Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Attention and Deep Learning Framework for Wearable Sensor-based Human Activity Recognition

Aishvarya Garg[1,3], Swati Nigam[2,3], Rajiv Singh[2,3]
[1]*Department of Physical Science, Banasthali Vidyapith, Rajasthan 304022, India*
[2]*Department of Computer Science, Banasthali Vidyapith, Rajasthan 304022, India*
[3]*Centre for Artificial Intelligence, Banasthali Vidyapith, Rajasthan 304022, India*
aishvaryagarg@gmail.com, swatinigam.au@gmail.com, jkrajivsingh@gmail.com

ARTICLE INFO

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

Human Activity Recognition (HAR) using wearable sensors has emerged as a significant research area that has found a key role in fitness tracking, ambient assisted living, and smart environments. Traditional machine learning (ML) methods with handcrafted features often exhibit limited ability in learning complex patterns and adaptability across datasets. To overcome this issue, deep learning (DL) techniques offer improved performance by automating feature extraction and capturing sequential patterns. However, DL-based HAR methods often face limitations such as high computational complexity and overfitting risks with deeper networks. To address these limitations, this paper proposes a novel, lightweight, attention-deep learning-based framework tailored for wearable sensor-based HAR (WHAR). The proposed method processes raw accelerometer readings through a convolutional autoencoder (ConvAE) architecture comprising an average pooling layer as a bottleneck layer for initial feature extraction. A self-attention layer is added to highlight the relevant, informative features, followed by two stacked long short-term memory (LSTM) layers to extract the deeper feature representation and long-term dependencies. These features are then passed through fully connected layers to classify activities. A scaling-based data augmentation technique is employed to address the imbalanced nature of datasets. The proposed method attained accuracies of 97.21%, 95.54%, and 99.84% on three publicly available datasets, namely, HAR70+, HARTH and MHealth, respectively. The experimental results demonstrate that the proposed framework achieved better performance across the wearable sensor-based application by introducing attention mechanism and augmentation techniques.

Keywords: WHAR, deep learning, ConvAE, self-attention, LSTM, augmentation technique

INTRODUCTION

The rapid advancements in wearable technologies have revolutionized the domain of healthcare and personalized medical interventions, smart environments and assisted living [1][2]. The primary objective of HAR is to identify and classify physical activities performed by individuals through sensor data, which enables real-time monitoring. In sensor-based HAR, wearable-based methods have gained significant attention due to their ease of use and continuous monitoring capabilities [3]. These applications include smart home environments, elderly care, medical diagnosis, and rehabilitation programs. However, classification based on precise and relevant information about human activities remains a computational challenge in sensor-based HAR. Wearable sensors typically leverage inertial measurement units (IMU) composed of tri-axial accelerometers and gyroscopes to measure body movements in terms of acceleration and angular velocity. Among these, accelerometers are the most widely utilized sensors in wearable-based HAR due to their ability to capture fine-grained motion data that help in learning variations in body movement and orientation, enabling the models to classify physical activities, particularly among elders [4].

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Many traditional handcrafted-based methods have been proposed. In those methods, first statistical features are extracted from the raw sensor data, followed by various ML-based classifiers such as support vector machines (SVM), random forests (RF), decision trees (DT) and K-nearest neighbor (KNN) for the classification of the activities. However, these methods often fail to capture the complex temporal dependencies in raw accelerometer signals [5]. Many DL-based methods have been exploited to overcome this problem of learning complex patterns. Among them, convolutional neural networks (CNN), long short-term memory (LSTM) and gated recurrent networks (GRUs), along with their variants, are primarily leveraged due to their remarkable performance in extracting both local and long-term dependencies from sensor-based data [6][7][8]. To further improve the feature representations, an attention mechanism is introduced in the deep learning methods to focus on the relevant patterns. Attention modules such as self-attention, multihead attention, and convolutional block attention module (CBAM) [9][10][11][12] are utilized after feature extraction to select the relevant features.

Considering the advantages of accelerometer readings, CNN, LSTM, and attention mechanisms, we propose a novel lightweight, attention-deep-based, sensor-based HAR tailored for healthcare applications. The proposed method leverages the raw accelerometer readings, which, in turn, undergo ConvAE architecture with an average pooling layer as the bottleneck layer for initial feature extraction, followed by a self-attention layer for learning the relevant features. These features are passed through two LSTM layers to extract deeper feature representations and long-term dependencies. The extracted features are flattened and passed to four fully connected layers to classify activities from sensor-based data. In the proposed framework, an augmentation technique called scaling is employed to tackle the problem of the imbalanced nature of datasets. The proposed framework is validated on publicly available datasets – HAR70+, HARTH and MHealth. Detailed performance and comparative analyses for each dataset are conducted. Furthermore, an ablation study is performed to investigate the impact of incorporating BiLSTM, two optimizers (Adam and RMSProp) and two activation functions (ReLU and Leaky ReLU), both with and without augmentation techniques.

The contributions of the paper are as follows:

1. A hybrid DL-attention-based framework combining ConvAE, self-attention and LSTM for sensor-based HAR.
2. The scaling augmentation technique is applied to address the imbalanced nature of datasets.
3. For the validation of the proposed methodology, three sensor-based datasets, namely, HAR70+, HARTH, and MHealth, are exploited in which the model gained an accuracy of 97.21%, 95.54%, and 99.84%, respectively.
4. In this paper, we also conducted an ablation study for the impact of BiLSTM layers, two different optimizers and activation functions with and without augmentation techniques.

