2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW) | 979-8-3503-0261-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSPW59220.2023.10193698

# ENHANCING HUMAN ACTIVITY RECOGNITION THROUGH SENSOR FUSION AND HYBRID DEEP LEARNING MODEL

*Adane Nega Tarekegn, Mohib Ullah, Faouzi Alaya Cheikh, Muhammad Sajjad*

The Software, Data and Digital Environments (SDDE) Research Group, Department of Computer Science, Norwegian University of Science and Technology (NTNU), Gjøvik, Norway

## ABSTRACT

Wearable-based human activity recognition (HAR) is essential for several applications, such as health monitoring, physical training, and rehabilitation. However, most HAR systems presently depend on a single sensor, typically a smartphone, due to its widespread use. To improve performance and adapt to various scenarios, this study focuses on a smart belt equipped with acceleration and gyroscope sensors for detecting activities of daily living (ADLs). The collected data was pre-processed, fused and used to train a hybrid deep learning model incorporating a CNN and BiLSTM network. We evaluated the effect of window length on recognition accuracy and conducted a performance analysis of the proposed model. Our framework achieved an overall accuracy of 96% at a window length of 5 seconds, demonstrating its effectiveness in recognizing ADLs. The results show that belt sensor fusion for HAR provides valuable insights into human behaviour and could enhance applications such as healthcare, fitness, and sports training.

**Index Terms—**sensor fusion, human activity recognition**,** deep learning, smart belt, wearable sensor.

## 1. INTRODUCTION

The field of human activity recognition (HAR) has witnessed significant progress thanks to the advancements in sensor technology and the growth of the Internet of Things (IoT) [1]. With the increasing availability of miniaturized sensing devices that are cost-effective and consume low energy, sensor-based activity recognition has become a critical and influential topic in various research domains [2]. HAR is a method of automatically detecting and analyzing human movements using information obtained from multiple sensing devices [3]. HAR can be classified into two main categories: vision-based and sensor-based. Vision-based HAR utilizes video sensor technologies, such as RGB cameras, to monitor and recognize the actions of the subject, while sensor-based recognition is based on body-worn sensors, such as those found in bands, smartwatches, clothes, and smartphones [4]. Wearable sensors are representative examples of state-of-the-art sensors for detecting human motion, vibrations, and orientation changes in three axes. Wearables are ubiquitous, preserve user privacy, and have less computational complexity compared to vision-based HAR. With the growing maturity of artificial intelligence (AI) and deep learning approaches, wearable-based HGR has become popular in various domains, including healthcare services [5], smart homes [6], athlete monitoring [7], security, and surveillance systems [8]. This paper focuses on wearable-based HAR using a fusion of sensors.

The use of wearable sensors for human activity recognition (WS-HAR) has gained traction in the healthcare system due to its ease of use, cost-effectiveness, and ability to provide continuous monitoring [9]. Moreover, these sensors can serve as a substitute for assessing the frailty phenotype in elderly people [10][11]. Numerous studies demonstrate that recognizing physical activity and regular monitoring can potentially reduce the risk of several diseases in people, such as neurological disorders, cardiovascular disease, and type 2 diabetes [12]. The significance of WS-HAR lies in its ability to not only recognize daily activities but also how those activities are performed. This can be helpful in monitoring a patient's recovery after surgery, diagnosing the state of diseases, or predicting falls. Deep learning techniques are exciting methods well-suited for wearable-based gait activity prediction.

Convolutional neural networks (CNNs) are specifically designed for image processing and analysis, but they can also be applied to other types of data, such as wearable sensor data [13]. The advantage of using CNNs for HAR is that they can learn to extract features automatically from the raw sensor data without the need for hand-crafted feature engineering. Wearable sensor-based activity recognition tasks have extensively utilized CNN and LSTM networks [14]. A recent review of such deep learning techniques for wearable-based HAR is provided in reference [15].

