Received: 27 August 2022 | Revised: 8 November 2022 | Accepted: 29 November 2022

DOI: 10.1111/coin.12565

ORIGINAL ARTICLE

Computational Intelligence WILEY

A lightweight deep learning with feature weighting for activity recognition

Ayokunle Olalekan Ige | **Mohd Halim Mohd Noor**

School of Computer Sciences, Universiti Sains Malaysia, George Town, Malaysia

Correspondence
Mohd Halim Mohd Noor, School of Computer Sciences, Universiti Sains Malaysia, 11800 George Town, Pulau Pinang, Malaysia.
Email: halimnoor@usm.my

Abstract

With the development of deep learning, numerous models have been proposed for human activity recognition to achieve state-of-the-art recognition on wearable sensor data. Despite the improved accuracy achieved by previous deep learning models, activity recognition remains a challenge. This challenge is often attributed to the complexity of some specific activity patterns. Existing deep learning models proposed to address this have often recorded high overall recognition accuracy, while low recall and precision are often recorded on some individual activities due to the complexity of their patterns. Some existing models that have focused on tackling these issues are always bulky and complex. Since most embedded systems have resource constraints in terms of their processor, memory and battery capacity, it is paramount to propose efficient lightweight activity recognition models that require limited resources consumption, and still capable of achieving state-of-the-art recognition of activities, with high individual recall and precision. This research proposes a high performance, low footprint deep learning model with a squeeze and excitation block to address this challenge. The squeeze and excitation block consist of a global average-pooling layer and two fully connected layers, which were placed to extract the flattened features in the model, with best-fit reduction ratios in the squeeze and excitation block.

The squeeze and excitation block served as channel-wise attention, which adjusted the weight of each channel to build more robust representations, which enabled our network to become more responsive to essential features while suppressing less important ones. By using the best-fit reduction ratio in the squeeze and excitation block, the parameters of the fully connected layer were reduced, which helped the model increase responsiveness to essential features. Experiments on three publicly available datasets (PAMAP2, WISDM, and UCI-HAR) showed that the proposed model outperformed existing state-of-the-art with fewer parameters and increased the recall and precision of some individual activities compared to the baseline, and the existing models.

KEYWORDS

activity recognition, deep learning, wearable

1 | INTRODUCTION

Advancements in pervasive computing have seen an increase in the interest of researchers in human activity recognition. Human activity recognition has numerous application areas, such as; elder healthcare,[1] child monitoring,[2] rehabilitation monitoring, and general well-being,[3] among other areas. Generally, activity data can be collected using the vision-based method or sensor-based method. The vision-based method involves placing a capturing device such as a camera in a strategic position to capture human movement and activities. However, this method may not work in situations where continuous monitoring of a person's activities is essential. Also, cameras are intrusive, and many individuals are uneasy about being continually watched. These limitations of the vision-based method have prompted the adoption of wearable sensors for capturing human activity data. Wearable sensors use inertia measurement units and Radio Frequency Identifications to collate human activities. Wearable sensors are worn directly on subjects, and they can be incorporated into clothing, eyewear, wristwatches, mobile devices, or placed on the body directly.[4] They are unaffected by the surroundings and have the potential to improve accuracy. Furthermore, wearable sensors cannot put users' privacy at risk. As a result, wearable sensors are better for recognizing human activities.[5,6] However, the recognition task is difficult due to the vast number of sensor modalities, noisy data, variances in the spatial and temporal dimensions of the feature space between people,[7] and also, the variability when a subject or different subjects perform the same task at various times, among other factors.[8] Examples of wearable sensors include accelerometers, gyroscopes, magnetometers, and others. Various machine learning models have been proposed to recognize the activities collected using these sensors; an example of such can be seen in Sani et al.[9] where the authors used K-Nearest Neighbor for

classification. Others include Random Forest,[10] Support Vector Machine (SVM),[11] and Decision Tree,[12] among many others. Before machine-learning techniques can be used for classification, features of the data must be extracted. Since the process of extracting features manually from wearable sensor data is laborious,[13] recent HAR researchers have adopted deep learning for activity classification.

Deep learning models can automatically extract features from wearable sensor datasets[14] since it allows the model to learn all layers of representation greedily. For activity recognition, researchers have proposed several deep learning models. Even though these deep learning models have achieved state-of-the-art recognition of the overall wearable sensor data, most researchers have focused on achieving effective feature extraction and high overall classification accuracy without concentrating on some activities' individual recall and precision. This is because the recall and precision of some individual activities are sometimes low, a challenge attributed to less training data, interclass similarity and the complexity of the pattern of such activities.[15] The precision measures the ability of a model to distinguish a particular activity class from all the other classes, while the recall shows the performance of the model in recognizing an activity class. To the best of our knowledge, attention mechanism has been proposed in the literature to address this challenge, as seen in Ma et al.[16] where the authors proposed a GRU with attention mechanism, also in Zeng et al.[17] where the researchers proposed long short-term memory (LSTM) with attention, and in Gao et al.[18] where convolutional neural network (CNN) was proposed with Dual attention. However, the bulky size of these models and the low individual recall and precision of some activities have made this area open to research contribution. Also, since most embedded systems have resource constraints in terms of their processor, memory and battery capacity, it is therefore, paramount to propose efficient, lightweight activity recognition models that require limited resource consumption. To tackle this, our research proposes a high performance, low footprint predictive model for activity recognition. The model is designed using a one-dimensional CNN with varying kernel size and a squeeze and excitation block[19] with best-fit reduction ratio, placed after the flattened feature map. The squeeze and excitation (SE) block is added to automatically rescale the weights of all features according to their relevance. The justification for the proposed model is based on four proven facts. Firstly, one-dimensional CNN has been shown to perform better on multivariate time series data.[13,20,21] Secondly, not all data features are of the same relevance for modeling,[22] and since CNN treats all features equally, the SE block will ignore irrelevant features and prioritize relevant and pertinent ones. Thirdly, the difficulty and cost of training RNN-based architectures, where the densely arranged trainable parameters in neuron units are challenging to tune made it paramount to propose models without RNN for Activity Recognition.[23] Lastly, the flatten layer reduces feature maps to a single one-dimensional vector; adding a SE Block after this layer can reduce the model size while focusing on essential features. Our main contribution is in three folds:

1. Firstly, we present a lightweight and efficient model for activity classification using one-dimensional CNN with varying kernel sizes and SE Block with best-fit reduction ratio to boost the importance of the discriminative features.
2. Secondly, we experimented on three public datasets to achieve overall state-of-the-art recognition performance.
3. Thirdly, we addressed the issues of low recall and precision on individual activities with complex patterns and overall performance using a smaller deep learning model with fewer parameters.

The remainder of this paper is organized as follows: Section 2 describes the related works, Section 3 presents the research methodology, Section 4 presents the experimental results, and Section 5 concludes.

2 | RELATED WORKS

Recent research on activity recognition has employed wearable sensors for activity data collection, with the most common sensor device being an accelerometer.[24,25,26] An accelerometer is a device that measures the acceleration of an entity. Other sensor devices include gyroscopes, magnetometers, electrocardiography monitors, and many others.[27] Some researchers have recently proposed HAR systems developed through sensor data obtained through smartphones. An example of this is in Ronao and Cho,[28] where the authors utilized the inherent properties of one-dimensional activity signals for activity recognition. The typical process of wearable sensor-based activity recognition consists of three critical stages: data segmentation, feature extraction, and activity classification.

Some researchers have proposed machine learning techniques to classify activities. For example, in Trost et al.[29] seven activities were recognized. The authors used a sensor worn on the wrist and the hips to collect activity data and used logistic regression as the classifier. Even though machine-learning methods functioned effectively in wearable sensor activity recognition, the need to achieve state-of-the-art and address the relative bottlenecks of the machine learning approach has led to the adoption of deep learning for wearable sensor activity recognition. Deep learning has been successful in a variety of fields such as image segmentation,[30] image feature extraction,[31] classification,[32] object detection,[33] and sentiment analysis,[34] among other areas. Deep learning models are generally capable of extracting features from wearable sensor datasets automatically[14] since it enables the model to learn all layers of representation jointly at the same time. Researchers for activity recognition have proposed numerous deep learning approaches. For example, Jantawong et al.[35] proposed a sensor-based HAR to classify high-performance activities using a deep learning model called the InceptTime network. The model was tested on the PAMAP2 dataset and achieved a classification accuracy of 88%. The following section describes some other literature that used deep learning models for activity recognition.

2.1 | Convolutional neural network models

Convolutional neural networks (CNN) have been the most adopted deep learning approach for automatic feature extraction and classification in activity recognition. For example, the authors in Reference 36 employed a single wrist-worn accelerometer to recognize five different physical activities: sitting, standing, lying, walking, and running. Similarly, a wrist-worn accelerometer was used in Reference 37 to distinguish eight different activities. Gomathi et al.[38] developed a Fuzzy associator rule-based fuzzified deep convolutional neural network architecture to classify wearable sensor-based human activity recognition. The lambda max method was fused for weight initialization to ensure data normalization and faster convergence. Experiments were carried out on the UCI HAR dataset, consisting of six activities. The model achieved overall recognition accuracy of 97.89%; however, the accuracy on some individual activities, such as walking upstairs and walking, was relatively low. In Rueda et al.[39] a CNN model that processes each wearable sensor data separately was proposed. The model experimented on two publicly available datasets and

an industrial dataset. The model achieved an improved recognition accuracy on a few activities. However, some activities with complex patterns were not well classified, and the model's size is a constraint.

In Qi et al.[40] the authors suggested a fast and reliable deep convolutional neural network for wearable sensor data. The model included a series of signal processing algorithms and a signal selection module to increase accuracy and extend the content of the acquired raw data from the accelerometer, gyroscope, and magnetometer data. Experiments on the collected dataset achieved a classification accuracy of 95.27%. However, individual accuracy on each activity showed that out of the 12 classified activities, the model was only able to achieve accuracies higher than 90% in only 7 of the activities. At the same time, some had a recognition accuracy as low as 50.60% and 75.50%. Also, in Huang et al.[15] the authors focused on overcoming the problem of low accuracy when walking upstairs and downstairs and proposed a two-stage end-to-end convolutional neural network. The model improved recognition accuracy on the two activities compared to a single-stage CNN. However, since wearable sensors are often in time series format, the long-term dependency of the time series is difficult to extract with CNN, making it difficult to increase the model's performance.[41] For this reason, some researchers have proposed hybrid deep learning models using Recurrent Neural Networks for activity recognition to achieve improved classification accuracy on individual activities with complex patterns. The following section presents these works.