The structure of the paper is as follows: Section 2 discusses HAR approaches based on deep and attention mechanisms. Section 3 details the proposed method, which utilizes deep learning algorithms and attention mechanisms. Section 4 describes the experimental setup for validations, which comprises details of datasets, performance analysis based on confusion matrices, training/validation loss/accuracy graphs, testing accuracy, comparative analysis, and ablation study for each dataset. Section 5 comprises the conclusions from the proposed method.

LITERATURE REVIEW

HAR using wearable sensors has become increasingly significant for healthcare monitoring and fitness tracking applications. Traditional ML approaches with handcrafted features often fail to generalize well due to their limited ability to capture complex spatial and temporal features. In contrast, DL methods demonstrate exemplary performance in feature extraction and classification phases. Further enhancing the feature representations, attention mechanisms have been adopted. This section discusses some DL and attention-based WHAR methods.

2.1. Deep Learning-based WHAR methods

Motivated by the automated feature of DL algorithms, Gupta (2021) [13] explores a DL approach that employs CNN and GRU algorithms to extract spatial and sequential dependencies. Similarly, Thu and Han (2021) [14] proposed a two-stage framework that focuses on the extraction of local features via CNN1D and BiLSTM from the window-based

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

sensor data at the data level, while global BiLSTM is employed for extracting the long-term dependencies from the adjacent windows as global features.

Another method is introduced by Luwe et al. (2022) [15] that leverages CNN1D and BiLSTM to extract high-level representative features and long-term dependencies sequentially. Chandramouli et al. (2024) [16] designed a hybrid CNN-BiLSTM method that leverages the undersampling technique as pre-processing for the detection of activities performed by the elderly. Additionally, Nazar and Jalal (2024) [17] proposed a HAR method in which six statistical features are extracted and passed to a binary grey wolf optimizer (GWO), followed by multilayer perceptron (MLP) for optimization and classification, respectively.

2.2. Attention-based WHAR methods

Inspired by the benefits of attention mechanisms, Nithin et al. (2021) [18] proposed an attention-based deep learning framework that focuses on the classification of activities performed by the elderly. The authors extended baseline deep convolutional long short-term memory (DeepConvLSTM) architecture by adding two attention modules. The proposed framework attained an improved accuracy. Similarly, Al-Qaness et al. (2023) [19] introduced a multilevel-based residual network that comprises an initial block and residual block in a parallel manner. The initial block consists of CNN1D, batch normalization, and ReLU, while the residual block consists of two CNN1D, batch normalization, ReLU layers and a residual connection. The resulting features are then concatenated and fed to BiGRU to extract temporal dependencies. An attention layer is added to extract the meaningful features for the classification of activities. Similarly, Zhang and Xu (2024) [20] introduced a multilevel network that leverages CNN and BiGRU to capture spatial and temporal dependencies. Along with it, spatial and temporal attention modules are designed to extract spatiotemporal attention maps in order to learn deeper feature representation. Meanwhile, a HAR method is introduced by Tang et al. (2023) [21], addressing the improvement in the CNN-based method without increasing the complexity and memory. In the proposed method, the hierarchical split idea is applied to extract multiscale features.

Ullah et al. (2024) [22] developed a CNN-LSTM DL-based method that leverages a squeeze-and-excite (SE) attention module to enhance interdependencies and sparse learning in the fully connected layers. To address the long-term dependencies challenges in the domains of IoT, AbdelRaouf et al. (2024) [23] presented a framework that utilizes CNN1D, GRU and MHA to extract the spatial and temporal dependencies. Another method proposed by Mekruksavanich et al. (2024) [11] utilizes CNN and BiLSTM for the sequential extraction of spatio-temporal features. The extracted features are then passed through the CBAM mechanism to select relevant features before the classification of sports and daily activities.

Despite notable advancements in DL and attention-based frameworks for WHAR, several critical challenges remain unaddressed. Firstly, some methods employ ensemble frameworks that lead to a complex architecture, high computational cost and overhead. Secondly, it is observed that most approaches utilized attention modules at the end of feature extraction to select the informative features. Furthermore, the issue of class imbalance in the publicly available HAR datasets adversely affects the classification performance for minority classes.

To address the aforementioned issues, this paper proposes a lightweight, attention-deep-based framework that combines ConvAE for initial feature extraction, a self-attention layer, and two stacked LSTM layers to extract the long-term dependencies. The model is designed to operate efficiently on raw accelerometer data and is validated on three publicly available datasets. Furthermore, an augmentation technique is applied to address the imbalanced nature of the dataset, which improves the performance of the proposed method.

PROPOSED METHOD

This section details the components of the proposed WHAR method: ConvAE, LSTM and self-attention mechanism. The proposed approach for the classification of activities is discussed in detail.

3.1. Convolutional Autoencoder (ConvAE)

Autoencoder (AE) is a type of neural network designed to learn efficient feature representations from the input data. An AE consists of two primary elements: an encoder, which compresses the input data into a lower-dimensional latent representation and a decoder, which reconstructs the input from the compressed representations [8]. The

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

ConvAE extends the concept of traditional AE by integrating convolutional layers in encoders and decoders for the extraction of local patterns, hence resulting in robust and generalized feature extraction [24]. The encoding and decoding processes are mathematically represented as Equations 1 and 2.

The input data is compressed into a latent space Z^l at the encoder.

$$Z^l = \phi\left(\sum_{l\in L} W_e^{(l)} * Z^{(l-1)} + b_e^{(l)}\right), \qquad Z^0 = X \tag{1}$$

The input data is reconstructed from the encoded features at the decoder.

$$\hat{Z}^{(l)} = \phi\left(\sum_{l\in L} W_d^{(l)} \circledast \hat{Z}^{(l-1)} + b_d^{(l)}\right), \qquad \hat{X} = \hat{Z}^{(l)} \tag{2}$$

where X or Z^0 denotes the input signal, which is compressed into a latent representation Z^l at the encoder while $\hat{X}$ or $\hat{Z}^{(l)}$ denotes the reconstructed input signal at decoder at layer l. The encoder and decoder involve the convolutional and deconvolutional operations, represented by $*$ and $\circledast$ respectively. Here W_e, and b_e refers to the weight and bias of the encoder, W_d, and b_d corresponds to those of decoder. The symbol ϕ denotes the activation functions applied during both processes.