In this paper, we propose a hybrid deep learning framework based on wearable sensor-level data fusion for activity recognition problems with improved performance. The proposed model is a combination of two powerful deep learning techniques, namely, 1D Convolutional Neural Networks (CNNs) and Bidirectional Long Short-Term Memory (BLSTM) networks. The robustness of the


Authorized licensed use limited to: Imperial College London. Downloaded on October 31,2023 at 17:04:42 UTC from IEEE Xplore. Restrictions apply.

resulting recognition system has been evaluated using an inertial measurement unit (IMU) sensor integrated and used as a smart belt. One belt sensor contains 3 IMU sensors involving both acceleration and gyroscope sensors to collect triaxial linear and triaxial angular data at different points in the human waist. The belt provides more possibilities and higher accuracy for HAR tasks with an adaptation of various scenarios and thus plays an essential role in obtaining posture information. Thus, the main contribution of this article is summarized as follows:

- To recognize activities of daily living and assess the effect of increasing window size on recognition performance, we employed a smart belt with three motion sensors mounted around the human waistline.
- The study proposes hybrid deep learning as a viable and efficient method of offering an accurate solution to human gait activity recognition.
- We introduce a new smart belt dataset that utilizes a fusion of sensors located at different points around the waist for HAR, which differs from current datasets that mainly depend on a single sensor, usually a smartphone.

## 2. PROPOSED HAR MODEL FRAMEWORK

The entire configuration of our proposed wearable sensor-based HAR framework is presented in Fig. 1. The process starts with data acquisition, which involves gathering data from smart belt sensors. The next step is performing sensor-level data fusion, which combines acceleration and gyroscope sensors from three different positions around the human waist. The data pre-processing is applied to the fused sensor data, including noise reduction, missing data filling, and data normalization. Data segmentation is also necessary to convert the multi-dimensional sensor data into sample data suitable for model training. This involves defining temporal windows, determining the overlap of temporal windows, and assigning class labels. The proposed framework primarily includes CNN-BiLSTM where its performance is evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. The confusion matrix is used to compare the results of the proposed model.

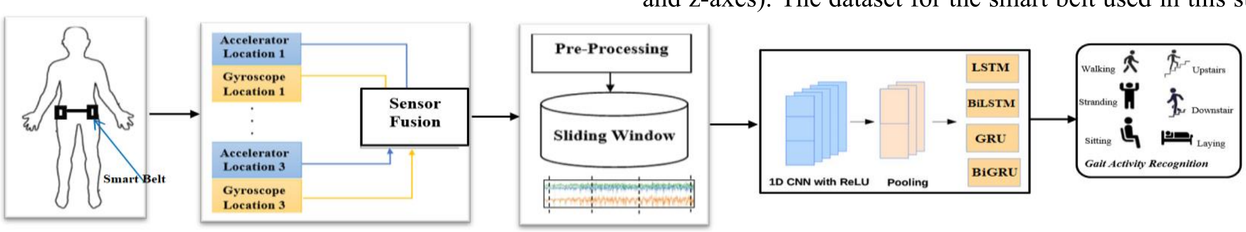


Fig.1 The proposed framework of wearable sensor based HAR.

### 2.1 Smart Belt Data Collection

Our source of data for this study is the smart belt motion tracking sensor, which was designed and developed at the Norwegian University of Science and Technology (NTNU). The smart belt data was collected at NTNU's university campus by performing experiments on twelve volunteers for performing activities of daily living (ADL). These activities include walking, walking downstairs, walking upstairs, sitting, lying, and standing, with each activity performed two times. Fig.2. shows an example of the daily activity samples collected from one IMU sensor mounted on the left hip of a volunteer performing a 'Walking' activity with two trials (repetitions).


Fig.2. Example data for 'Walking' activity one user.

Each subject wore three IMU (inertial measurement unit) sensors simultaneously at specific body locations around the waist, including left, right, and middle hip, to measure the physical activities of the user. The sampling rate was 100 Hz, and a total of 22 attributes were collected from sensors for each sample. The sensors worn by the subjects were securely mounted on a regular elastic waist belt. This resulted in minimal movement between the sensor and the elastic waist belt. All participants were required to make all the six ADLs in the way they used to. The sensors recorded both linear accelerometer and gyroscope sensor data using the three types of oriental motions (forward-backwards, upward-downward, and sideways with relation to the x, y, and z-axes). The dataset for the smart belt used in this study will be made publicly available in the future.