2.2 | Hybrid models

Several researchers have proposed adopting convolutional neural networks with recurrent neural networks. For example, Dua et al.[42] proposed a multi-input CNN with Gated recurrent units (GRU). The model concatenated three CNN-GRU architectures for classification, benchmarked on the PAMAP2, UCI-HAR, and WISDM datasets, and achieved 95.27%, 96.20%, and 97.21% accuracies, respectively. However, the confusion matrix of the model on the three datasets showed that some activities achieved low recognition performance, based on their precision, recall, and F1 score, despite the size of the model. Challa et al.[43] proposed CNN with Bidirectional-LSTM (Bi-LSTM) to classify multi-activities on the PAMAP2 dataset and achieved recognition accuracy of 94.29%. However, some activities such as rope jumping, ironing, and ascending stairs still achieved low recognition accuracy in the range of 80%. In Li et al.[41] the authors introduced residual block and Bi-LSTM. The model worked by using the residual block to automatically extract spatial features from multidimensional inertial sensor inputs, then used Bi-LSTM to retrieve the forward and backward dependencies of the feature sequence before feeding the features into the Softmax layer for classification. Also, the recognition accuracy of this model on some individual activities was low. In Noor et al.[44] a Conv-LSTM model which utilized the temporal features of sensor-based activity recognition data together with sliding window relationship was proposed. The model concatenated various window features, then used a sequence-learning module to learn temporal features, and achieved an accuracy of 91.6% on the benchmarking dataset. In Ullah et al.[45] the authors proposed a stacked LSTM network for human activity recognition. The model achieved an accuracy of 93.13%, but very low accuracy was also recorded on some individual activities. Nafea et al.[46] proposed a deep CNN with many convolutional layers and varying kernel sizes, and Bi-LSTM. Experiments were performed on the WISDM and UCI-HAR datasets, comprising six activities each, and accuracies of 98.53% and 97.05% were recorded, respectively. Even though the model was bulky and achieved high overall recognition accuracy, results showed that

the model recorded low precision and accuracy when classifying walking downstairs, walking, and walking upstairs on the two datasets.

The performance evaluation of LSTM against Bi-LSTM for activity recognition using wearable sensor data was done in Reference 41, and the result showed that the Bi-LSTM achieves more recognition accuracy. However, the issue of low precision and recall of some individual activities, which is often caused by such activity's complex pattern, was not addressed. Recent researchers have proposed an attention mechanism to improve activity classification in wearable sensor-based activity recognition to tackle this issue.

The attention mechanism works in parallel and can perceive the entire chain of events. The attention mechanism's fundamental notion is to assign varying weights to different types of information. As a result, giving relevant data a more significant weight draws the deep learning model's attention to it.[19] The formal notion of attention mechanism is that weights are learned over k vectors, each of dimension d, using a collection of linear layers and non-linearities. The dimension of these d-dimensional vectors is mapped to a one-dimensional score in most architectures, and these scores are then passed through the Softmax function to yield the set of k weights. Recent activity recognition research has adopted an attention mechanism to improve recognition accuracy. In Murahari,[22] a DeepConvLSTM with attention was proposed to extend the work in Ordóñez,[47] and a relative improvement of 87.5% was recorded on the PAMAP2 dataset using the model with attention, and an accuracy of 74.8% on the model without attention. Wang et al.[48] used a three-CNN architecture that consisted of five one-dimensional (1D) convolutional layers and three max-pooling layers. Following the convolution procedure, three attention submodules were employed to give high weights to significant features and low weights to irrelevant features in order to maximize SoftMax classification capabilities. On the UCI HAR dataset and a weakly labeled dataset, the model achieved a recognition accuracy of 93.83%, with low precision and recall in walking and going downwards. Gao et al.[18] proposed an attention mechanism for HAR using CNN and Recurrent Neural Network. The model used channel and temporal attention to increase feature learning and classification performance of wearable sensors. The model was evaluated on five datasets, and the result showed an increased performance when compared to baseline models and was able to achieve a recognition accuracy of 92.45% and 93.16% on the PAMAP2 dataset using CNN with attention and residual network with attention, respectively. Although the model achieved a high recognition accuracy on the datasets; however, unlike our proposed model, three full architectures were connected to achieve this performance, which made the model size bulky and complex compared to our model. And also, after 500 epochs, the precision achieved on some activities with complex patterns, such as vacuum cleaning (80%), was relatively low.

Also, in Abdel-Basset,[49] a dual-channel model was proposed. The model comprised LSTM layers with attention and convolutional layers with a SE block. The model was tested on two publicly available datasets in UCI-HAR and WISDM. Experiments showed that the model achieved 97.70% and 98.90% recognition accuracy on the UCI-HAR and WISDM datasets. The result showed that walking upstairs and downstairs had the lowest precision and recall. Unlike our present study, three (3) SE blocks were stacked with CNN and LSTM, making the model size bulky. Also, the model was trained for 250 epochs. Also, in Liu et al.[50] the researchers combined bidirectional long short-term memory (BiLSTM) with an attention mechanism for human activity recognition. The result showed that the model outperformed existing models. However, some activities achieved low individual recall and precision. In summary, the limitations of the presented related works can be summarized as follows; CNN models have limitations in learning discriminative features, since they treat all features as equally important. Also, such deep neural networks often come with increased computational cost. RNN and hybrid models have limitations in terms of

computational time, thereby limiting their edge inference applicability. Also, existing models with attention mechanisms often come with high model complexity, due to the number of features that are passed to the fully connected layers.

To tackle this, our proposed model connected 1D convolutional layer to fully connected SE block. Similar to the work in Gao et al.[18] where temporal and channel-wise attention was used together with 2D convolutions with same kernel size and batch normalization in three concatenated architectures, our proposed model used channel-wise attention with standalone 1D convolutional layers architecture with varying kernel sizes, batch normalization, max-pooling layers, and best-fit SE reduction ratio. By leveraging on the best-fit SE reduction ratio, the parameters of the fully-connected layers were reduced, which helped the model increase the responsiveness to the essential features with a smaller model size, and increased the recognition accuracy, precision, and recall of some individual activities with complex patterns.

3 | METHODOLOGY

This section presents the method adopted in the signal segmentation phase, data pre-processing, the description of the proposed model, and the model parameters. The block diagram of the overall approach is presented in Figure 1.

3.1 | Signal segmentation

In sensor-based activity recognition, the activity signals need to be segmented into a sequence of windows for classification. In this study, the fixed sliding window is adopted whereby the signals are segmented into windows of fixed size with a degree of overlap. Let $D = \left\{\left(\boldsymbol{x}^1 y_c^1\right), \left(\boldsymbol{x}^2 y_c^2\right), \ldots, \left(\boldsymbol{x}^N y_c^N\right)\right\}$ denotes the activity dataset, which is the sequence of N windows with each window is labeled with activity class y_c that is performed during the interval where $c = 1, \ldots, C$. Let $\boldsymbol{x}^i \in \mathbb{R}^{w \times K}$ denotes an i-th window with K sensor channels. The data of the k-th channel of a sensor is given as follows:

$$\boldsymbol{x}_{\boldsymbol{k}}^{i} = [x_1, x_2, \ldots, x_w] \tag{1}$$

where w is the size of the window segmentation.

3.2 | Proposed model

In general, the issue with CNN is that they treat each feature channel equally. To get around this, we solve the problem by adding a SE block that serves as a channel-wise attention. This allowed every feature channel to be re-weighted and as a result of this, improved the accuracy of the features. In our proposed model, we first built a compressed representation for the features using 1D convolutional layers with kernel sizes of 5, 7, 9, and 11, batch normalization layer, and max-pooling layer. Then the features are flattened to reduce the feature maps to a single one-dimensional vector, with a 0.5 dropout rate after the flattening operation. In the second step, the responsiveness of the extracted features is recalibrated according to their relevance in predicting the activities. This is achieved by adopting the SE block to model the relationship

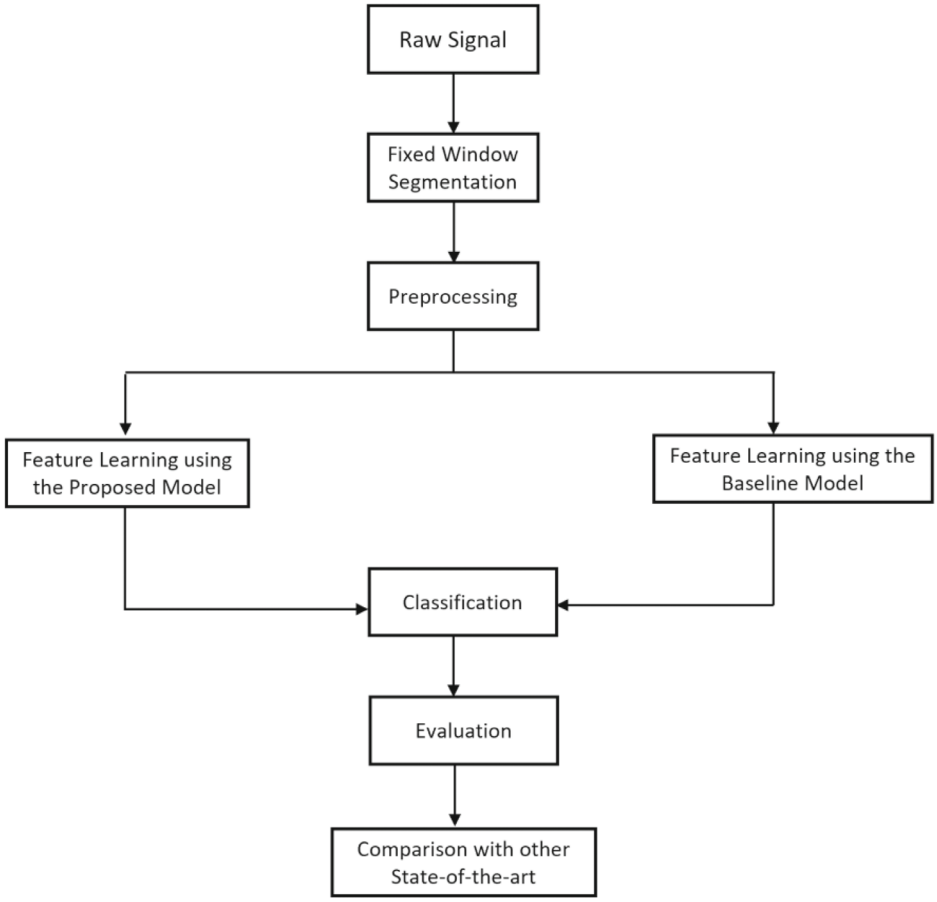

FIGURE 1 Block diagram of the proposed approach

between the feature maps. The SE block consists of a Global Average Pooling layer to perform global average pooling on each flattened feature map and sum out the spatial features. Then, the features are reduced using varying reduction ratio and later extended using a 2-layered fully connected network with rectified linear unit (ReLU) activation function in the first and sigmoid in the next. This produces a feature vector used to scale the feature channels. Figure 2 illustrates an overview of the proposed model.

