*Journal of Computer Science*

Research Article

# Edge Computing Enabled Human Activity Recognition (ECEHAR) Using LSTM and CNN

**Suresh Kumar and M Y. Mohamed Parvees**

*Department of Computer and Information Science, Annamalai University, Chidambaram-608002, India*

*Article history*
Received: 11-06-2025
Revised: 14-10-2025
Accepted: 24-12-2025

**Corresponding Author:**
Suresh Kumar
Department of Computer and Information Science Annamalai University Chidambaram-608002, India
Email: sureshkumarphdau@gmail.com

**Abstract:** Human Activity Recognition (HAR) is an important research area for various application domains such as healthcare, gaming, telemonitoring, and sports. However, executing HAR algorithms on remote servers or in the cloud have challenges in terms of high latency, increased bandwidth demand, and high energy consumption. Moving the computation to edge-assisted HAR is more effective and flexible solution to address the limitations of conventional approaches. In this paper, a set of salient points are identified on the human body and are represented mathematically as triangles. Human activities affect the angles of the triangle, and the resulting deformation is used for classifying the activity. Both Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) are used for human action classification and have good performance with accuracy as 99.8%. The performance of Edge Computing Enabled Human Activity Recognition (ECEHAR) is evaluated on both benchmark and real-time datasets using precision, recall, F1-score, and accuracy. The model has shown promising results compared to contemporary methods.

**Keywords:** Human Activity Recognition, Edge Computing, Deep Learning, Multiclass Classification

## Introduction

Human Activity Recognition (HAR) is an important research domain useful for many real-time applications. Researchers have made significant contributions in HAR in terms of data collection, model development, post-processing techniques, and result interpretation. Most of the traditional approaches (Firouzi et al., 2021; Koch et al., 2022) for HAR uses cloud infrastructure. A classifier is trained on predefined activities of human/objects that are captured from an edge device and stored in cloud for further processing and inference. However, it is observed that cloud-based approach has three important issues such as high latency, delayed user-cloud communication and lack of flexibility, personalization and privacy control.

As a result, researchers have increasingly focused on shifting tasks and computation such as model training, inference, and learning processes from the cloud to edge devices. However, there are challenges due to the inherent limitations of edge devices. These challenges can be addressed by considering factors such as model size, data volume, and energy consumption. It is crucial for the model to be compact enough to fit within the memory constraints of the edge device, taking the storage capacity into account. Similarly, the training process also has to be effective to minimize energy consumption.

In this research work, we present a framework of HAR with Artificial Intelligence based algorithm that is suitable to be deployed on the edge devices (Awotunde et al., 2021) The core of computational processing is moved to the edge devices such that portion of training and inference is handled by the edge devices itself. A deep learning framework is proposed for data collection, model for adaptation, training and retraining, procedure for calibration, inference engine, visual result analysis. All these components are processed and executed on the edge device and thus the data exchange between cloud and edge is avoided. The focus of this work is on human action recognition that is represented in terms of specific points on the human body and is commonly known as salient points. The salient points are identified in the body that corresponds to body joints as well as hands and legs and torso sections. A total of 17 significant points establish the system used to track walking, standing, sitting, and running movements. The neural network employs

convolutional layers paired with Long Short-Term Memory (LSTM) to extract features from activities through an approach that maintains efficient classification while minimizing model parameter number. LSTM is the modified version of Recurrent Neural Network (RNN) which works efficiently with time-sensitive inputs. LSTM and convolutional layers process the input sequence from the edge devices or surveillance cameras. The proposed model distinguishes various actions such as walking, running, standing and sitting effectively on benchmark and real-time datasets.

### *Literature Review*

Human Activity Recognition uses Machine Learning (ML) algorithm for classifying human actions and they are captured from different sensors such as wearable devices, smart phones and IoT devices. It is observed that these algorithms have limitations in terms of latency, privacy, bandwidth and traffic while it is executed on specialized devices (Mukherjee et al., 2018a; Sharma and Wang, 2019). These issues can be handled through edge computing by performing data analysis and processing at the edge devices. As a result, latency and energy consumption are reduced and improves privacy, and optimizes network resource usage. It also improves the real-time analysis and responses for performance enhancement of wearable devices by handling latency concerns (Aran et al., 2016; Yahaya et al., 2019).