### 2.2 Pre-processing

Before analysis and modelling, the acquired multi-dimensional data needs to undergo a pre-processing stage. In the case of the smart belt dataset, several techniques were employed to pre-process the data, including noise filtering, normalization, segmentation, and data balancing. To reduce noise and improve the quality of the signal, a median filter and a third-order low-pass Butterworth filter were employed with a cut-off frequency of 20Hz, as most of the information contained in human body movement is below this frequency [16]. Min-max normalization was used to scale the data into

Authorized licensed use limited to: Imperial College London. Downloaded on October 31,2023 at 17:04:42 UTC from IEEE Xplore. Restrictions apply.

the range between 0 and 1 with mean and standard deviation**.** Segmentation was done using a fixed-size sliding window to extract relevant features or patterns from the data, and the most frequently occurring activity within a window was selected as its label. Data balancing was also implemented to address any imbalances in the data and improve the performance of the classification or prediction model. To annotate the collected signals from the IMU sensors in the smart belt, the NOVA annotation tool was used [17]. The goal was to recognize the six activities of daily living: walking downstairs, walking, walking upstairs, standing, sitting, and lying. Although time-consuming, using the NOVA annotation tool provided a more precise labelling process.

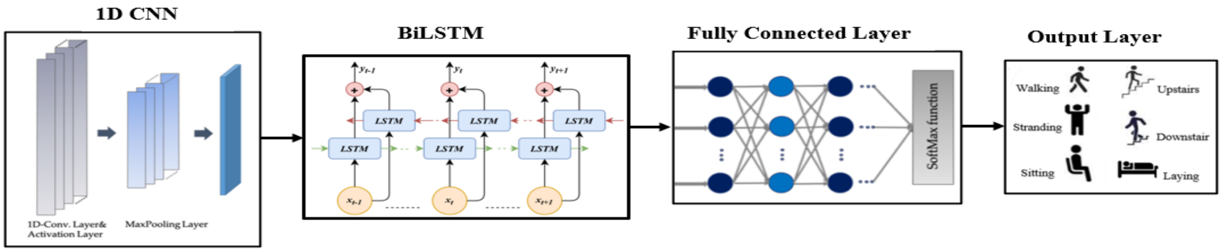


Fig.3. Proposed architecture of CNN-BiLSTM for activity recognition problem.

### 2.3 Hybrid Deep Learning Model

The proposed CNN-BiLSTM architecture for the activity recognition problem is presented in Fig.3. In this paper, 1D CNN is combined with BiLSTM as a powerful technique for the HAR task. The 1D CNN is applied to extract local features from the input sequence by applying a set of filters to sliding windows of the input. It consists of one or more convolutional layers, followed by a max-pooling layer. The output of the CNN is a sequence of feature maps. The BiLSTMs, on the other hand, are good at processing sequential data, such as time series data from sensors. It takes the sequence of feature maps from the CNN as input and learns the temporal dependencies in the data. It consists of one or more LSTM layers, each with a forward and backward direction. The output of the BiLSTM is a sequence of hidden states. The classifier takes the final hidden state from the BiLSTM and uses it to classify the activity. The classifier can be a simple, fully connected layer with softmax activation. The final output of the network is a prediction of the activity performed by the person.

## 3. EXPERIMENTAL RESULTS AND DISCUSSIONS

This section presents the experimental results of the proposed model. It includes the experimental setup, results, and discussions, as well as a comparison of our proposed model with other hybrid models.

### 3.1 Experimental setup

The implementation and experiments were carried out with various frameworks and libraries, such as Anaconda with Python 3.10.5 and TensorFlow 2.9 with Keras frontend. The proposed model was trained and evaluated using the smart belt dataset, which was segmented and reshaped into suitable input dimensions. The generated samples were split into a training dataset and a testing dataset with a ratio of 80% and 20%. On the training dataset, grid search with 5-fold cross-validation was utilized to tune the models with hyperparameters.