3.2. Long Short-Term Memory (LSTM)

LSTM is a type of recurrent neural network (RNN) designed to capture and retain the long-term dependencies in sequential data [8][25]. It addresses the vanishing gradients problem by introducing a unique memory cell that selectively stores or forgets the information as needed. It consists of four components: input gate, forget gate, output gate and memory, which regulate the flow and storage of relevant data while discarding unnecessary information. The mathematical representation of the gates and working in the LSTM cell is presented in Equations 3 to 8.

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \tag{3}$$

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \tag{4}$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \tag{5}$$

$$\tilde{c}_t = \tanh(w_c[h_{t-1}, x_t] + b_c) \tag{6}$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \tag{7}$$

$$h_t = \tanh(c_t) * o_t \tag{8}$$

where i_t, f_t and o_t represent input gate, forget gate, and output gate, respectively. The variable x_t denotes the input at current timestamp, while h_{t-1} is the output of previous LSTM block at timestamp t-1 and h_t is the current hidden state. The weights and biases associated with each gate are denotes as w_x and b_x, where x refers to the specific gate. The candidate cell state at timestamp t is represented by $\tilde{c}_t$ and the actual cell state (memory cell) at timestamp t is denoted by c_t.

3.3. Self-Attention Mechanism

The self-attention mechanism focuses on the relationships between different elements of the input data, computing a score for each pair of elements to capture long-term dependencies and contextual information [26]. This mechanism is done by transforming the input into three matrices- queries, keys and values, calculating the attention scores of each element with the help of the dot product of query and key matrices and multiplying the SoftMax output with the value matrix. This mechanism helps model to dynamically emphasize the most relavant information from the input sequence. The working of self-attention is depicted in Equations 9 and 10.

$$Q = XW^Q, K = XW^K, V = XW^V \tag{9}$$

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$Attention\ (Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (10)$$

where X denotes the input sensor sequence, that transformed into Q, K, V matrices with learnable weight matrices denoted as W^Q, W^K, and W^V for query, key, and value transformations, respectively. The attention score matrix is computed as QK^T with scaling factor, $\sqrt{d_k}$ where d_k is the dimensionality of the key vectors, to stabilize gradients during training.

3.4. Proposed WHAR Framework

The proposed method comprises three steps: pre-processing, feature extraction and classification of human activities. The architecture of the proposed method is illustrated in Figure 1.

Pre-Processing

The accelerometer readings are chosen from the raw data in three directions for the proposed method. Further, an augmentation technique called scaling is applied for the imbalanced classes. The scaling technique incorporates the variations in the amplitude of the sensor signal with a random scaling factor. These augmented readings are then converted into frames using the sliding window technique, which has a window size of 50 with a fixed overlap of 10.

Fe ature Extraction

The formulated frames undergo the attention-deep-based model for feature extraction. Initially, the frames are passed through an encoder framework that comprises two convolutional layers with 64 and 32 filters, a max pooling layer and a batch normalization layer to capture the low-level features. This encoder representation is then further compressed using an average pooling bottleneck layer. The decoder subsequently reconstructs the features with convolutional and upsampling layers, which aim to preserve the essential information from the original input.

A self-attention layer with 64 units is introduced to extract relevant features. The attention-based features are then processed through two LSTM layers with 64 and 32 units for richer feature extraction.

Classification

The extracted features are then passed to three fully connected layers with 32, 16, and 8 units with the Leaky ReLU activation function. The final layer is introduced with units the same as a number of classes and SoftMax activation function. To address the problem of overfitting, a dropout layer is added with 0.2%. The proposed method is presented in Algorithm 1.

Algorithm 1: Proposed Method

Step 1: Initialize

- Define input shape as (frame_size, 6)

Step 2: Apply ConvAE architecture

- Create Conv1D (64) → LeakyReLU → MaxPooling1D
- Add Conv1D (32) → LeakyReLU
- Apply BatchNormalization
- Downsample using AveragePooling1D
- Apply BatchNormalization again
- Add Conv1D (32) → UpSampling1D → Conv1D (64)

Step 3: Apply Self-Attention Layer

- If (attention is active) then

  Search:

    Compute Q, K, and V matrices

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Calculate scaled dot-product attention

Output attention-weighted values

End

Step 4: Compute deeper feature representations

- Stack LSTM (64) → LSTM (32) with return_sequences = True

Step 5: Classification

- Flatten the output tensor
- Pass through Dense (32) → Dense (16) → Dropout → Dense (8) and use LeakyReLU activation function and L2 regularization
- Pass through Dense (number of classes) and use SoftMax activation function

