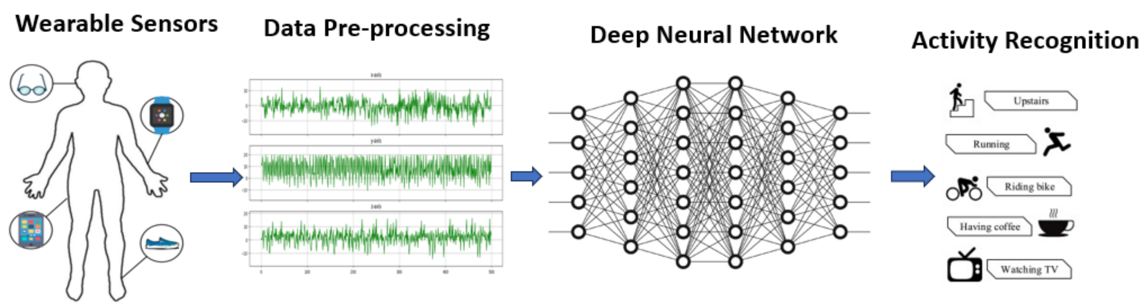

Figure 1. Overview of Deep Learning based Human Activity Recognition Pipeline.

trees,[19], k-nearest neighbor (KNN) [11], and support vector machine (SVM) [20] have been used.

Vepakomma et al. [21] extracted mean, and variance features from sensors (accelerometer and gyroscope) and then fed them into a deep neural network and obtained a precision of 90% for HAR. Zeng et al. [15] used CNN for HAR. Furthermore, Yang et al. [14] used 1D CNN based on weight sharing among different sensors.Tang et al. [22] proposed a lightweight CNN for HAR that used Lego filters to reduce memory and computation costs. Shi et al. [23] proposed a multichannel CNN that used appropriate convolutional kernels to generate deep features and obtained a significant performance enhancement. Almaslukh et al. [24] used stacked auto-encoders (SAE) for HAR along with greedy layer-wise pertaining. The authors used an in-house dataset with accelerometer and gyroscope signals and obtained an accuracy of 97.5%.

Inoue et al. [25] introduced a technique to recognize human activities from accelerometer data that employs an Deep Recurrent Neural Network (DRNN) and obtained a recognition rate of 95.42% for single activities and 83.43% for sequential activities, outperforming traditional methods. Uddin and Torresen [26] proposed an RNN-based inertial sensor technique capable of recognizing 12 human activities. Wang and Liu [27] used a hierarchical deep LSTM to classify HAR. Mekruksavanich and Jitpattanakul [28] developed a four-layer CNN-LSTM model on the UCI-HAR dataset to improve HAR performance. Li and Wang [29] developed a deep learning model using residual blocks and bidirectional LSTM (Bi-LSTM) that significantly improved HAR performance.The **aforementioned** existing studies ([14, 22–29]) obtained an average HAR accuracy of about 90.73% when evaluated on different HAR datasets.

3. Models and Training Hyper-parameters

This study uses deep-learning models such as LSTM, BI-LSTM, 1D CNN-based, 1D CNN-LSTM based on single and multi-sensor data fusion.

Hyperparameters: All the models were trained using cross-entropy loss function with class weight to address potential class imbalance. Adam optimizer is used for weight optimization, a learning rate of 0.001, a batch size of 32, and the number of epochs were determined using the early stopping mechanism. These hyper-parameters were selected using a grid-search approach.

4. Experimental Validation

4.1. Dataset: ScientISST MOVE Dataset

The ScientISST MOVE dataset [18] contains biosignals data from 17 healthy participants recorded during activities such as lifting a chair, greeting, gesticulating, walking, and jogging. Using three wearables—the ScientISST-Chest, ScientISST-Forearm, and Empatica E4 wristband—the dataset captures dual-view ECG, EDA, PPG, biceps EMG, wrist temperature, and actigraphy (ACC). The ScientISST devices sampled signals at 500 Hz, while the Empatica E4 recorded EDA (4 Hz), PPG (64 Hz), temperature (4 Hz), and ACC (32 Hz). The study was approved by the Ethics Committee of Instituto Superior Técnico, Lisbon (ref: 22/2022).