Wearable devices are increasingly common in real-time applications particularly in healthcare for monitoring and analyzing vital health parameters of patients to improve healthcare quality and reduce patient risks. Activity of Daily Life (ADL) monitoring systems observe the habits of elderly patients to determine high-risk events of Parkinson's symptoms and Chemotherapy-Induced Peripheral Neuropathy (CIPN) symptoms (Gia et al., 2018; Demrozi et al., 2020a-b; Mantovani et al., 2020). Early detection of chronic diseases allows personalized care and improved safety (Jin et al., 2022). Various studies on effective edge-based HAR demonstrate how different sensors, algorithms, and devices can track the human activities and analyze the same with good accuracy (Goodfellow et al., 2016).

It is observed that the deep learning techniques found to be powerful, face challenges on time awareness and training speed. These methods work in loosely coupled environments and handle issues on communication, and data privacy. However, models like CNN and RNN require significant computational resources and can slow training speeds while processing large-scale data. The Convolutional Recurrent Neural Network (CRNN) structure addresses this issue by combining the strengths of both CNN and RNN for high-accuracy on time-series data. In applications like telerehabilitation, HAR systems are used for closely monitoring the patients that perform motor activities. In fitness, these systems integrate performance metrics with physiological responses (Trotta et al., 2023a; Alam et al., 2023). In addition, the parking recommendation system and object mobility tracking applications in smart city can be complimented by HAR through mobile activities.

Federated Learning (FL) along with semi-supervised learning and clustering methods have been used to solve problems caused by unlabelled data and non-independent, identical data distribution (Presotto et al., 2023; Trotta et al., 2024). Jin et al. (2022) investigated dynamic range quantization as a method to minimize the size of Convolutional Neural Networks (CNNs) for microcontroller units (Luo et al., 2023). However, the data collection from wearable IoT devices for human activity recognition is found to be difficult due to constraint on computational devices and memory (Khatun et al., 2022; Ghibellini et al., 2022). Though the datasets are publicly available (Zhang et al., 2022), this challenge is particularly relevant for supervised Deep Learning (DL) algorithms, since they worked on labelled dataset and it is observed that the performance of DL algorithm has outperformed the ML schemes with encouraging F1 score.

Recent advancements in HAR have increasingly focused on combining deep learning models with edge computing to enable real-time, low-latency applications. Several works have investigated this issue and Trotta et al. (2023b) have highlighted the potential to deploy HAR models on low-power devices with the trade-off between accuracy and computational efficiency. Similarly, Mukherjee et al. (2018b) have discussed latency and privacy challenges in cloud-based HAR which edge devices can handle.

Demrozi et al. (2020c) have demonstrated the effectiveness of combining convolutional layers for feature extraction with recurrent architectures such as LSTM for sequence modelling in wearable sensor–based HAR and this is the reason this paper uses CNN–LSTM for improved sequence recognition.

Sharma and Wang (2017) have explored HAR in IoT environments and highlighted the need for efficient methods adaptable to diverse conditions. Recent studies using YOLO-based detection (Redmon and Farhadi, 2018) also have showed improvements in real-time action localization and however it is not feasible to integrate with sequence-based deep learning.

The recent and contemporary works (Chen et al., 2021; Li et al., 2022) have evaluated RNN, GRU, and LSTM in HAR and have found that the performance of LSTM is encouraging in capturing temporal dependencies (Chen et al., 2021; Li et al., 2022).

Despite these contributions, few studies have integrated pose-based angular representations with CNN–LSTM models on edge devices and found that there are

research gaps. Based on the insights from the literature review, it is observed that HAR algorithms and models has to be executed on edge devices. These devices enable seamless integration of data collection and processing with low latency. The edge is communicated in isolated location and ensures the privacy with personalization. However, due to the limited computational resources of edge devices, careful design of machine learning models is also essential for effectively executing the model.