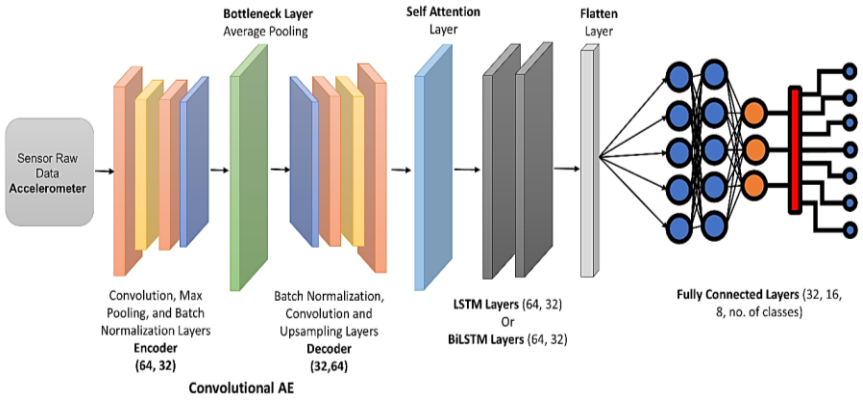

Figure 1: Architecture of the proposed WHAR framework

RESULTS AND DISCUSSIONS

This section provides details of three utilized sensor-based datasets, namely, HAR70 +, HARTH, and MHealth. For the performance analysis, the confusion matrix, accuracy, training/validation accuracy, and training/validation per epoch graph are exploited. A study on the impact of optimizers and activation functions in the proposed method with LSTM and BiLSTM configurations is also presented. The implementations of the proposed method are conducted in the Google Colaboratory Pro environment with the Keras library. The proposed framework exploits the Adam optimizer and sparse categorical cross-entropy loss for each dataset. In each dataset, the split ratio (train: validation: test) is taken as 80:15:5.

4.1. Datasets

This section provides the details of utilized datasets to validate the proposed method. The details are summarized in Table 1.

Table 1. Details of utilized datasets (Accelerometers (Acc), Gyroscope (Gryo))

	Datasets		
	HAR70+	**HARTH**	**MHealth**
Type	Wearable	Wearable	Wearable
Number of Activities	7	12	12
Number of Participants	18	22	10
Placements	Lower Back, Thigh	Lower Back, Thigh	Wrist, Ankle, Chest
Sensor Type	Acc	Acc	Acc, Gryo

HAR70 +

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

This dataset consists of accelerometer sensor readings in three axes from the device placed on the lower back and thighs of 18 adults aged between 70-85. The participants perform seven activities: walking, shuffling, ascending stairs (upstairs), descending stairs (downstairs), standing, sitting, and lying [27].

HARTH

This dataset consists of readings from two accelerometer devices worn on the thigh and lower back in three axes in free-living settings. Twenty-two participants perform 12 activities, namely, walking, running, shuffling, ascending stairs (upstairs), descending stairs (downstairs), standing, sitting, lying, cycling (sit), cycling (stand), cycling (inactive and sit), cycling (inactive and standing) [28].

MHealth

This dataset comprises the readings of ten participants performing twelve activities: standing, sitting, lying, walking, ascending stairs (upstairs), bending, handwaving, crouching, jogging, running, and jumping. The readings are collected with accelerometers and gyroscopes placed on the wrist, ankle, and chest [29].

4.2. Performance Analysis

The proposed method is validated with three publicly benchmark datasets, viz., HAR70+, HARTH and MHealth. This section provides insights into the experimental results based on the confusion matrix and training/validation accuracy/loss curves. The details of hyperparameters leveraged in the implementations of models with and without augmentation techniques are presented in Table 2.

Table 2. Details of hyperparameters leveraged with and without augmentation technique for each dataset

Hyperparameters	With Augmentation			Without Augmentation		
	Datasets			Datasets		
	HAR70+	HARTH	MHealth	HAR70+	HARTH	MHealth
Considered Number of Activities	7	10	12	7	10	12
Batch Size	128	256	128	32	64	32
Epochs	70	80	45	70	70	45
Readings per Activity	65650	221800	70000	4300	55450	10100
Total Readings	459550	2218000	840000	30100	554500	121200

The HAR70+ dataset attained a strong performance with an accuracy of 97.21%. The confusion matrix, shown in Figure 2(a), highlights some confusion among activities with similar patterns, namely, walking, shuffling and downstairs, which reduces accuracy. At the same time, other classes achieve high classification rates. The training/validation curves in Figure 2(b) reveal that the validation accuracy stabilizes near 97% after about 10 epochs, with both training/validation losses converging effectively. However, slight fluctuations are observed in validation loss, suggesting the generalization gaps are due to inter-class similarities.

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

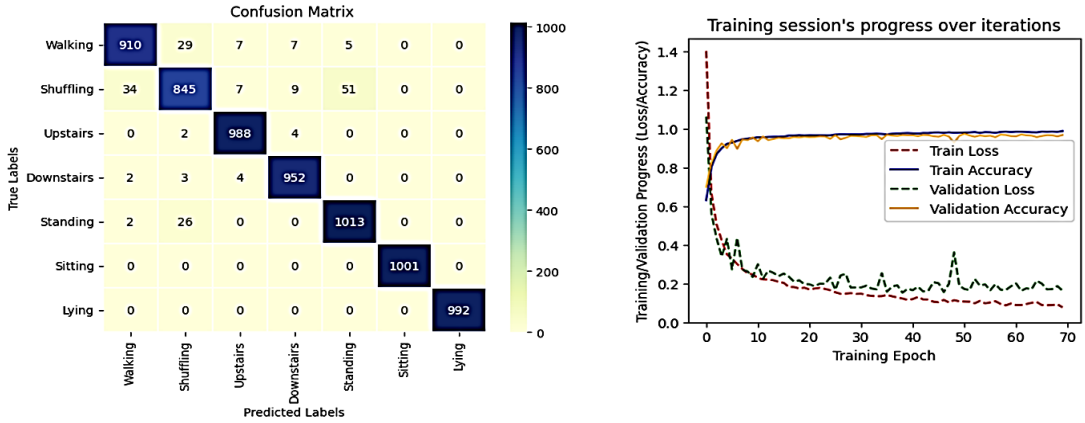

Figure 2: (a) Confusion Matrix (b) Training /Validation Accuracy/Loss curves for HAR70+ dataset

The HARTH dataset achieves an accuracy of 95.54%. The confusion matrix illustrated in Figure 3(a) indicates notable misclassifications between walking and shuffling and standing and running. These misclassifications may be caused by overlapping motion patterns when recorded by inertial sensors. The training/validation curves depicted in Figure 3(b) show stable convergence, indicating better generalization ability of the proposed method.