4.2. Data Preprocessing and Segmentation

ECG signals are preprocessed using a band-pass filter with a frequency of $1, 40$ Hz; Acc signals were passed via a low-pass filter with a cutoff frequency of 10 Hz; EDA was preprocessed with a low-pass filter at 1 Hz. The PPG signal was bandpass filtered with freq of 0.5 and 5 Hz. The preprocessed signals were resampled at 100 Hz. The preprocessed sensor signals were initially normalized using z-score normalization and data segmentation with 1 sec window along with 50% overlap was used.

5. Results and Discussion

Performance metrics such as accuracy, precision, recall, and F1-score were used for evaluation of the proposed models in this study; Table 1 shows the HAR per-

Table 1. HAR of deep learning models on wrist, chest, and forearm sensor data from the ScientISST MOVE dataset.

Models	Wrist Sensor Data				Chest Sensor Data				Forearm Sensor Data			
	ACC	P	R	F1	ACC	P	R	F1	ACC	P	R	F1
LSTM	93.8	0.95	0.94	0.94	**96.07**	0.96	0.96	0.96	75.7	0.57	0.76	0.65
BiLSTM	95.47	0.96	0.96	0.96	95.59	0.96	0.96	0.96	79	0.82	0.79	0.80
1D CNN	**97.17**	0.97	0.97	0.97	95.9	0.96	0.96	0.96	57.27	0.81	0.57	0.65
1D CNN-LSTM	95.85	0.96	0.96	0.96	94.30	0.95	0.94	0.94	64.13	0.77	0.64	0.69

formance of LSTM, BI-LSTM, 1D CNN, and 1D CNN-LSTM models trained on wrist sensor data from the ScientISST MOVE Dataset. As can be seen from the Table, the average accuracy, precision, recall, and F1-score of about 95.57%, 0.96, 0.95, and 0.96, respectively, was obtained. The best performance was obtained by the 1D CNN model with an increment in accuracy of about 2.13% over other models. Table 1 shows the HAR performance of LSTM, BI-LSTM, 1D CNN, and 1D CNN-LSTM models trained on chest sensor data from the ScientISST MOVE Dataset. As can be seen in the table, the average accuracy, precision, recall, and F1 score of about 95.47%, 0.96, 0.95, and 0.95, respectively, was obtained. The best performance was obtained by the LSTM model with an increment in accuracy of about 1.00% over other models.

Table 1 shows the HAR performance of LSTM, BI-LSTM, 1D CNN, and 1D CNN-LSTM models trained on forearm sensor data from the ScientISST MOVE Dataset. As can be seen in the table, the average accuracy, precision, recall, and F1 score of about 69.03%, 0.74, 0.69, and 0.70, respectively, was obtained. The best performance was obtained by the Bi-LSTM model with an increase in accuracy of about 13% over other models. Table 2 shows the HAR performance of the LSTM, BI-LSTM, 1D CNN, and 1D CNN-LSTM models trained on sensor fusion data from the ScientISST MOVE dataset. As can be seen in the table, the average accuracy, precision, recall, and F1 score of about 95.83%, 0.96, 0.95, and 0.96, respectively, was obtained. The best performance was obtained by the 1D CNN model with an increment in accuracy of about 4.0% over other models.

Table 2. HAR of the deep learning models on sensor-fusion (Wrist + Chest+ Forearm fusion) data from the ScientISST MOVE dataset.