### 3.2 Experimental Results

In our experiment, we first examined the influence of increasing window sequence length on the performance of the CNN-BiLSTM model for action recognition. In this paper, we validate the recognition performance of the proposed hybrid model by varying the window sequence lengths (0.5,1,2, 3,…, 9 in seconds). A comparison is made between different sequence lengths to determine their impact on the performance of the proposed CNN-BiLSTM model, and the results are presented in Table 1. We realized

Table 1. Recognition performance of CNN-BiLSTM model across different window sequence length

| | **Window Size (seconds)** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **0.5** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
| **Accuracy** | 79.72 | 84.52 | 88.21 | 92.47 | 92.32 | **96.02** | 91.88 | 93.59 | 93.87 | 94.58 |
| **Precision** | 81.40 | 85.20 | 88.70 | 92.61 | 92.65 | **96.00** | 91.95 | 93.54 | 93.95 | 94.53 |
| **Recall** | 79.50 | 84.60 | 88.31 | 92.57 | 92.19 | **95.97** | 92.05 | 93.28 | 93.97 | 94.66 |
| **F1-Score** | 80.00 | 84.70 | 88.20 | 92.38 | 92.32 | **95.96** | 91.82 | 93.33 | 93.85 | 94.55 |

Authorized licensed use limited to: Imperial College London. Downloaded on October 31,2023 at 17:04:42 UTC from IEEE Xplore. Restrictions apply.

that when the sequence length is increased, the performance of the proposed CNN-BiLSTM model is improved significantly on the smart belt dataset. The performance of the model is increased for larger window sizes where the best accuracy (96.02%) and F1-score (95.96%) are attained at a window length of 5s. After the sequence length of 5s, the performance becomes stagnant and starts decreasing for all evaluation metrics. Although larger window sizes provide more past and future information to analyze and identify more complex activities, they contain redundant information and lead to large recognition latency [18][19]. In this paper, a cut-off window size of 5s is used as the best trade-off between accuracy and latency to evaluate the performance of the proposed CNN-BiLSTM. Fig. 4 presents the confusion matrix obtained from the evaluation of the CNN-BiLSTM at a window size of 5s, which summarizes the number of correct and incorrect predictions made by the model for each activity class. As shown in Fig.4, except for the 'laying' and 'walking' activities, all the other activities were perfectly classified with high accuracy (>=95%) using the CNN-BiLSTM model. All the samples in the 'Standing' activity were correctly classified with an accuracy of 98.4%,