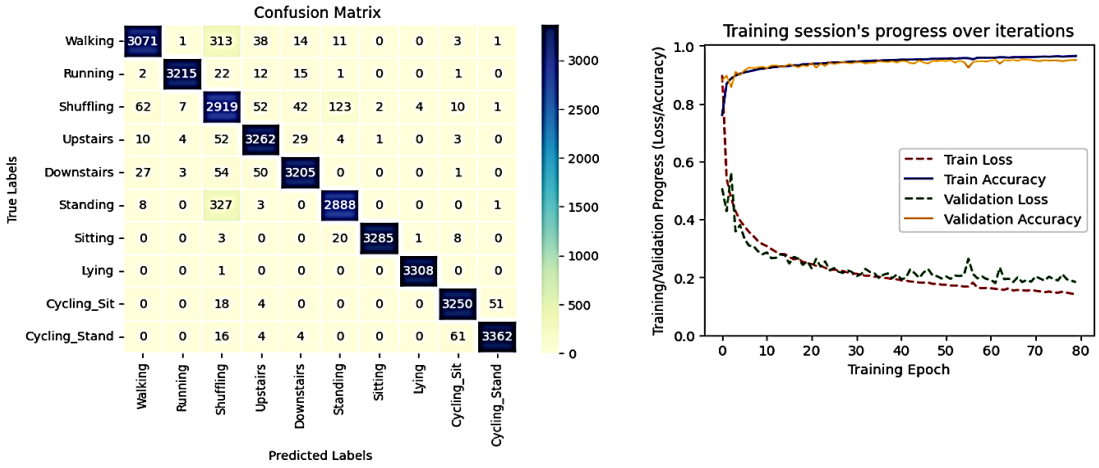

Figure 3: (a) Confusion Matrix (b) Training /Validation Accuracy/Loss curves for HARTH dataset

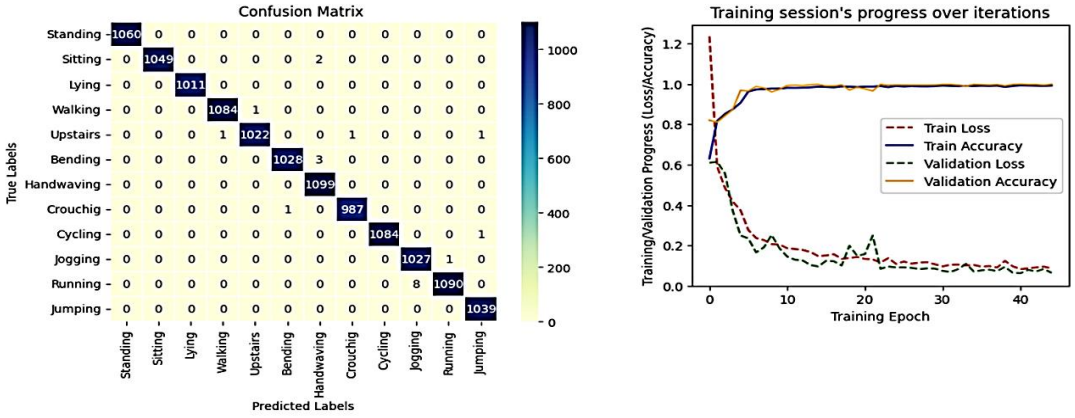

Figure 4: (a) Confusion Matrix (b) Training /Validation Accuracy/Loss curves for MHealth dataset

The MHealth dataset demonstrates superior classification performance with an accuracy of 99.84%. The confusion matrix depicted in Figure 4(a) reveals that nearly all activities are correctly classified with minimal confusion. The training/validation accuracy curves in Figure 4(b) indicate rapid convergence within the first 10 epochs.

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Furthermore, the loss curve consistently decreases and stabilizes, depicting a better generalization of the proposed method.

4.3. Comparative Analysis

This section provides the comparative analysis for three datasets with machine- and deep-learning-based methods based on the attained accuracy.

The HAR70+ dataset has been analyzed against two DL-based methods, SelfPAB and CNN + Residual BiGRU, out of which the latter demonstrated good performance in capturing temporal dependencies, as presented in Table 3. The proposed method showed comparable efficiency by integrating convolutional autoencoding with the self-attention mechanism.

Table 3. Comparative Analysis for HAR70+ Dataset

Year	Ref.	Methods	Accuracy (%)
2023	[30]	SelfPAB	93.8
2024	[31]	CNN + Residual BiGRU	97.39
		Proposed Method	97.21

The HARTH dataset has been explored using traditional machine and DL-based methods. Classical approaches utilized statistical features combined with different classifiers such as multilayer perceptron (MLP), random forests (RF), K-nearest neighbors (KNNs) and decision trees (DT). DL-based methods are compared with the proposed method, including MobileNetV3 + Efficient B0 + Wrapper Optimization, SelfPAB, and CNN-LSTM. It is observed from Table 4 that the proposed method outperformed existing methods with the help of captured local and long-term dependencies through the attention mechanism.