Models	**ACC**	**P**	**R**	**F1**
LSTM (sensor-fusion)	93.32	0.95	0.93	0.94
BI-LSTM (sensor-fusion)	92.82	0.95	0.93	0.94
1D CNN (sensor-fusion)	98.76	0.99	0.99	0.99
1D CNN-LSTM (sensor-fusion)	**98.43**	0.98	0.98	0.98

In cross-comparison with single sensor-based models (i.e., wrist, chest, and forearm sensor-based models) in Tables 1 the overall increase in HAR precision, precision, recall, and F1 score of 9.14%, 7.30%, 8.70%, 9.00%. The existing work of [30] obtained an accuracy of 94.47% (which is 4% less than those obtained by our models) using an LSTM-based model on sensor data from different modalities using the voting ensemble technique. The precision of individual modalities are accelerometer (92.18%), magnetometer (85.22%) and gyroscope (78.33%). Thus, demonstrating the efficacy in the fusion of data from multiple sensors for HAR.

6. Conclusion and Future Work

The majority of the datasets for HAR capture multimodal time series data from multiple wearable sensors. However, for most studies, these multimodal signals have been processed individually or combined before being fed to deep learning models in the existing literature [14]. Thus, the true potential of complementary information from multimodal time series data captured from multiple sensors has not been utilized in existing HAR models. In this work, we proposed the use of deep learning models based on sensor fusion that fuse complementary signals from multiple wearable sensors for human activity classification. Our experiments on two widely adopted datasets suggest an improvement in the performance of about 9% using sensor-fusion models over those based on single-sensor data for HAR. As part of future work, fine-grained feature extraction techniques together with more advanced attention-based fusion modules will be investigated.

References

[1] Bianchi V, Bassoli M, Lombardo G, Fornacciari P, Mordonini M, De Munari I. Iot wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment. IEEE Internet of Things Journal 2019;6(5):8553–8562.

[2] Shi L, Xu H, Ji W, Zhang B, Sun X, Li J. Real-time human activity recognition system based on capsule and lora. IEEE Sensors Journal 2020;21(1):667–677.

[3] Chen Z, Jiang C, Xiang S, Ding J, Wu M, Li X. Smartphone sensor-based human activity recognition using feature fusion and maximum full a posteriori. IEEE Transactions on Instrumentation and Measurement 2019;69(7):3992–4001.

[4] Hossain T, Inoue S. Sensor-based daily activity understanding in caregiving center. In 2019 IEEE International Confer-