3.2.1 | Squeeze and excitation block for feature weighting

Given an input window $\boldsymbol{x}$ that is fed to the proposed model to produce feature maps $\boldsymbol{u} \in \mathbb{R}^{L \times D}$. The weightage of the feature maps is calculated via several operations. First, global-average pooling is used to produce channel-wise statistics. The statistic $\boldsymbol{z} \in R^D$ is generated by squeezing $\boldsymbol{u}$ through its temporal dimensions L. Then the d-th element of $\boldsymbol{z}$ is given as:

$$z_d = F_{sq}(\boldsymbol{u}_d) = \frac{1}{L}\sum_{j=1}^{L} u_d^j \tag{2}$$

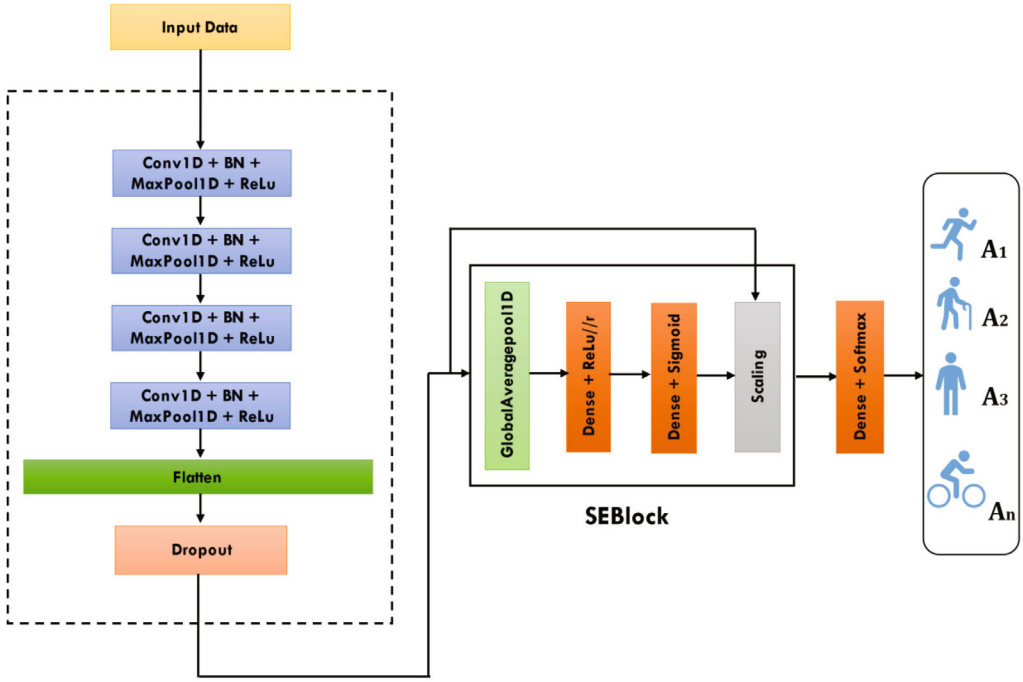

FIGURE 2 Overview of the proposed model

where F_{sq} is given as the squeeze function and L is the length of the feature maps. The aggregated information acquired using the squeeze operation is then passed to the excitation operation to capture channel-wise dependencies using a gating mechanism with a sigmoid activation function, given as Reference 49:

$$\boldsymbol{s} = F_{ex}(\boldsymbol{z}, W) = \sigma(g(\boldsymbol{z}, W)) = \sigma\left(W_2 \delta\left(W_1 \boldsymbol{z}\right)\right) \tag{3}$$

where σ is the Sigmoid activation function and δ is the ReLU activation function. $W_1 \in \mathbb{R}^{D/r \times D}$, $W_2 \in \mathbb{R}^{D \times D/r}$ and r is the reduction ratio. $\boldsymbol{s}$ is a vector of size equal to the number of feature maps. Thus, the values can be interpreted as the weights indicating the importance of the feature maps. Using $\boldsymbol{s}$, the feature map $\boldsymbol{X}$ is rescaled as follows:

$$F_{(scale)}\left(u_d, s_{\mathrm{d}}\right) = s_d, u_d \tag{4}$$

where $F_{(scale)}\left(u_d, s_d\right)$ is the channel-wise multiplication of a scalar s_d and feature map u_d.

3.2.2 | Model parameters

As shown in Table 1, the input stage takes in the data input and passes to the one-dimensional convolutional layer with batch normalization; since research[51] has proved that batch normalization can minimize the number of training steps necessary for model convergence and to allow for a higher learning rate without focusing on initialization parameters and dropout rate. As a result, a batch normalization layer simplifies and speeds up the model's training. The activation

TABLE 1 Architecture summary of the proposed model

Layer	Configuration	Output shape
Input		171 × 36
1D Conv	Kernel size = 5, Activation = ReLU	171 × 32
Batch normalization	—	171 × 32
1D max-pooling	Pool size = 2	57 × 32
1D Conv	Kernel size = 7, Activation = ReLU	57 × 64
Batch normalization	—	57 × 64
1D max-pooling	Pool size = 2	19 × 64
1D Conv	Kernel size = 9, Activation = ReLU	19 × 128
Batch normalization	—	19 × 128
1D max-pooling	Pool size = 2	6 × 128
1D Conv	Kernel size = 11, Activation = ReLU	6 × 256
Batch normalization	—	6 × 256
1D Max-pooling	Pool size = 2	2 × 256
Flatten	—	512
Dropout	Rate = 0.5	512
SEBlock (Dense)	Ratio, Activation = ReLU	85
SEBlock (Dense)	Activation = Sigmoid	512
Multiply	—	512
Dense	Activation = Softmax	12

layer uses a ReLu activation, and a one-dimensional max-pooling layer is included. Our SE block consists of two fully connected layers. The first layer has a ReLU activation function and varying reduction ratio. The second fully connected layer used a sigmoid activation function to compute probabilities before the multiply layer. The code of the model architecture is available online at https://github.com/AOige/SEConv1D.

4 | EXPERIMENTAL RESULTS

4.1 | Datasets

We investigated the most widely used wearable sensor-based datasets in HAR research within the last 5 years (2017–2021) to evaluate the proposed model. We conducted a thorough search on various libraries and databases such as Scopus, IEEExplore, Springer, Web of Science, and ACM, through google scholar using the following Boolean keywords; "Wearable HAR", "HAR OR Wearable sensor" "PAMAP2", "WISDM AND HAR", "mHealth AND HAR AND dataset", "UCI-HAR", "UniMib-Shar AND Dataset".

The result showed that the PAMAP2,[52] WISDM,[53] and UCI HAR[54] datasets are the most used datasets, appearing in 442, 419, and 297 activity recognition research papers respectively

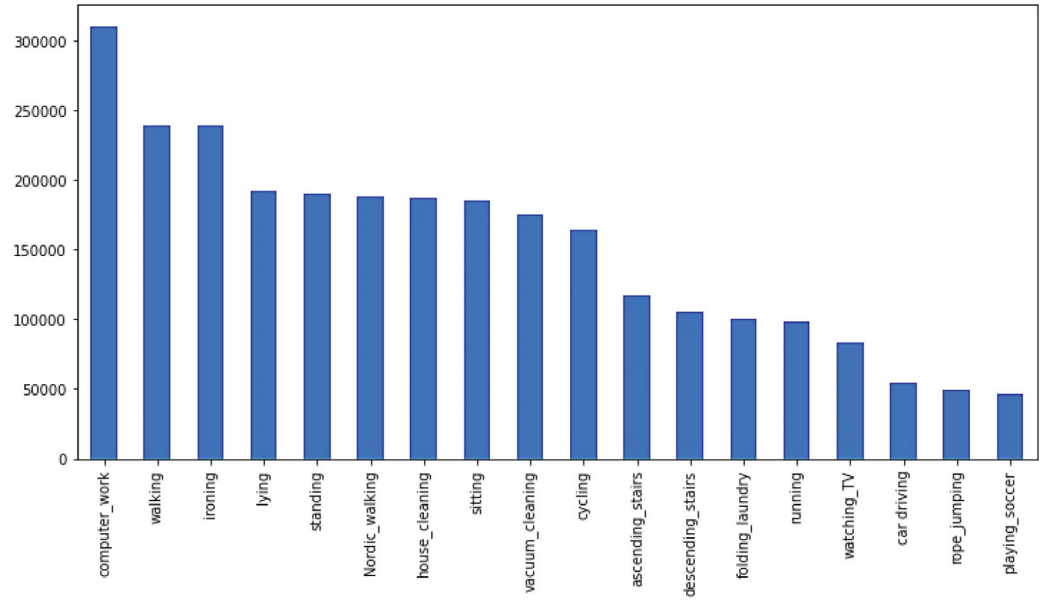

FIGURE 3 PAMAP2 data distribution

between 2017 and 2021. UniMib-SHAR and mHealth appear 146 and 138 times respectively. For this reason, the proposed model was benchmarked on these three datasets.

4.1.1 | PAMAP2 dataset description

The PAMAP2 dataset,[52] has nine participants, who were all required to participate in 12 protocol activities (including six [6] optional activities). The dataset consists of nine participants' basic and complex activities. Basic activities include sitting, standing, running, descending stairs, ascending stairs, cycling, walking, and Nordic walking, with complex activities such as vacuum cleaning, computer work, car driving, ironing, folding laundry, house cleaning, playing soccer and rope jumping. The data sample count before segmentation is shown in Figure 3.

Gyroscopes, accelerometers, magnetometers, heart rate monitors, and temperature measurements were used for data collection. In this work, we considered the protocol activities, and used 36 features of 3 IMUs, with a sliding window size of 5.12 s, sampling rate of 33.3 Hz, and 78% overlap.

4.1.2 | WISDM dataset description

The WISDM dataset[53] is an activity recognition dataset gathered from 36 participants who go about their daily lives. Accelerometer data from three-axis was considered. The dataset consists of 6 activities: Walking, Sitting, Standing, Jogging, Ascending stairs, and Descending stairs. The data was collected at a 20 Hz sampling rate using a smartphone accelerometer sensor, sliding window size of 10 s, and 95% overlap. The distribution of sample numbers of each activity is shown in Table 2.

TABLE 2 Samples in WISDM using three-axis accelerometer data

S/N	Activity	Number of samples
1	Walking	2082
2	Jogging	1626
3	Upstairs	633
4	Downstairs	529
5	Sitting	307
6	Standing	247

TABLE 3 Samples in UCI-HAR dataset

S/N	Activity	Number of samples
1	Laying	1944
2	Standing	1906
3	Sitting	1777
4	Walking	1772
5	Walking upstairs	1544
6	Walking downstairs	1406

4.1.3 | UCI-HAR dataset description

The UCI-HAR[54] dataset was gathered through 30 participants participating in performing experiments. Each participant wore a smartphone around their waist and completed six different tasks (laying, standing, sitting, walking downstairs, walking upstairs, and walking). The accelerometer and gyroscope sensor data streams recorded the volunteers' activities at a sampling rate of 50 Hz. There are nine features in the raw time-series data (i.e., body acceleration, total acceleration, and gyroscope signals in all three directions). The data signals were segmented into 2.56 s windows with a 50% overlap, yielding a total of 10,299 samples. The training dataset consists of the activities of 21 volunteers, whereas the test dataset consists of the activities of 9 volunteers. There are 7352 samples in the training dataset and 2947 samples in the test set. The distribution of sample numbers of each activity is shown in Table 3.