As a result, we have proposed a new framework, ECEHAR, that employs deep learning techniques on the edge device to recognize human activities. The performance is found to be encouraging compared to the conventional classical ML algorithms in terms of computational complexity and accuracy.

## Methods

### *Data Preprocessing*

The data preprocessing pipeline consists of the following steps:

- Salient Point detection: The body joints are detected in each frame using YOLO v3 in real time
- Angular representation: Triplets of joints form triangles and the corresponding angles are calculated to represent human posture that is invariant to scale and Orientation
- Normalization: Angles are normalized to the range [0, 1] to ensure it is consistent to the input across varying subjects and conditions
- Feature Extraction: Angular sequences are aggregated and passed to CNN and LSTM for feature learning and temporal modeling

### *ECEHAR Salient Points on Human Body*

Human Activity Recognition is an important task that uses specific points on a human (image) and it is referred to as salient points. These points represent various parts of the human object such as joints, hands, legs, body, etc. The salient points are represented by a set of coordinates in 2D ($x$, $y$) or 3D ($x$, $y$, $z$) space.

Sample salient points on the human body are depicted in Fig. 1. As shown in Fig. 1(a), the salient points are connected and represented as lines on human body (image). Similarly, in Fig. 1(b) the points and lines are shown without human body in skeletal structure. The proposed method identifies 17 significant body points encompassing the nose and all four eyes, two ears, eight shoulders, elbows and wrists as well as hips, knees, and ankles. The body map is created using these essential points and is represented as human skeleton. The detailed skeletal structure of the human body with various salient points is shown in Fig. 2. It is observed from Fig. 2 that the human body is decomposed as various salient points and the variation and changes on these salient points can be used for estimating the human activity and action recognition.

The angle between the salient points is used and the way by which these points are connected is depicted in Fig. 3. It demonstrates how changes in angles of body and corresponding triangles allow to monitor specific human body activities. The theoretical concept of the scheme is presented.

### *Angular Representation of Salient Points*

As mentioned earlier, the angles between the salient points provide significant information on human movements. The angles are calculated using the positions of the salient points in consecutive frames of a video sequence. The key angles to be considered for various actions are presented in Table 1.