ence on Pervasive Computing and Communications Workshops (PerCom Workshops). IEEE, 2019; 439–440.
[5] Gholamiangonabadi D, Kiselov N, Grolinger K. Deep neural networks for human activity recognition with wearable sensors: Leave-one-subject-out cross-validation for model selection. Ieee Access 2020;8:133982–133994.
[6] Coelho YL, dos Santos FdAS, Frizera-Neto A, Bastos-Filho TF. A lightweight framework for human activity recognition on wearable devices. IEEE Sensors Journal 2021; 21(21):24471–24481.
[7] Surek GAS, Seman LO, Stefenon SF, Mariani VC, Coelho LdS. Video-based human activity recognition using deep learning approaches. Sensors 2023;23(14):6384.
[8] Kulchyk J, Etemad A. Activity recognition with wearable accelerometers using deep convolutional neural network and the effect of sensor placement. In 2019 IEEE SENSORS. IEEE, 2019; 1–4.
[9] Vidya B, Sasikumar P. Wearable multi-sensor data fusion approach for human activity recognition using machine learning algorithms. Sensors and Actuators A Physical 2022;341:113557.
[10] Zhang S, Li Y, Zhang S, Shahabi F, Xia S, Deng Y, Alshurafa N. Deep learning in human activity recognition with wearable sensors: A review on advances. Sensors 2022; 22(4):1476.
[11] Kunze K, Lukowicz P. Dealing with sensor displacement in motion-based onbody activity recognition systems. In Proceedings of the 10th international conference on Ubiquitous computing. 2008; 20–29.
[12] Bulling A, Roggen D. Recognition of visual memory recall processes using eye movement analysis. In Proceedings of the 13th international conference on Ubiquitous computing. 2011; 455–464.
[13] Van Kasteren T, Noulas A, Englebienne G, Kröse B. Accurate activity recognition in a home setting. In Proceedings of the 10th international conference on Ubiquitous computing. 2008; 1–9.
[14] Yang J, Nguyen MN, San PP, Li X, Krishnaswamy S. Deep convolutional neural networks on multichannel time series for human activity recognition. In Ijcai, volume 15. Buenos Aires, Argentina, 2015; 3995–4001.
[15] Zeng M, Nguyen LT, Yu B, Mengshoel OJ, Zhu J, Wu P, Zhang J. Convolutional neural networks for human activity recognition using mobile sensors. In 6th international conference on mobile computing, applications and services. IEEE, 2014; 197–205.
[16] Jiang W, Yin Z. Human activity recognition using wearable sensors by deep convolutional neural networks. In Proceedings of the 23rd ACM international conference on Multimedia. 2015; 1307–1310.
[17] Reiss A, Indlekofer I, Schmidt P, Van Laerhoven K. Deep ppg: Large-scale heart rate estimation with convolutional neural networks. Sensors 2019;19(14):3079.
[18] Saraiva JA, Abreu M, Carmo AS, da Silva HP, Fred A. Scientisst move: Annotated wearable multimodal biosignals recorded during everyday life activities in naturalistic environments. Circulation 2023;101:e215–e220.
[19] Kwapisz JR, Weiss GM, Moore SA. Activity recognition using cell phone accelerometers. ACM SigKDD Explorations Newsletter 2011;12(2):74–82.
[20] Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL, et al. A public domain dataset for human activity recognition using smartphones. In Esann, volume 3. 2013; 3.
[21] Vepakomma P, De D, Das SK, Bhansali S. A-wristocracy: Deep learning on wrist-worn sensing for recognition of user complex activities. In 2015 IEEE 12th International conference on wearable and implantable body sensor networks (BSN). IEEE, 2015; 1–6.
[22] Tang Y, Teng Q, Zhang L, Min F, He J. Layer-wise training convolutional neural networks with smaller filters for human activity recognition using wearable sensors. IEEE Sensors Journal 2020;21(1):581–592.
[23] Shi W, Fang X, Yang G, Huang J. Human activity recognition based on multichannel convolutional neural network with data augmentation. IEEE Access 2022;10:76596–76606.
[24] Almaslukh B, AlMuhtadi J, Artoli A. An effective deep autoencoder approach for online smartphone-based human activity recognition. Int J Comput Sci Netw Secur 2017; 17(4):160–165.
[25] Inoue M, Inoue S, Nishida T. Deep recurrent neural network for mobile human activity recognition with high throughput. Artificial Life and Robotics 2018;23:173–185.
[26] Uddin MZ, Torresen J. Activity recognition using smartphone sensors, robust features, and recurrent neural network. In 2019 13th international symposium on medical information and communication technology (ISMICT). IEEE, 2019; 1–6.
[27] Wang L, Liu R. Human activity recognition based on wearable sensor using hierarchical deep lstm networks. Circuits Systems and Signal Processing 2020;39:837–856.
[28] Mekruksavanich S, Jitpattanakul A. Lstm networks using smartphone data for sensor-based human activity recognition in smart homes. Sensors 2021;21(5):1636.
[29] Li Y, Wang L. Human activity recognition based on residual network and bilstm. Sensors 2022;22(2):635.
[30] Chung S, Lim J, Noh KJ, Kim G, Jeong H. Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning. Sensors 2019;19(7):1716.

Address for correspondence:

Ajita Rattani, Ph.D.
Dept of Computer Science and Engineering,
University of North Texas at Denton, Texas, USA.
email: ajita.rattani@unt.edu