4.2 | Implementation details

Keras with Python 3 on Google Colab with Tensorflow 2.7.0, 12 GB RAM, and GPU was utilized to build and train the network model. The filters were set at 32, 64, 128, and 256, with a kernel size of 5, 7, 9, and 11. 1D convolutional layer with batch normalization and ReLU activation function, and 1D max-pooling layer were stacked, with the SE block after the Flatten and Dropout layer. The model aims to reduce the categorical cross-entropy loss. An epoch of 100 was set, and an early stopping mechanism was used to stop the training once the model stopped improving. During experiments, ablation studies were done to determine the best fit SE ratio for each dataset, and the best performance was recorded. The hyper parameters are shown in Table 4.

TABLE 4 Hyper parameters of the proposed model

Hyper parameters	Details
Optimizer	Adam
Epoch	100
Batch size	PAMAP2–32, WISDM - 16, UCIHAR – 10
Learning rate	Initial learning rate = $1e^{-4}$
	Minimum learning rate = $1e^{-7}$
	Patience = 5
Model loss	Categorical cross entropy
	Early stopping patience = 20
Best-fit SE reduction ratio	PAMAP2–6, WISDM – 18 UCI-HAR – 28
Kernel size	5, 7, 9, 11

4.3 | Performance metric

Since the performance of the proposed model on each activity is to be evaluated, our performance metrics were based accuracy, individual precision, recall, and F1-Score. Since the datasets are imbalanced, and these are given as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 - Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{8}$$

where TP – true positive, TN – true negative, FP – false positive, and FN – false negative.

Accuracy is defined as the overall fraction of the activities correctly recognized, Precision is the ratio of positives predicted correctly to the total number of samples classified as positives, Recall is the ratio of accurately predicted positives to the actual number of positive samples, and F1-Score is the harmonic mean of recall and precision.

4.4 | PAMAP2 experiments

4.4.1 | Baseline on PAMAP2

Using a baseline model of Conv1D without the SEBlock, the model achieved an overall testing accuracy of 97.16% after training for 62 epochs. The baseline model's training accuracy and loss are shown in Figure 4, and the confusion matrix is shown in Table 5.

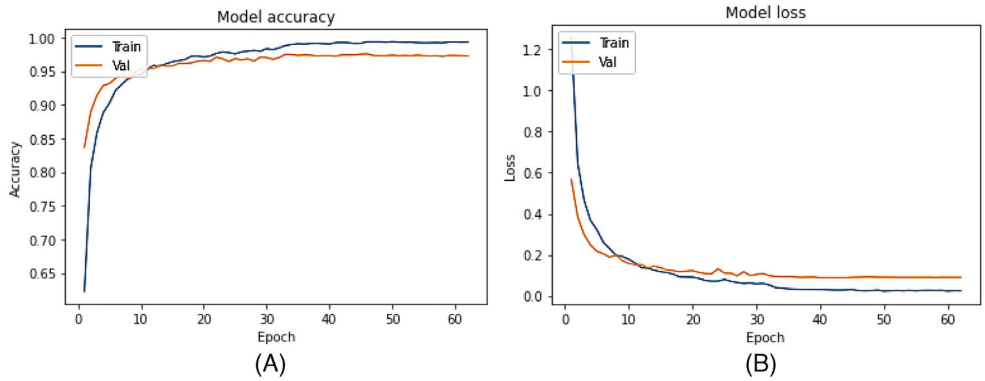

FIGURE 4 Training accuracy and loss of the baseline model on PAMAP2 (A) Accuracy versus epochs of the baseline model (B) Loss versus epochs plot of the baseline model

TABLE 5 Confusion matrix on baseline model

Activity	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
A1	308	1	0	0	0	0	0	2	0	2	0	0
A2	1	288	2	0	0	0	0	0	0	3	0	0
A3	0	1	307	0	0	0	0	0	0	0	1	0
A4	0	0	0	382	0	0	1	0	1	1	0	0
A5	0	0	0	1	157	0	0	0	0	0	0	0
A6	0	0	0	0	1	280	1	0	1	2	0	0
A7	0	0	0	0	0	0	296	0	1	0	0	1
A8	0	0	1	2	1	0	1	171	6	2	0	0
A9	0	0	1	2	0	0	1	4	154	3	0	0
A10	0	3	3	0	0	3	0	2	2	265	8	0
A11	0	0	10	0	0	2	0	0	1	0	403	0
A12	0	0	0	0	0	1	1	0	0	5	1	73

Abbreviations: A1, lying; A2, sitting; A3, standing; A4, walking; A5, running; A6, cycling; A7, Nordic walking; A8, ascending stairs; A9, descending stairs; A10, vacuum cleaning; A11, ironing; A12, rope jumping.

As shown in the confusion matrix of the baseline model in Table 5, 308 lying activities were correctly classified, with 1 misclassified as sitting, 2 as ascending stairs and 2 as vacuum cleaning. 288 sitting activities were correctly classified out of 304 samples, with 2 classified as standing, 3 as vacuum cleaning, and 1 as lying. However, activities such as Ascending stairs, descending stairs, vacuum cleaning, and rope jumping had high misclassifications.

4.4.2 | Proposed model on PAMAP2

By using the proposed model, the model trained for 69 epochs and achieved an overall testing accuracy of 97.76%. The model's training accuracy and loss are shown in Figure 5, and the confusion matrix is presented in Table 6.

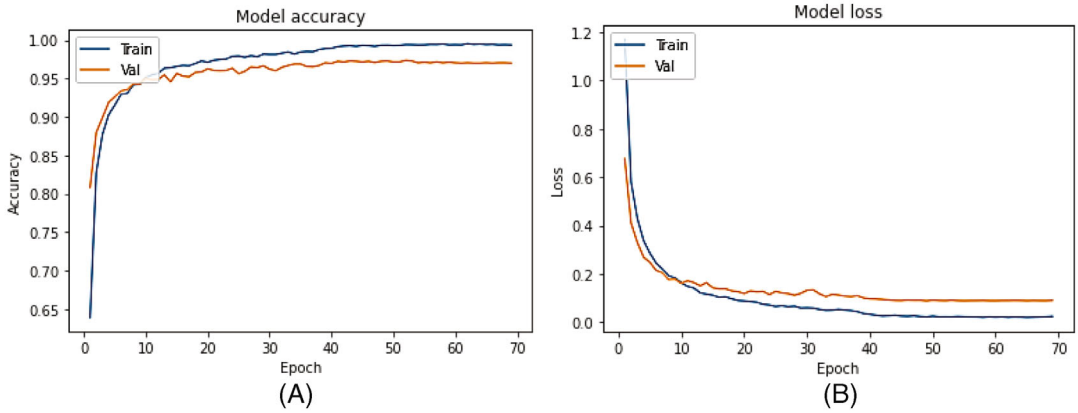

FIGURE 5 Training accuracy and loss of the proposed model on PAMAP2 (A) Accuracy versus epochs of the proposed model (B) Loss versus epochs plot of the proposed model

TABLE 6 Confusion matrix of the proposed model on PAMAP2

Activity	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
A1	316	0	0	0	0	0	0	0	0	4	0	0
A2	6	318	3	1	0	0	0	1	0	1	1	0
A3	0	0	309	0	0	0	0	0	1	0	0	0
A4	0	1	0	385	1	0	2	1	1	0	0	0
A5	0	0	0	0	148	0	1	0	0	0	0	0
A6	0	0	0	0	0	271	0	0	0	3	0	0
A7	0	0	0	0	0	0	303	0	1	0	0	0
A8	0	0	0	3	0	0	1	177	2	2	0	0
A9	0	0	0	1	0	0	0	6	168	6	0	0
A10	1	4	2	0	1	2	0	3	2	272	1	0
A11	0	0	2	0	0	0	0	0	0	2	353	0
A12	0	0	0	0	0	0	0	0	0	1	0	83

Abbreviations: A1, lying; A2, sitting; A3, standing; A4, walking; A5, running; A6, cycling; A7, Nordic walking; A8, ascending stairs; A9, descending stairs; A10, vacuum cleaning; A11, ironing; A12, rope jumping.

As shown in the confusion matrix of the proposed model in Table 6, 316 lying activities were correctly classified with 4 misclassified as vacuum cleaning, 309 standing activities were correctly classified, while 1 was misclassified as descending stairs, 148 running activities were correctly classified out of the total sample of 149, 271 cycling activities were also correctly classified, with 3 misclassified as vacuum cleaning. 177 descending stairs were correctly classified out of the total sample of 205, with 3 misclassified as walking, 1 as Nordic walking, 2 as ascending stairs, and 2 as vacuum cleaning. Likewise, 168 ascending stairs activities were correctly classified out of 181 samples, with 1 misclassified as walking, 6 as descending stairs, and 6 as vacuum cleaning. Also, rope jumping activities had 83 correctly classified out of the total 84 samples, with 1 misclassified

TABLE 7 Evaluation comparison of baseline model versus proposed model on PAMAP2

PAMAP2											
	Baseline model: 97.16%				Proposed model: 97.76%				Relative improvement		
Activity	Precision	Recall	F1-Score	MS	Precision	Recall	F1-Score	MC	Precision	Recall	F1-Score
A1	1.00	0.98	0.99	5	0.98	0.99	0.98	4	−2%	+1%	−1%
A2	0.98	0.98	0.98	6	0.98	0.96	0.97	13	0%	−2%	−1%
A3	0.95	0.99	0.97	2	0.98	1.00	0.99	1	+3%	+1%	+2%
A4	0.99	0.99	0.99	3	0.99	0.98	0.99	6	0%	−1%	0%
A5	0.99	0.99	0.99	1	0.99	0.99	0.99	1	0%	0%	0%
A6	0.98	0.98	0.98	5	0.99	0.99	0.99	3	+1%	+1%	+1%
A7	0.98	0.99	0.99	2	0.99	1.00	0.99	1	+1%	+1%	0%
A8	0.93	0.93	0.94	13	0.94	0.96	0.95	8	+1%	+3%	+1%
A9	0.93	0.93	0.93	11	0.96	0.93	0.94	13	+3%	0%	+1%
A10	0.94	0.93	0.93	21	0.93	0.94	0.94	18	−1%	+1%	+1%
A11	0.98	0.97	0.97	13	0.99	0.99	0.99	4	+1%	+2%	+2%
A12	0.99	0.90	0.94	8	1.00	0.99	0.99	1	+1%	+9%	+5%

Abbreviations: A1, lying; A2, sitting; A3, standing; A4, walking; A5, running; A6, cycling; A7, Nordic walking; A8, ascending stairs; A9, descending stairs; A10, vacuum cleaning; A11, ironing; A12, rope jumping; MS, misclassified samples.

vacuum cleaning. A more detailed comparison based on test accuracy, individual precision, recall and the number of misclassified samples are presented in Table 7.