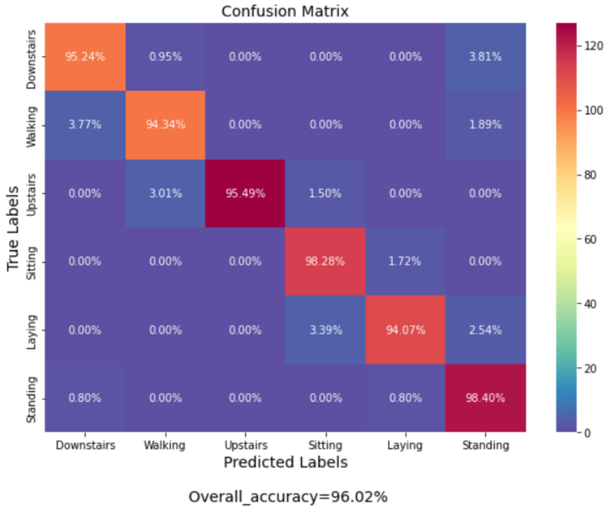


Fig.4. Confusion matrix of the CNN-BiLSTM model with only 1.6% of the samples wrongly classified as 'laying' and 'downstairs. It is also evident that the CNN-BiLSTM model had the most difficulty in distinguishing the 'Laying' from the 'Standing' and 'Sitting' activities. This is partly because the 'Laying' and 'Sitting' is very similar for some of the subjects, thus generating similar motion signals. For the 'Walking' activity, where 5.66% of the samples are wrongly classified as 'Downstairs' and 'Standing'. Fig. 5 shows the accuracy loss of the model for 50 epochs where a gradual decrease in loss can be observed throughout the training session. The CNN-BiLSTM model was compared with other hybrid models in terms of accuracy, precision, recall and F1-score. Table 2 depicts the comparison of the CNN-BiLSTM with the other three hybrid models, namely CNN-LSTM, CNN-GRU, and CNN-BiGRU, in terms of accuracy, precision, recall, and F1 score using a window length of 5s. The comparative results in Table 2 demonstrate that the CNN-BiLSTM model achieved the highest performance across all the evaluation measures.

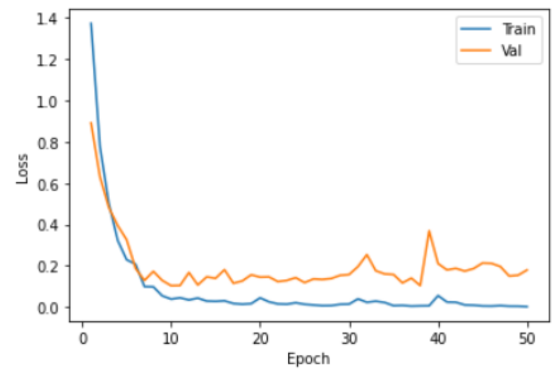


Fig.5. Accuracy loss of the proposed model

Table 2. Performance of the proposed model

| Models | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| CNN-LSTM | 95.59 | 99.95 | 95.50 | 95.50 |
| CNN-GRU | 95.73 | 95.64 | 95.69 | 95.70 |
| CNN-BiGRU | 95.02 | 95.08 | 95.00 | 94.99 |
| CNN-BiLSTM | **96.02** | **96.00** | **95.97** | **95.96** |

## 4. CONCLUSIONS

The study proposes a 1D CNN-BiLSTM hybrid model for human activity recognition based on sensor-level data fusion. The proposed model utilizes CNN layers for feature extraction and BiLSTM for sequence prediction on the smart belt dataset. The study demonstrates that the proposed hybrid model achieved good performance on HAR. One of the significant advantages of the proposed model is that it employs three motion sensors mounted around the waist as a smart belt, which differs from traditional activity recognition methods that rely on a single sensor, usually on a smartphone. Additionally, the study investigates the impact of increasing window length on the recognition performance of the proposed model and finds that longer window lengths lead to better performance. The results of the study show that the proposed 1D CNN-BiLSTM hybrid model can effectively recognize human activities using smart belt sensors, indicating its potential applications in healthcare.

## 5. ACKNOWLEDGEMENT

The European Union funded this research through the Horizon 2020 Research and Innovation Programme, in the context of the ALAMEDA (Bridging the Early Diagnosis and Treatment Gap of Brain Diseases via Smart, Connected, Proactive and Evidence-based Technological Interventions) project under grant agreement No GA 101017558.

Authorized licensed use limited to: Imperial College London. Downloaded on October 31,2023 at 17:04:42 UTC from IEEE Xplore. Restrictions apply.

## 6. REFERENCES


[1] D. Sehrawat and N. Singh Gill, “IoT Based Human Activity Recognition System Using Smart Sensors,” Adv. Sci. Technol. Eng. Syst., 2020, doi: 10.25046/aj050461.

[2] F. Serpush, M. B. Menhaj, B. Masoumi, and B. Karasfi, “Wearable Sensor-Based Human Activity Recognition in the Smart Healthcare System,” Computational Intelligence and Neuroscience. 2022. doi: 10.1155/2022/1391906.

[3] N. Gupta, S. K. Gupta, R. K. Pathak, V. Jain, P. Rashidi, and J. S. Suri, “Human activity recognition in artificial intelligence framework: a narrative review,” Artif. Intell. Rev., 2022, doi: 10.1007/s10462-021-10116-x.