Table 4. Comparative Analysis for HARTH Dataset

Year	Ref.	Methods	Accuracy (%)
2023	[32]	MobileNetV3 + Efficient B0 + Wrapper Optimization	88.89
2023	[30]	SelfPAB	94.6
2023	[33]	CNN - LSTM	94.56
2023	[34]	Statistical Features + MLP	92.92
		Statistical Features + RF	92.42
		Statistical Features + KNN	86.86
		Statistical Features + DT	84.34
2025	[35]	GAF + Deep CNN	90.7
		Proposed Method	95.54

In the case of the MHealth dataset, the proposed method competes with deep and transfer learning-based methods, including deep CNN (DCNN), CNN-LSTM, Deep CNN + BiLSTM, Deep CNN – LSTM +Self-Attention and GAF + Deep CNN. It can be seen in Table 5 that the proposed method surpassed all existing methods with superior performance in the modeling of temporal patterns in sensor-based data.

Table 5. Comparative Analysis for MHealth Dataset

Year	Ref.	Methods	Accuracy (%)
2022	[36]	Deep Transfer Learning	98.63

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Year	Ref.	Methods	Accuracy (%)
2022	[37]	Deep CNN-LSTM + Self Attention	98.76
2022	[38]	DCNN	87.43
2023	[33]	CNN - LSTM	94.56
2024	[16]	Deep CNN + BiLSTM	99.5
2025	[35]	GAF + Deep CNN	90.7
		Proposed Method	99.84

4.4. Ablation Study

This section provides the studies of the impact of the augmentation technique with two optimizers and activation functions, namely, Adam, RMSProp, ReLU, and Leaky ReLU, across three datasets. The efficiency of the proposed method is also evaluated by integrating BiLSTM layers in place of LSTM layers. In the case of without augmentation technique-based implementations, the overlap is taken as 5. Also, in the implementation of the proposed method with BiLSTM configuration, training is performed for 70 epochs in the case of the HARTH dataset.

LSTM without Augmentation using Different Activation Functions and Optimizers

Tables 6 and 7 present the results of the proposed method with LSTM layers without augmentation, exploited for RMSProp and Adam optimizers. When optimized RMSProp with Leaky ReLU activation function, the method delivered better accuracy than ReLU across all datasets with accuracies of 95.08%, 94.20%, and 99.09% on HAR70+, HARTH, and MHealth, respectively. Conversely, when utilizing the Adam optimizer, ReLU outperformed Leaky ReLU in the case of the HAR70+ and MHealth datasets with an accuracy of 91.57% and 98.10%, respectively.

Table 6. Accuracy attained by the proposed method with LSTM and RMSProp (Without Augmentation)

Activation Function	Datasets		
	HAR70+	**HARTH**	**MHealth**
ReLU	90.35	94.06	97.82
Leaky ReLU (0.1)	95.08	94.20	99.09

Table 7. Accuracy attained by the proposed method with LSTM and Adam (Without Augmentation)

Activation Function	Datasets		
	HAR70+	**HARTH**	**MHealth**
ReLU	91.57	94.97	98.10
Leaky ReLU (0.1)	90.35	95.08	97.24

LSTM with Augmentation using Different Activation Functions and Optimizers

As depicted in Tables 8 and 9, the augmented dataset showed better results when applied to the LSTM-based model. When the RMSProp optimizer is utilized with Leaky ReLU, the model attained the highest accuracies of 96.67% and 94.74% on HAR70+ and HARTH, respectively. However, in the case of the MHealth dataset, the method attained the highest accuracy of 99.80% with the ReLU function. In contrast, when the Adam optimizer is employed, the Leaky ReLU configuration again yielded a superior performance across all datasets with accuracies of 97.21%, 95.54%, and 99.84% on HAR70+, HARTH, and MHealth datasets, respectively. It is also observed that ReLU achieved a marginally better accuracy of 99.88% on the MHealth dataset.

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 8. Accuracy attained by the proposed method with LSTM and RMSProp (With Augmentation)

Activation Function	Datasets		
	HAR70+	HARTH	MHealth
ReLU	96.32	94.53	99.80
Leaky ReLU (0.1)	96.67	94.74	99.72

Table 9. Accuracy attained by the proposed method with LSTM and Adam (With Augmentation)

Activation Function	Datasets		
	HAR70+	HARTH	MHealth
ReLU	96.62	93.90	99.88
Leaky ReLU (0.1)	97.21	95.54	99.84

BiLSTM without Augmentation using Different Activation Functions and Optimizers

Without the augmentation technique, the BiLSTM-based configuration's performance is summarized in Tables 10 and 11. While leveraging the RMSProp optimizer, Leaky ReLU attained better accuracies of 94.56, 94.88, and 98.37% on HAR70+, HARTH, and MHealth datasets, respectively, compared to the ReLU function. In contrast, the Adam optimizer yielded mixed outcomes. Although Leaky ReLU still performed well on HAR70+ and HARTH datasets with accuracies of 92.23% and 95.07%, respectively, it underperformed slightly on the MHealth dataset with an accuracy of 97.90% compared to the ReLU activation with an accuracy of 98.59%.

Table 10. Accuracy attained by the proposed method with BiLSTM and RMSProp (Without Augmentation)

Activation Function	Datasets		
	HAR70+	HARTH	MHealth
ReLU	91.46	94.66	98.21
Leaky ReLU (0.1)	94.56	94.88	98.37

Table 11. Accuracy attained by the proposed method with BiLSTM and Adam (Without Augmentation)

Activation Function	Datasets		
	HAR70+	HARTH	MHealth
ReLU	88.47	93.29	98.59
Leaky ReLU (0.1)	92.23	95.07	97.90

BiLSTM with Augmentation using Different Activation Functions and Optimizers

The performance of the proposed method with BiLSTM configuration with augmentation technique is presented in Tables 12 and 13 with different optimizers and activation functions. As depicted in the tables, the highest accuracy of 97.34% on the HAR70+ dataset is achieved with Leaky ReLU activation and the Adam optimizer, slightly outperforming the ReLU counterpart. Similarly, on the HARTH dataset, Leaky ReLU with Adam achieved better performance with an accuracy of 95.30%, while ReLU with RMSProp achieved a comparable accuracy of 94.37%. For the MHealth dataset, the best accuracy of 99.77% is obtained with Leaky ReLU and Adam, outperforming ReLU with Adam at 96.73%.