4.4.3 | PAMAP2 discussion

As shown in the analysis presented in Table 7, the baseline model achieved an overall testing accuracy of 97.16%. The individual precision on the 12 activities classified showed that 8 activities had high precision ranging from 0.98 to 1.00. The lowest precisions were recorded in standing (0.95), ascending stairs (0.93), descending stairs (0.93), and vacuum cleaning (0.94). Likewise, a recall of 0.93, 0.93, 0.93, and 0.90 was recorded on ascending stairs, descending stairs, vacuum cleaning, and rope jumping. The highest misclassifications were also present in these activities, ranging between 11 to 21 misclassified samples. The comparison with the proposed model showed that the overall accuracy of 97.76% achieved on the PAMAP2 dataset outperformed the baseline, improved the precision, recall, and F1 score, and reduced the number of misclassifications in each individual activity. A relative improvement of +3% was achieved in the precision of the standing and descending stairs activities, and +1% in cycling, Nordic walking, ascending stairs, ironing, and rope jumping activities. At the same time, there was a negative precision improvement of −2% and −1% on lying and vacuum cleaning activities. The recall comparison between the baseline and the proposed model showed that relative improvements of +1% was achieved on lying, standing, cycling, Nordic walking and vacuum cleaning. +2% on ironing, +3% on ascending stairs, and +9% on rope jumping. Also, the F1 score improved in the proposed model compared to the baseline, and the number of misclassified activities was

TABLE 8 Performance comparison with existing models on PAMAP2

Model	Year	Accuracy	No. of parameters
Predsim[55]	2020	92.97%	2.60 M
ResNet + HC[56]	2022	92.97%	1.37 M
Conv2D Dual Attention[18]	2021	93.16%	3.51 M
Multibranch CNN-BiLSTM[43]	2021	94.29%	—
CNN-GRU[42]	2021	95.27%	—
Multichannel CNN-GRU[57]	2022	96.25%	—
Residual block with Bi-LSTM[41]	2022	97.15%	0.185 M
RCNN[58]	2021	97.37%	—
Sub-window CNN[59]	2021	97.22%	3.701 M
Ensem-HAR[60]	2022	97.73%	—
Proposed model	—	97.76%	0.549 M

reduced. This result showed that the proposed model could recognize most of the individual activities better than the baseline model. A comparison with the existing models is shown in Table 8.

As shown in Table 8, Predsim, proposed in Teng et al.[55] achieved an overall accuracy of 92.97% on the protocol activities of the PAMAP2 dataset, the multibranch CNN-BiLSTM proposed in Challa et al.[43] achieved an accuracy of 94.29%, and the model on the 12 protocol activities in the dataset. The precision and recall on ascending stairs, descending stairs, vacuum cleaning, ironing, and rope jumping were between 82% and 93%. Our proposed model achieved higher recall and precision. Conv2D dual attention proposed in Gao et al.[18] achieved an accuracy of 93.16% with 3.51 M parameters. Residual Block with Bi-LSTM proposed in Li & Wang,[41] has a recognition accuracy of 97.15%, with a parameter of 0.185 M. Even though the model's parameters are smaller than ours, the model achieved a recognition accuracy of 97.15% and had a low recall of 0.93 on ascending stairs, 0.90 on standing and 0.94 on running. To the best of our knowledge, the best overall accuracy recorded on the PAMAP2 dataset was presented in Bhattacharya et al.[60] where the authors proposed Ensem-HAR and achieved an accuracy of 97.73%. The accuracy of 97.76% recorded by our proposed model outperformed the state-of-the-art by 0.03 with 0.549 M parameters.

4.5 | Experiments on WISDM dataset

4.5.1 | Baseline on WISDM

The baseline model achieved a recognition accuracy of 97.48% on the WISDM dataset, after training for 29 epochs. The model's training accuracy and loss are shown in Figure 6. The confusion matrix is presented in Table 9.

As shown in Table 9, 158 downstairs activities were correctly classified out of the total 173 samples, with 14 misclassified as upstairs, and 1 as standing. 734 jogging activities were correctly classified, while 1 was misclassified as upstairs and 1 as walking, 113 standing activities were

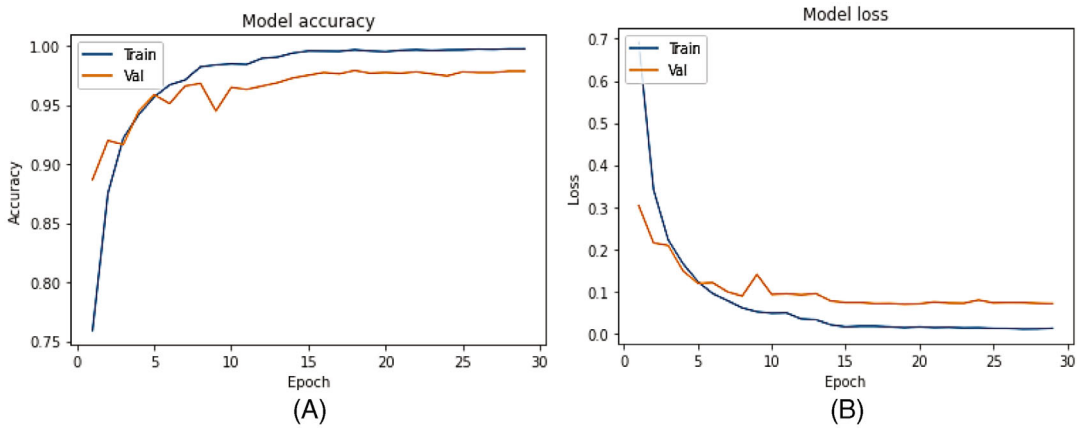

FIGURE 6 Training accuracy and loss of the baseline model on WISDM (A) Accuracy versus epochs of the baseline model (B) Loss versus epochs plot of the baseline model

TABLE 9 Confusion matrix of baseline model on WISDM

Activity	A1	A2	A3	A4	A5	A6
A1	158	0	0	0	14	1
A2	0	734	0	0	1	1
A3	0	0	113	2	0	1
A4	0	0	0	82	1	0
A5	16	2	1	0	221	1
A6	4	1	0	0	0	843

Abbreviations: A1, downstairs; A2, jogging; A3, sitting; A4, standing; A5, upstairs; A6, walking.

correctly classified, but 2 were misclassified as standing and 1 as walking. A total of 20 upstairs activities were misclassified, with 14 misclassified as downstairs, 2 as jogging, 1 as sitting and 1 as walking.

4.5.2 | Proposed model on WISDM

By using the proposed model, an overall testing accuracy of 98.907% was achieved after training for 40 epochs. The model's training and loss vs epoch are shown in Figure 7, and the confusion matrix is shown in Table 10.

As shown in Table 10, the proposed model correctly classified 187 downstairs activities, then misclassified 7 as upstairs, 1 as jogging and 1 as walking, 129 sitting activities were also correctly classified out of the total 130 samples, the whole 103 samples of standing activities were correctly classified, and 255 upstairs activities were correctly classified, while 6 were misclassified as downstairs, 2 as jogging and 1 as sitting. A comparison of the baseline model with the proposed model on the WISDM dataset is shown in Table 11.

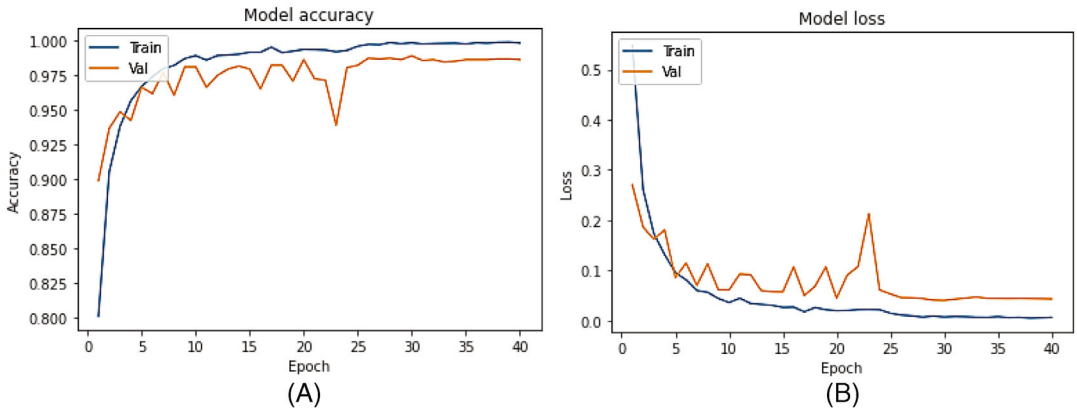

FIGURE 7 Training accuracy and loss of the proposed model on WISDM (A) Accuracy versus epochs of the proposed model (B) Loss versus epochs plot of the proposed model

TABLE 10 Confusion matrix of the proposed model on WISDM

Activity	A1	A2	A3	A4	A5	A6
A1	187	1	0	0	7	1
A2	1	648	0	0	1	0
A3	1	0	129	0	0	0
A4	0	0	0	103	0	0
A5	6	2	1	0	255	0
A6	2	2	0	1	0	851

Abbreviations: A1, downstairs; A2, jogging; A3, sitting; A4, standing; A5, upstairs; A6, walking.

TABLE 11 Evaluation comparison of baseline model versus proposed model on WISDM dataset

WISDM											
	Baseline: 97.48%				Proposed model: 98.907%				Relative improvement		
Activity	Precision	Recall	F1-Score	MS	Precision	Recall	F1-Score	MS	Precision	Recall	F1-Score
A1	0.89	0.91	0.90	15	**0.95**	**0.96**	**0.96**	**8**	+6%	+5%	+6%
A2	**1.00**	1.00	1.00	2	0.99	1.00	1.00	2	−1%	0%	0%
A3	0.99	0.97	0.98	3	0.99	**1.00**	**1.00**	**0**	0%	+3%	+2%
A4	0.98	0.99	0.98	1	**0.99**	**1.00**	**1.00**	**0**	+1%	+1%	+2%
A5	0.93	0.92	0.92	20	**0.97**	**0.97**	**0.97**	**8**	+4%	+5%	+5%
A6	1.00	0.99	0.99	5	1.00	0.99	**1.00**	5	0%	0%	+1%

Abbreviations: A1, downstairs; A2, jogging; A3, sitting; A4, standing; A5, upstairs; A6, walking; MS, misclassified samples.