[4] L. Minh Dang, K. Min, H. Wang, M. Jalil Piran, C. Hee Lee, and H. Moon, “Sensor-based and vision-based human activity recognition: A comprehensive survey,” Pattern Recognit., 2020, doi: 10.1016/j.patcog.2020.107561.

[5] F. Serpush, M. B. Menhaj, B. Masoumi, and B. Karasfi, “Wearable Sensor-Based Human Activity Recognition in the Smart Healthcare System,” Computational Intelligence and Neuroscience. 2022. doi: 10.1155/2022/1391906.

[6] J. Suto, S. Oniga, C. Lung, and I. Orha, “Comparison of offline and real-time human activity recognition results using machine learning techniques,” Neural Comput. Appl., 2020, doi: 10.1007/s00521-018-3437-x.

[7] Y. L. Hsu, S. C. Yang, H. C. Chang, and H. C. Lai, “Human Daily and Sport Activity Recognition Using a Wearable Inertial Sensor Network,” IEEE Access, 2018, doi: 10.1109/ACCESS.2018.2839766.

[8] T. Keerthana, K. Kaviya, S. D. Priya, and A. S. Kumar, “Retraction: AI enabled smart surveillance system,” Journal of Physics: Conference Series. 2021. doi: 10.1088/1742-6596/1916/1/012034.

[9] E. Ramanujam, T. Perumal, and S. Padmavathi, “Human Activity Recognition with Smartphone and Wearable Sensors Using Deep Learning Techniques: A Review,” IEEE Sensors Journal. 2021. doi: 10.1109/JSEN.2021.3069927.

[10] A. Tarekegn, F. Ricceri, G. Costa, E. Ferracin, and M. Giacobini, “Predictive modeling for frailty conditions in Elderly People: Machine learning approaches,” JMIR Med. Informatics, 2020, doi: 10.2196/16678.

[11] A. Tarekegn, F. Ricceri, G. Costa, E. Ferracin, and M. Giacobini, “Detection of Frailty Using Genetic Programming: The Case of Older People in Piedmont, Italy,” 2020. doi: 10.1007/978-3-030-44094-7_15.

[12] L. Sigcha et al., “Deep learning approaches for detecting freezing of gait in parkinson’s disease patients through on-body acceleration sensors,” Sensors (Switzerland), 2020, doi: 10.3390/s20071895.

[13] Y. Tang, Q. Teng, L. Zhang, F. Min, and J. He, “Layer-Wise Training Convolutional Neural Networks with Smaller Filters for Human Activity Recognition Using Wearable Sensors,” IEEE Sens. J., 2021, doi: 10.1109/JSEN.2020.3015521.

[14] S. K. Challa, A. Kumar, and V. B. Semwal, “A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data,” Vis. Comput., 2021, doi: 10.1007/s00371-021-02283-3.

[15] S. Zhang et al., “Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances,” Sensors, 2022, doi: 10.3390/s22041476.

[16] D. Garcia-Gonzalez, D. Rivero, E. Fernandez-Blanco, and M. R. Luaces, “A public domain dataset for real-life human activity recognition using smartphone sensors,” Sensors (Switzerland), 2020, doi: 10.3390/s20082200.

[17] A. Heimerl, T. Baur, F. Lingenfelser, J. Wagner, and E. Andre, “NOVA-A tool for eXplainable Cooperative Machine Learning,” 2019. doi: 10.1109/ACII.2019.8925519.

[18] Banos, J. M. Galvez, M. Damas, H. Pomares, and I. Rojas, “Window size impact in human activity recognition,” Sensors (Switzerland), 2014, doi: 10.3390/s140406474.

[19] M. Ullah, H. Ullah, S. D. Khan, and F. A. Cheikh, “Stacked Lstm Network for Human Activity Recognition Using Smartphone Data,” 2019. doi: 10.1109/EUVIP47703.2019.8946180.


Authorized licensed use limited to: Imperial College London. Downloaded on October 31,2023 at 17:04:42 UTC from IEEE Xplore. Restrictions apply.