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 12. Accuracy attained by the proposed method with BiLSTM and RMSProp (With Augmentation)

Activation Function	Datasets		
	HAR70+	HARTH	MHealth
ReLU	97.04	94.37	99.42
Leaky ReLU (0.1)	94.34	90.96	98.72

Table 13. Accuracy attained by the proposed method with BiLSTM and Adam (With Augmentation)

Activation Function	Datasets		
	HAR70+	HARTH	MHealth
ReLU	97.01	95.20	96.73
Leaky ReLU (0.1)	97.34	95.30	99.77

From the implementations of the proposed method with different optimizers and activation functions and comparative analysis, key findings are summarized as follows:

1. The proposed attention-deep-based HAR method achieved superior performance on three benchmark datasets – 97.21% on HAR70+, 95.54% on HARTH, and 99.84% on MHealth. These results demonstrate the proposed method's good efficiency and generalization capabilities across sensor-based activity recognition tasks.
2. The combination of ConvAE, self-attention mechanism, and LSTM layers enabled the model to learn richer and more discriminative feature representations comprising both local and long-range temporal dependencies.
3. From the results obtained, it is observed that the augmentation technique significantly improved performance on imbalanced datasets.
4. The comparative analysis shows that the proposed method outperformed existing DL and attention-based HAR frameworks in the case of the HARTH and MHealth datasets.
5. From the ablation study, it can be inferred that there is a slight fall in the proposed method's performance while utilizing the BiLSTM configuration. Also, it is noted that the Leaky ReLU with Adam optimizer delivered the best results.

CONCLUSIONS

In this work, we proposed an attention-DL-based lightweight WHAR framework. The framework comprises three steps: pre-processing, training, and classification. In the pre-processing step, a scaling augmentation technique is employed to tackle the imbalanced nature of datasets. The sliding window technique is applied to convert the raw accelerometer readings into frames with a fixed window size of 50 and an overlap of 10. For the extraction of features, the frames were initially passed to ConvAE with an average pooling layer as a bottleneck. The extracted features undergo a self-attention mechanism emphasizing relevant, informative features, followed by two LSTM layers to capture long-term dependencies and deeper feature representations. The extracted features are processed through four fully connected layers to classify activities from sensor-based data. The proposed method is validated on three publicly available datasets: HAR70+, HARTH, and MHealth. The proposed method demonstrated good performance with accuracies of 97.21%, 95.54% and 99.84% on the above-mentioned datasets, respectively.

For the evaluation of the efficiency of the proposed method, a comparative analysis is performed for each dataset. An ablation study is presented for the study of the impact on the proposed method with the introduction of BiLSTM layers in place of LSTM layers, augmentation techniques with two optimizers and activation functions- RMSProp, Adam, ReLU and Leaky ReLU. From the attained results, it is clearly seen that with the introduction of augmentation techniques, there is a significant improvement in the performance of the proposed method with both LSTM and BiLSTM configurations. It can also be inferred that the model with the Leaky ReLU activation function and Adam optimizer provides superior performance compared to RMSProp and ReLU functions. The plotted graphs of the

above-mentioned implementations indicate the presence of minimal overfitting. Future directions include integrating multiple sensor modalities and exploring the proposed method with different attention mechanisms and temporal modelling algorithms.

REFERENCES

[1] Gomez N., Pato M., Lourenco A. R., and Datia N. (2023) A survey on wearable sensors for mental health monitoring. Sensors, 23(3), 1330.
[2] Dang L. M., Min K., Wang H., Piran M. J., Lee C. H., and Moon H. (2020) Sensor-based and vision-based human activity recognition: a comprehensive survey. Pattern Recognition, 108, 107561.
[3] Anike C. V. et al. (2022) Mobile and wearable sensors for data-driven health monitoring system: state-of-the-art and future prospect. Expert Systems with Applications, 202, 117362.
[4] Serpush F., Menhaj M. B., Masoumi B., and Karasfi K. (2022) Wearable sensor-based human activity recognition in the smart healthcare system. Computational Intelligence Neuroscience, 2022(1), 1391906.
[5] Saha A., Rajak S., Saha J., and Chowdhury C. (2024) A survey of machine learning and meta-heuristics approaches for sensor-based human activity recognition systems. Journal of Ambient Intelligence and Humanized Computing, 15(1), 29-56.
[6] Ramanujam E., Perumal T., and Padmavathi S. J. I. (2021) Human activity recognition with smartphones and wearable sensors using deep learning techniques: a review. IEEE Sensors Journal, 21(12), 13029-13040.
[7] Chen K. et al. (2021) Deep learning for sensor-based human activity recognition: overview, challenges, and opportunities. ACM Computing Surveys, 54(4), 1-40.
[8] Dua N. et al. (2022) A survey on human activity recognition using deep learning techniques and wearable sensor data. In International Conference on Machine Learning, Image Processing, Network Security and Data Sciences, Springer Nature, 52-71.
[9] Tao S., Goh W. L., and Gao Y. (2023) A convoluted self-attention model for IMU-based gait detection and human activity recognition. In 2023 IEEE 5th International Conference for Artificial Intelligence Circuits and Systems (AICAS) IEEE, 1-5.
[10] Zheng G. (2021) A novel attention-based convolution neural network for human activity recognition. IEEE Sensors Journal, 2(23), 27015-27025.
[11] Mekruksavanich S., Jantawong P., Phaphan W., and Jitpattanakul A. (2024) Hybrid attention with CNN-BiLSTM and CBAM for efficient wearable activity recognition. In 2024 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT and NCON), 572-576.
[12] Chen X. et al. (2024) A novel CNN-BiLSTM ensemble model with attention mechanism for sit-to-stand phase identification using wearable inertial sensors. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 32, 1068-1077.
[13] Gupta S. (2021) Deep learning based human activity recognition (HAR) using wearable sensor data. International Journal of Information Management Data Insights, 1(12), 100046.
[14] Thu N. T H., and Han D. S. (2021) HiHAR: a hierarchical hybrid deep learning architecture for wearable sensor-based human activity recognition. IEEE Access, 9, 145271-145281.
[15] Luwe Y. J., Lee C. P., and Lim K. M. (2022) Wearable sensor-based human activity recognition with hybrid deep learning model. Informatics, 9(3), 56.
[16] Chandramouli A. et al. (2024) Enhanced human activity recognition in medical emergencies using a hybrid CNN and bi-directional LSTM model with wearable sensors. Scientific Reports, 14(1), 30979.
[17] Nazar F., and Jalal A. (2024) Wearable sensor-based activity recognition over statistical features selection and MLP approach. In 2024 3rd International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE), IEEE, 1-7.
[18] Nithin G. R. et al. (2021) Sensor-based human activity recognition for elderly in-patients with a luong self-attention network. In 2021 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 97-101.