TABLE 12 Performance comparison with existing and baseline model on WISDM

Model	Year	Accuracy (%)	No. of parameters
PSDRNN[61]	2020	93.06	—
LSTM-CNN[62]	2020	95.75	2.89 M
Multi-head attention[63]	2020	96.40	2.77 M
Multichannel CNN-GRU[57]	2022	96.41	—
CNN-GRU[42]	2021	97.21	—
Attention induced CNN[64]	2021	98.18	1.0415 M
Ensem-HAR[60]	2022	98.70	
PredSim[55]	2020	98.82	2.60 M
Attention ResNet[18]	2021	98.85	2.33 M
ST-Deep HAR[49]	2021	98.90	—
Proposed model		98.90	1.514 M

4.5.3 | WISDM discussion

Experiment on the WISDM dataset showed that the proposed model achieved overall recognition accuracy of 98.907%. As shown in the analysis presented in Table 11, high precision was achieved by the baseline model on jogging, sitting, standing and walking, with precisions of 1.00, 0.99, 0.98, and 1.00, respectively. However, low precision was recorded on downstairs (0.89) and upstairs (0.93). Likewise, the recall of 0.91 and 0.92 recorded on these two activities was the lowest compared to that of the other activities on the baseline model. However, the proposed model increased the precision of the downstairs activity to 0.95 and the recall to 0.96, which shows a relative improvement of +5% and +6% in the precision and recall of the downstairs activities, respectively. Also, the upstairs activity saw an increase in precision of 0.93 on the baseline to 0.97 on the proposed model and a recall increment of 0.97 in the proposed model, compared to the 0.92 achieved on the baseline model. This shows a relative improvement of +4% and +5% in precision and recall respectively on upstairs activities. The other relative improvement achieved include +1% in the precision of standing activities, +3% on the recall of sitting, +1% on the recall of standing, and F1 increment of +6%, +2%. +2%, +5%, and +1% on downstairs, sitting, standing, upstairs and walking, respectively. This result showed that the proposed model outperformed the baseline on the WISDM dataset. A comparison of the proposed model based on accuracy and number of parameters with some existing models is presented in Table 12.

As shown in Table 12, the LSTM-CNN model presented in Li et al.[62] achieved an accuracy of 95.75% with 2.89 M parameters, while the Multi-head attention model proposed in Zhang et al.[63] achieved a recognition accuracy of 96.40% with 2.77 M parameters. The Predsim model proposed in Teng et al.[55] achieved an accuracy of 98.82%, with 2.60 M parameters. To the best of our knowledge, the best recognition accuracy presented on the WISDM dataset is in Abdel-Basset,[49] where the authors proposed ST-Deep HAR, and recorded a recognition accuracy of 98.90%. The proposed model achieved equal accuracy with the state-of-the-art with a recognition accuracy of 98.90%, with increased precision and recall on some individual activities.

4.6 | Experiments on UCI-HAR dataset

4.6.1 | Baseline on UCI-HAR

After 39 epochs, the baseline model without the SEBlock achieved a recognition accuracy of 95.07%. The model's training and model loss are shown in Figure 8, and the confusion matrix is presented in Table 13.

As shown in Table 13, the baseline model correctly classified 536 walking activities, and misclassified 1 as walking downstairs. From the total 491 samples of walking upstairs activities, a total of 6 were misclassified as walking, 41 as walking downstairs, and 1 as laying. Likewise, 35 walking downstairs activities were misclassified as walking upstairs. Fewer misclassifications were recorded in sitting and standing, but a total of 17 laying activities were misclassified as sitting, and 24 as standing.

4.6.2 | Proposed model on UCI-HAR

After 34 epochs, the proposed model achieved an accuracy of 95.89%. The model's training accuracy and loss is presented in Figure 9, and the confusion matrix in Table 14.

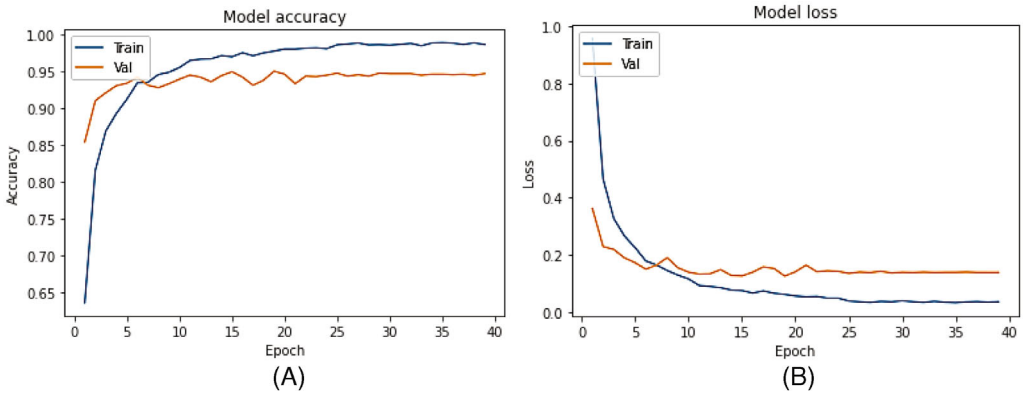

FIGURE 8 Training and model loss of baseline model on UCI-HAR. (A) Accuracy versus epochs of the baseline model (B) Loss versus epochs plot of the baseline model

TABLE 13 Confusion matrix of the baseline model on UCI-HAR

Activity	A1	A2	A3	A4	A5	A6
A1	536	0	1	0	0	0
A2	6	443	41	0	0	1
A3	0	36	496	0	0	0
A4	0	0	0	487	9	0
A5	0	0	0	5	410	5
A6	0	0	0	17	24	430

Abbreviations: A1, walking; A2, walking upstairs; A3, walking downstairs; A4, sitting; A5, standing; A6, laying.

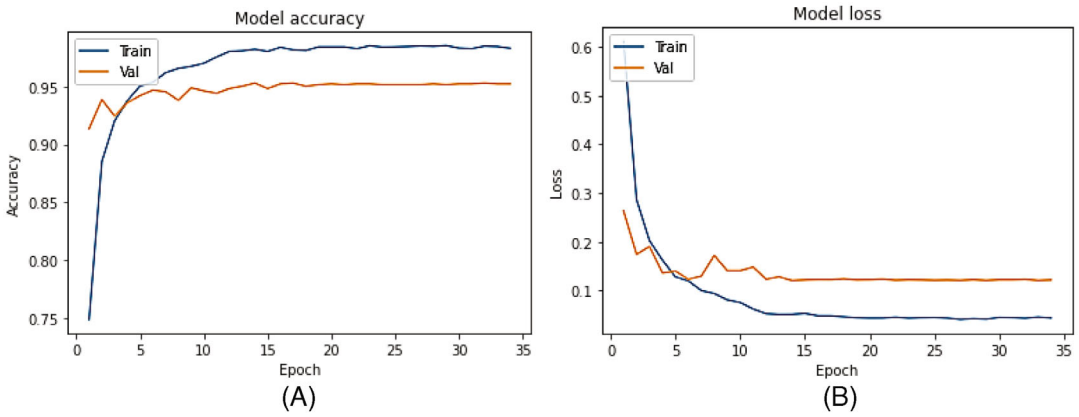

FIGURE 9 Training accuracy and loss of the proposed model on UCI-HAR (A) Accuracy versus epochs of the proposed model (B) Loss versus epochs plot of the proposed model

TABLE 14 Confusion matrix of the proposed model on UCI-HAR

Activity	A1	A2	A3	A4	A5	A6
A1	537	0	0	0	0	0
A2	4	445	40	0	0	2
A3	0	30	502	0	0	0
A4	0	0	0	484	10	2
A5	0	0	0	3	400	17
A6	0	0	0	9	4	458

Abbreviations: A1, walking; A2, walking upstairs; A3, walking downstairs; A4, sitting; A5, standing; A6, laying.

As shown in the confusion matrix of the proposed model on UCI-HAR in Table 14, the total 537 samples of walking activities were correctly classified, while a total of 40 walking upstairs activities were misclassified as walking downstairs, 4 as walking, and 2 as laying. 30 walking downstairs activities were misclassified as walking upstairs, 10 sitting activities were misclassified as standing, and 2 as laying, 17 standing activities were also misclassified as laying, and 3 as sitting, while 9 laying activities were misclassified as sitting, and 4 as standing. A full comparison on the performance of the baseline model with the proposed model on the UCI-HAR dataset is shown in Table 15.

4.6.3 | UCI-HAR discussion

As shown in the result of the analysis presented in Table 15, the baseline model achieved an overall accuracy of 95.07%, while the proposed model had an overall accuracy of 95.89%. The number of misclassified activities reduced in the proposed model compared to the baseline. High precision was recorded on the walking activity by the baseline model, but 0.92 precision was achieved on walking upstairs and walking downstairs, 0.96 on sitting, 0.93 on standing, and 0.95 on laying. The proposed model increased these precisions by +2%, +1%, +2%, +4% and +1% on walking

TABLE 15 Evaluation comparison of baseline model versus proposed model on UCI-HAR dataset

UCI-HAR												
	Baseline Conv1D: 95.07%					Proposed model: 95.89%				Relative improvement to baseline		
Activity	Precision	Recall	F1-Score	Samples	MS	Precision	Recall	F1-Score	MS	Precision	Recall	F1-Score
A1	0.99	1.00	0.99	537	1	0.99	1.00	1.00	0	0%	0%	+1%
A2	0.92	0.90	0.91	491	48	0.94	0.91	0.92	46	+2%	+1%	+1%
A3	0.92	0.93	0.93	532	36	0.93	0.94	0.93	30	+1%	+1%	0%
A4	0.96	0.98	0.97	496	9	0.98	0.98	0.98	12	+2%	0%	+1%
A5	0.93	0.98	0.95	420	10	0.97	0.95	0.96	20	+4%	−3%	+1%
A6	0.95	0.91	0.95	471	41	0.96	0.97	0.96	11	+1%	+5%	+1%

Abbreviations: A1, walking; A2, walking upstairs; A3, walking downstairs; A4, sitting; A5, standing; A6, laying; MS, misclassified samples.

TABLE 16 Performance comparison with existing and baseline model on UCI-HAR

Model	Year	Accuracy	Number of parameters
Bi-LSTM[65]	2020	91.21%	—
InnoHAR[66]	2019	91.70%	—
Net-att3[48]	2019	93.83%	—
Attention induced multi-head[64]	2021	95.38%	1.511 M
LSTM-CNN[62]	2020	95.78%	—
CNN-GRU[42]	2021	96.20%	—
Deep CNN-LSTM + self attention[67]	2022	93.11%	—
Ensem-HAR[60]	2022	95.05%	—
Proposed model		95.89%	1.157 M

upstairs, walking downstairs, sitting, standing and laying, respectively. Likewise, the recall of the proposed model, when compared to the baseline on the UCI-HAR dataset showed an improvement of +1% on walking, 1% on walking upstairs, and +5% on laying activities. However, the recall of the proposed model on the standing activity was better than the proposed model, while no improvements occurred in walking and sitting activities. This result showed that our proposed model outperformed the baseline model. The performance comparison of our proposed model with the existing state-of-the-art on UCI-HAR dataset is shown in Table 16.

As shown in Table 16, the proposed model outperformed the Bi-LSTM model proposed in Li et al.[65] by 4.68%, InnoHAR[66] by 4.19%, Attention induced multi-head proposed in Khan and Ahmad[64] by 0.51%, and LSTM-CNN proposed in Xia et al.[62] by 0.11%. Also, the proposed model outperformed the Deep CNN-LSTM + self-attention model proposed in Khatun[67] and Ensem-HAR proposed in Bhattacharya et al.[60] However, the existing model in Dua et al.[42] where CNN-GRU was proposed achieved a slightly better accuracy than our proposed model, the recall and precision recorded by our proposed model on some of the activities are quite better, and also, the proposed model has smaller number of parameters.

4.7 | Ablation studies

In order to find the best fit SE reduction ratio, ablation studies were done by increasing the reduction ratio starting from ratio 2 to ratio 30, and the performance on the PAMAP2 benchmarking dataset is presented per experiment, based on the accuracy and number of parameters recorded. Figure 10 shows the plot of reduction ratio against accuracy. The highest accuracy is achieved for reduction ratio equal to 6 whereby the recorded accuracy is 97.76%.

5 | CONCLUSION

In this research, we have proposed a lightweight and efficient predictive model for wearable sensor-based activity recognition using a 1D convolutional and max-pooling layers with squeeze and excitation block with best fit varying reduction ratio as a channel-wise attention. The channel-wise attention layer was able to automatically rescale the weights of features according

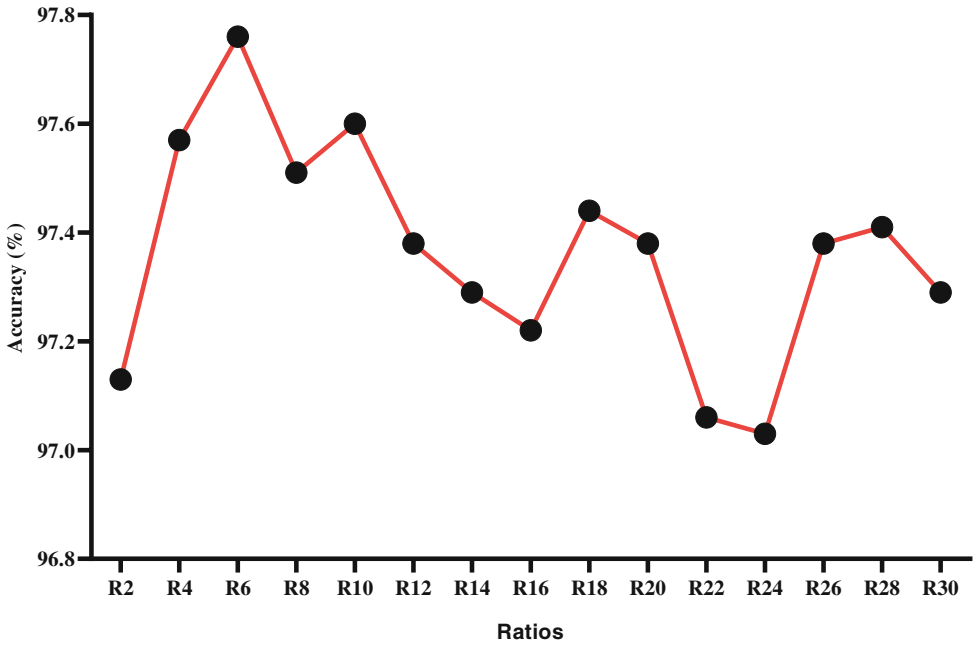

FIGURE 10 Accuracy based on reduction ratio on PAMAP2

to their relevance. Therefore, addressing the challenging task of achieving high recall and precision on activities with complex patterns. Experiments on the benchmarking datasets showed that our model outperformed the baseline and existing models by achieving a recognition accuracy of 97.76%, 98.90%, and 95.89% on PAMAP2, WISDM and UCI-HAR, respectively. This is an improvement of 0.6%, 1.427%, and 0.82% on the performance recorded by the baseline model in this work. However, high improvement of up to +9% over the baseline model was achieved by the proposed model in the recall and precision of some activities. These improvements were also achieved with minimal number of parameters compared to the baseline. By doing this, we have addressed the limitations of discriminative feature learning in activities with complex patterns. Even though some existing models achieved higher overall recognition accuracy than our proposed model on the UCI-HAR dataset, the precision and recall recorded on some activities with inter-class similarities, such as walking, upstairs and downstairs improved compared to the existing models, and this was also achieved with fewer parameters. Therefore, this research has been able to propose a smaller model for activity recognition, which can be easily implemented in wireless devices and capable of achieving state-of-the-art activity recognition. For future work, we plan to validate the model on large-scale sensor-based transient datasets, which will use adaptive window segmentation method. Also, other models that adopt self-attention such as transformers will be explored to investigate how more salient features can be learnt, to improve the performance of activity recognition models.

AUTHOR CONTRIBUTIONS

Study conception, design, and experiments were performed by Ayokunle Olalekan Ige. The first draft of the manuscript was written by Ayokunle Olalekan Ige and Mohd Halim Mohd Noor reviewed the results, revised, and approved the final manuscript.

CONFLICT OF INTEREST

The authors declare no competing financial interests. The authors declare that they have no conflicts of interest to report regarding the present study.

DATA AVAILABILITY STATEMENT

All code for data cleaning and analysis associated with the current submission is available at https://github.com/AOige/SEConv1D. The datasets that support the findings of this study are available in/from UCI repository https://archive.ics.uci.edu/ml/index.php. These data were derived from the following resources available in the public domain: PAMAP2, https://archive.ics.uci.edu/ml/datasets/pamap2+physical+activity+monitoring; UCI-HAR, https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones; WISDM, https://www.cis.fordham.edu/wisdm/dataset.php.

ORCID

Mohd Halim Mohd Noor https://orcid.org/0000-0002-3300-3270

REFERENCES

1. Tan JS, Beheshti BK, Binnie T, et al. Human activity recognition for people with knee osteoarthritis—a proof-of-concept. *Sensors*. 2021;21(10):3381.
2. Kasahara Y, Nishiyama Y, Sezaki K. Detecting childcare activities using an off-the-shelf smartwatch. Paper presented at: 2022 IEEE International Conference on Smart Computing (SMARTCOMP); 2022:159–161. doi:10.1109/SMARTCOMP55677.2022.00038
3. Askari MR, Rashid M, Sun X, et al. Meal and physical activity detection from free-living data for discovering disturbance patterns of glucose levels in people with diabetes. *BioMedInform*. 2022;2(2): 297-317.
4. Ertuğrul ÖF, Kaya Y. Determining the optimal number of body-worn sensors for human activity recognition. *Soft Comput*. 2017;21(17):5053-5060. doi:10.1007/s00500-016-2100-7
5. Casale P, Pujol O, Radeva P. Human activity recognition from accelerometer data using a wearable device. Paper presented at: Lect Notes Comput Sci Subser Lect Notes Artif Intell Lect Notes Bioinforma, 6669 LNCS. 2011:289–296. doi:10.1007/978-3-642-21257-4_36
6. Ige AO, Mohd Noor MH. A survey on unsupervised learning for wearable sensor-based activity recognition. *Appl Soft Comput*. 2022;127:109363. doi:10.1016/j.asoc.2022.109363
7. Mahmud S, Tanjid Hasan Tonmoy M, Kumar Bhaumik K, et al. Human activity recognition fromwearable sensor data using self-attention. *Front Artif Intell Appl*. 2020;325:1332-1339. doi:10.3233/FAIA200236
8. Jimale AO, Mohd Noor MH. Subject variability in sensor-based activity recognition. *J Ambient Intell Humaniz Comput*. 2021. doi:10.1007/s12652-021-03465-6
9. Sani S, Wiratunga N, Massie S, Cooper K. kNN sampling for personalised human activity recognition. Paper presented at: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10339 LNAI; 2017:330–344. doi:10.1007/978-3-319-61030-6_23
10. Uddin MT, Uddiny MA. A guided random forest based feature selection approach for activity recognition. Paper presented at: 2nd International Conference on Electrical Engineering and Information and Communication Technology, ICEEiCT 2015; 2015:1–6. doi:10.1109/ICEEICT.2015.7307376
11. Manosha Chathuramali KG, Rodrigo R. Faster human activity recognition with SVM. Paper presented at: Int Conf Adv ICT Emerg Reg ICTer 2012 - Conf Proc; 2012:197–203. doi:10.1109/ICTer.2012.6421415
12. Fan L, Wang Z, Wang H. Human activity recognition model based on decision tree. Paper presented at: 2013 International Conference on Advanced Cloud and Big Data, CBD 2013; 2013:64–68. doi:10.1109/CBD.2013.19
13. Mohd Noor MH. Feature learning using convolutional denoising autoencoder for activity recognition. *Neural Comput Appl*. 2021;33(17):10909-10922. doi:10.1007/s00521-020-05638-4
14. Wang J, Chen Y, Hao S, Peng X, Hu L. Deep learning for sensor-based activity recognition: a survey. *Pattern Recognit Lett*. 2019;119:3-11. doi:10.1016/j.patrec.2018.02.010