[19] Al-Qaness M. A., Dahou A., Abd Elaziz M., and Helmi A. M. (2022) Multi-ResAtt: Multilevel residual network with attention for human activity recognition using wearable sensors. IEEE Transactions on Industrial Informatics, 19 (1), 144-152.
[20] Zhang H., and Xu L. (2024) Multi-STMT: multi-level network for human activity recognition based on wearable sensors. IEEE Transactions on Instrumentation and Measurement, 73, 1-12.
[21] Tang Y., Zhang L., Min F., and He J. (2022) Multiscale deep feature learning for human activity recognition using wearable sensors. IEEE Transactions on Industrial Electronics, 20(2), 2106-2116.
[22] Ullah S., Pirahandeh M., and Kim D. H. (2024) Self-attention deep ConvLSTM with sparse-learned channel dependencies for wearable sensor-based human activity recognition. Neurocomputing, 571, 127157.
[23] AbdelRaouf H., Abouyoussef M., and Ibrahem M. I. (2024) An innovative approach for human activity recognition based on a multi-head attention mechanism. In 2024 International Conference on Machine Learning and Applications (ICMLA),1559-1563.
[24] Thakur D., Biswas S., Ho E. S., and Chattopadhyay S. (2022) Convae-lstm: convolutional autoencoder long short-term memory network for smartphone-based human activity recognition. IEEE Access, 10, 4137-4156.
[25] Olah A. (2015) Understanding LSTM Networks. Colah's Blog [Online] Available at https://colah.github.io/posts/2015-08-Understanding-LSTMs/.
[26] Guo M. H. et al. (2022) Attention mechanism in compute vision: a survey. Computer Vision Media, 8(3), 33-368.
[27] Ustad A. et al. (2023) Validation of an activity type recognition model classifying daily physical behaviour in older adult: the HAR70+ model. Sensors, 23(5), 2368.
[28] Logacjov A., Bach K., Kongsvold A., Bårdstu H. B., and Mork P. J. (2021) HARTH: a human activity recognition dataset for machine learning. Sensors, 21(23), 7853.
[29] Banos O. et al. (2014) mHealthDroid: a novel framework for agile development of mobile applications. Ambient Assisted Living and Daily Activities, 88868(14), 91-98.
[30] Logacjav A., Herland S., Ustad A., and Bach K. (2023) SelfPAB: large scale pre-training for dual-accelerometer human activity recognition.
[31] Mekruksavanich S., Phaphan W., and Jitpattanakul A. (2024) Harnessing deep learning for activity recognition in seniors' daily routines with wearable sensors. In 2024 47th International Conference on Telecommunications and Signal Processing (TSP), IEEE, 164-167.
[32] Sahoo K. K. et al. (2023) Wrapper-based dee feature optimization for activity recognition in the wearable sensor networks of healthcare systems. Scientific Reports, 13(1), 965.
[33] Sharma A. et al. (2023) A hybrid deep learning-based approach for human activity recognition using wearable sensors. In Innovations in Machine and Deep Learning: Case Studies and Applications, Springer, 231-259.
[34] Khan S., Noorani S. H., Arsalan A., Mahmood A., Rauf U., and Ali Z. (2023) Classification of human physical activities and postures during everyday life. In 2023 18th International Conference on Emerging Technologies (ICET), IEEE, 98-103.
[35] Paul A., Khan S., Mondal D., and Singh P. K. (2025) Recognizing human activities in ambient assisted environment from wearable sensor data using gramian angular field and deep CNN. In Enabling Person-Centric Healthcare Using Ambient Assistive Technology, Volume 2: Personalized and Patient-Centric Healthcare Services in AAT, Springer, 199-226.
[36] Varshney N., Bakariya B., and Kushwala A. K. S. (2022) Human activity recognition using deep transfer learning of cross position sensor based on vertical distribution of data. Multimedia Tools and Applications, 81(16), 22307-22322.
[37] Khatun M. A. et al. (2022) Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor. IEEE Journal of Translational Engineering in Health and Medicine, 10, 1-16.
[38] Davidashvilly S., Hssayeni M., Chi C., Jimenez-Shahed J., and Ghoraani B. (2022) Activity recognition in Parkinson's patients from motion data using a CNN model trained by healthy subjects. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 3199-3202.