15. Huang J, Lin S, Wang N, Dai G, Xie Y, Zhou J. TSE-CNN: a two-stage end-to-end CNN for human activity recognition. *IEEE J Biomed Health Inform*. 2020;24(1):292-299. doi:10.1109/JBHI.2019.2909688
16. Ma H, Li W, Zhang X, Gao S, Lu S. Attnsense: multi-level attention mechanism for multimodal human activity recognition. *Int Jt Conf Artif Intell*. 2019;3109-3115. doi:10.24963/ijcai.2019/431
17. Zeng M, Gao H, Yu T, et al. Understanding and improving recurrent networks for human activity recognition by continuous attention. Paper presented at: Proc - Int Symp Wearable Comput ISWC; 2018:56–63. doi:10.1145/3267242.3267286
18. Gao W, Zhang L, Teng Q, He J, Wu H. DanHAR: dual attention network for multimodal human activity recognition using wearable sensors. *Appl Soft Comput*. 2021;111:107728. doi:10.1016/j.asoc.2021.107728
19. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2018;7132-7141. doi:10.1109/CVPR.2018.00745
20. Ismail Fawaz H, Forestier G, Weber J, Idoumghar L, Muller PA. Deep learning for time series classification: a review. *Data Min Knowl Discov*. 2019;33(4):917-963. doi:10.1007/s10618-019-00619-1
21. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. doi:10.1038/nature14539
22. Murahari VS, Plotz T. On attention models for human activity recognition. Paper presented at: International Symposium on Wearable Computers, ISWC; 2018:100–103. doi:10.1145/3267242.3267287
23. Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks. Paper presented at: PMLR, ed. 30th International Conference on Machine Learning, ICML 2013, PMLR; 2013:2347–2355.
24. Banos O, Galvez JM, Damas M, Pomares H, Rojas I. Window size impact in human activity recognition. *Sens Switz*. 2014;14(4):6474-6499. doi:10.3390/s140406474
25. Scheurer S, Tedesco S, Brown KN, O'flynn B. Using domain knowledge for interpretable and competitive multi-class human activity recognition. *Sens Switz*. 2020;20(4):1-25. doi:10.3390/s20041208
26. Zhu Q, Chen Z, Soh YC. A novel semisupervised deep learning method for human activity recognition. *IEEE Trans Ind Inform*. 2019;15(7):3821-3830. doi:10.1109/TII.2018.2889315
27. Lara ÓD, Labrador MA. A survey on human activity recognition using wearable sensors. *IEEE Commun Surv Tutor*. 2013;15(3):1192-1209. doi:10.1109/SURV.2012.110112.00192
28. Ronao CA, Cho SB. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst Appl*. 2016;59:235-244. doi:10.1016/j.eswa.2016.04.032
29. Trost SG, Zheng Y, Wong WK. Machine learning for activity recognition: hip versus wrist data. *Physiol Meas*. 2014;35(11):2183-2189. doi:10.1088/0967-3334/35/11/2183
30. Ragb HK, Dover IT, Ali R. Deep convolutional neural network ensemble for improved malaria parasite detection. *Proc - Appl Imag Pattern Recognit Workshop*. 2020:1-10. doi:10.1109/AIPR50011.2020.9425273
31. Prabha P.A., Priya M.D., Jeba Malar A.C., Karthik S., Dakshin GKSD. Improved ResNet-based image classification technique for malaria detection springer, Singapore. In: Mahapatra R.P., Panigrahi B.K., Kaushik B.K., Roy S. (Eds) *Proceedings of 6th International Conference on Recent Trends in Computing. Lecture Notes in Networks and Systems*. Vol 177. Springer; 2021. doi:10.1007/978-981-33-4501-0_73
32. Joy Rakesh Y, Kavitha R, Julian J. Human activity recognition using wearable sensors. *Adv Intell Syst Comput*. 2021;1177(3):527-538. doi:10.1007/978-981-15-5679-1_51
33. Delgado-Ortet M, Molina A, Alférez S, Rodellar J, Merino A. A deep learning approach for segmentation of red blood cell images and malaria detection. *Entropy*. 2020;22(6):1-16. doi:10.3390/e22060657
34. Li W, Shao W, Ji S, Cambria E. BiERU: bidirectional emotional recurrent unit for conversational sentiment analysis. *Neurocomput*. 2022;467:73-82. doi:10.1016/j.neucom.2021.09.057
35. Jantawong P, Jitpattanakul A, Mekruksavanich S. Enhancement of human complex activity recognition using wearable sensors data with inception time network. Paper presented at: 2021 2nd International Conference on Big Data Analytics and Practices, IBDAP 2021; 2021:12–16. doi:10.1109/IBDAP52511.2021.9552133
36. Chernbumroong S, Atkins AS, Yu H. Activity classification using a single wrist-worn accelerometer. Paper presented at: SKIMA 2011 - 5th International Conference on Software, Knowledge Information, Industrial Management and Applications; 2011:21–25. doi:10.1109/SKIMA.2011.6089975
37. Catal C, Tufekci S, Pirmit E, Kocabag G. On the use of ensemble of classifiers for accelerometer-based activity recognition. *Appl Soft Comput J*. 2015;37:1018-1022. doi:10.1016/j.asoc.2015.01.025
38. Gomathi V, Kalaiselvi S, Thamarai SD. Sensor-based human activity recognition using fuzzified deep CNN architecture with λmax method. *Sens Rev*. 2022;42:250-262. doi:10.1108/SR-06-2021-0195

39. Rueda FM, Grzeszick R, Fink GA, Feldhorst S, Ten Hompel M. Convolutional neural networks for human activity recognition using body-worn sensors. *Informatics*. 2018;5(2):1-17. doi:10.3390/informatics5020026
40. Qi W, Su H, Yang C, Ferrigno G, De Momi E, Aliverti A. A fast and robust deep convolutional neural networks for complex human activity recognition using smartphone. *Sens Switz.* 2019;19(17):3731. doi:10.3390/s19173731
41. Li Y, Wang L. Human activity recognition based on residual network and BiLSTM. *Sensors.* 2022;22(2):1-18. doi:10.3390/s22020635
42. Dua N, Singh SN, Semwal VB. Multi-input CNN-GRU based human activity recognition using wearable sensors. *Comput Secur*. 2021;103(7):1461-1478. doi:10.1007/s00607-021-00928-8
43. Challa SK, Kumar A, Semwal VB. A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. *Vis Comput*. 2021. doi:10.1007/s00371-021-02283-3
44. Mohd Noor MH, Tan SY, Ab Wahab MN. Deep temporal conv-LSTM for activity recognition. *Neural Process Lett*. 2022;54:4027-4049. doi:10.1007/s11063-022-10799-5
45. Ullah M, Ullah H, Khan SD, Cheikh FA. Stacked Lstm network for human activity recognition using smartphone data. *Proc - Eur Workshop Vis Inf Process EUVIP*. 2019;175-180. doi:10.1109/EUVIP47703.2019.8946180
46. Nafea O, Abdul W, Muhammad G, Alsulaiman M. Sensor-based human activity recognition with spatio-temporal deep learning. *Sensors.* 2021;21(6):1-20. doi:10.3390/s21062141
47. Ordóñez FJ, Roggen D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sens Switz.* 2016;16(1):115. doi:10.3390/s16010115
48. Wang K, He J, Zhang L. Attention-based convolutional neural network for weakly labeled human activities' recognition with wearable sensors. *IEEE Sens J.* 2019;19(17):7598-7604. doi:10.1109/JSEN.2019.2917225
49. Abdel-Basset M, Hawash H, Chakrabortty RK, Ryan M, Elhoseny M, Song H. ST-DeepHAR: deep learning model for human activity recognition in IoHT applications. *IEEE Internet Things J*. 2021;8(6):4969-4979. doi:10.1109/JIOT.2020.3033430
50. Liu L, He J, Ren K, Lungu J, Hou Y, Dong R. An information gain-based model and an attention-based RNN for wearable human activity recognition. *Entropy.* 2021;23(12):1635. doi:10.3390/e23121635
51. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Paper presented at: 32nd Int Conf Mach Learn ICML 2015; 2015;1:448–456. doi:10.48550/arXiv.1502.03167
52. Reiss A, Stricker D. Introducing a new benchmarked dataset for activity monitoring. Paper presented at: Proc - Int Symp Wearable Comput ISWC 2012; 2012:108–109. doi:10.1109/ISWC.2012.13
53. Kwapisz JR, Weiss GM, Moore SA. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*. 2011;12:74-82. doi:10.1145/1964897.1964918
54. Garcia-Gonzalez D, Rivero D, Fernandez-Blanco E, Luaces MR. A public domain dataset for real-life human activity recognition using smartphone sensors. *Sensors (Switzerland)*. 2020;20:24-26. doi:10.3390/s20082200
55. Teng Q, Wang K, Zhang L, He J. The layer-wise training convolutional neural networks using local loss for sensor-based human activity recognition. *IEEE Sens J.* 2020;20(13):7265-7274. doi:10.1109/JSEN.2020.2978772
56. Han C, Zhang L, Tang Y, Huang W, Min F, He J. Human activity recognition using wearable sensors by heterogeneous convolutional neural networks. *Expert Syst Appl*. 2022;198(2):116764. doi:10.1016/j.eswa.2022.116764
57. Lu L, Zhang C, Cao K, Deng T, Yang Q. A multichannel CNN-GRU model for human activity recognition. *IEEE Access.* 2022;10:66797-66810. doi:10.1109/ACCESS.2022.3185112
58. Li X, Nie L, Si X, Ding R, Zhan D. Enhancing representation of deep features for sensor-based activity recognition. *Mob Netw Appl*. 2021;26(1):130-145. doi:10.1007/s11036-020-01689-y
59. Gil-Martín M, San-Segundo R, Fernández-Martínez F, Ferreiros-López J. Time analysis in human activity recognition. *Neural Process Lett.* 2021;53(6):4507-4525. doi:10.1007/s11063-021-10611-w
60. Bhattacharya D, Sharma D, Kim W, Ijaz MF, Singh PK. Ensem-HAR: an ensemble deep learning model for smartphone sensor-based human activity recognition for measurement of elderly health monitoring. *Biosensors*. 2022;12(6):393. doi:10.3390/bios12060393
61. Li X, Wang Y, Zhang B, Ma J. PSDRNN: an efficient and effective HAR scheme based on feature extraction and deep learning. *IEEE Trans Ind Inform.* 2020;16(10):6703-6713. doi:10.1109/TII.2020.2968920

62. Xia K, Huang J, Wang H. LSTM-CNN architecture for human activity recognition. *IEEE Access*. 2020;8:56855-56866. doi:10.1109/ACCESS.2020.2982225
63. Zhang H, Xiao Z, Wang J, Li F, Szczerbicki E. A novel IoT-perceptive human activity recognition (HAR) approach using multihead convolutional attention. *IEEE Internet Things J*. 2020;7(2):1072-1080. doi:10.1109/JIOT.2019.2949715
64. Khan ZN, Ahmad J. Attention induced multi-head convolutional neural network for human activity recognition. *Appl Soft Comput*. 2021;110:107671. doi:10.1016/j.asoc.2021.107671
65. Li H, Shrestha A, Heidari H, Le Kernec J, Fioranelli F. Bi-LSTM network for multimodal continuous human activity recognition and fall detection. *IEEE Sens J*. 2020;20(3):1191-1201. doi:10.1109/JSEN.2019.2946095
66. Xu C, Chai D, He J, Zhang X, Duan S. InnoHAR: a deep neural network for complex human activity recognition. *IEEE Access*. 2019;7:9893-9902. doi:10.1109/ACCESS.2018.2890675
67. MstA K, Yousuf MA, Ahmed S, et al. Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor. *IEEE J Transl Eng Health Med*. 2022;10:1-16. doi:10.1109/JTEHM.2022.3177710

How to cite this article: Ige AO, Mohd Noor MH. A lightweight deep learning with feature weighting for activity recognition. *Computational Intelligence*. 2022;1-29. doi: 10.1111/coin.12565