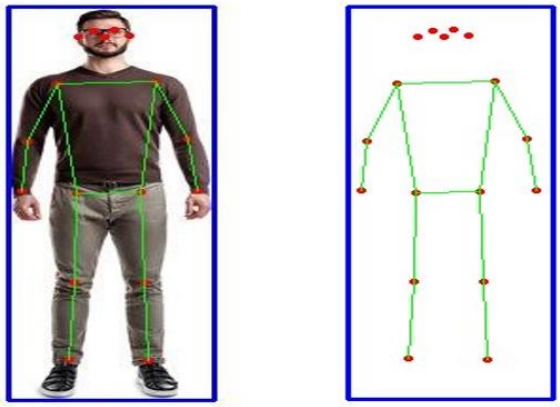


**Fig. 1:** Salient points (a) Salient points superimposed on human body (b) Skeletal structure and salient points

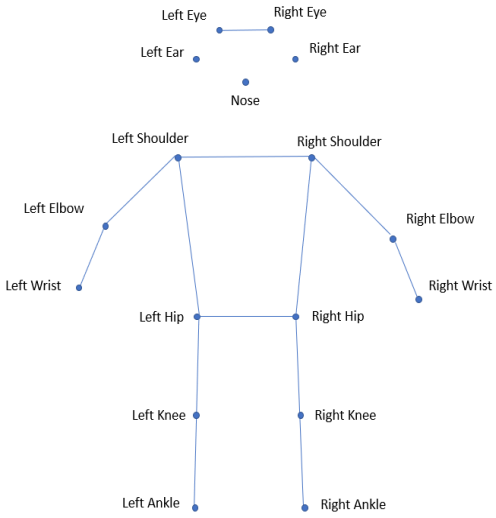



**Fig. 2:** Salient points on Human body

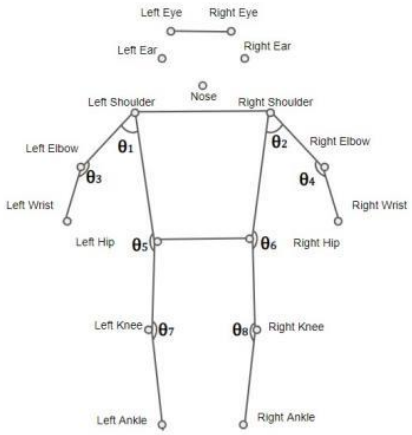



**Fig. 3:** Illustration of the various angle on a human body

In this subsection, the angular representations of salient points are presented. It is noticed in Fig. 3 that various angles are shown over human body and these are derived based on the salient points. The degree of deviation/changes of these angles can be collectively used for human activity recognition. Let's consider three points *A*, *B*, and *C* in a two-dimensional space with coordinates *A* ($x_1$, $y_1$), *B* ($x_2$, $y_2$), and *C* ($x_3$, $y_3$) and is presented in Fig. 4.

The following can be derived as vector from Fig. 4:

$$\overrightarrow{BA}=\vec{A}-\vec{B} \quad (1)$$

$$\overrightarrow{BC}=C-\vec{B} \quad (2)$$

In the above Eq. *A*, *B* and *C* are different points. Given the coordinates of the points, the vectors can be represented as:

$$\overrightarrow{BA}=(x_1-x_2),\ (y_1-y_2) \quad (3)$$

**Table1:** Key angles for HAR

| Area on Human Body | Points to be considered |
|---|---|
| Upper body angles | Shoulder to elbow (both left and right) and elbow to wrist (both left and right) |
| Lower body angles | Hip to knee (both left and right) and knee to ankle (both left and right) |
| Core body angles | Shoulder to hip (both left and right) |

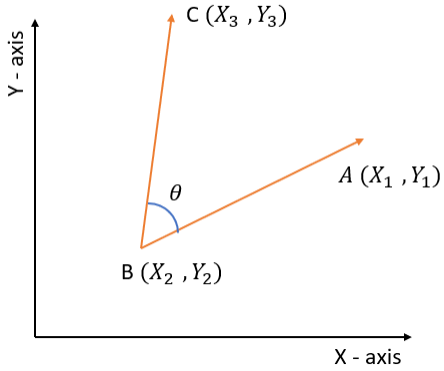



**Fig. 4:** Sample representation of angle in 2D plane

$$\overrightarrow{BC}=(x_3-x_2),\ (y_3-y_2) \quad (4)$$

The angle is derived from the dot product and magnitude of the vectors and is given in Eq. 5 and 6:

$$\overrightarrow{BA}\cdot\overrightarrow{BC}=(x_1-x_2)\,(x_3-x_2)+(y_1-y_2)\,(y_3-y_2) \quad (5)$$

In Eq. 5 $\overrightarrow{BA}\ .\ \overrightarrow{BC}$ is the dot product.

The magnitude (norms) of the vectors $\overrightarrow{BA}$ and $\overrightarrow{BC}$ are given by:

$$\left|\left|\overrightarrow{BA}\right|\right| = \sqrt{(x_1-x_2)^2+(y_1-y_2)^2} \quad (6)$$

$$\left|\left|\overrightarrow{BC}\right|\right| = \sqrt{(x_3-x_2)^2+(y_3-y_2)^2} \quad (7)$$

The Eq. 5 can be rewritten to obtain the cosine of the angle (θ) and is presented in Eq. 8 as:

$$\overrightarrow{BA}\cdot\overrightarrow{BC}=\left|\overrightarrow{BA}\right|\cdot\left|\overrightarrow{BC}\right|\cdot\cos\theta \quad (8)$$

Solving for cosθ, we get:

$$\cos\theta=\frac{\overrightarrow{BA}.\overrightarrow{BC}}{|\overrightarrow{BA}|.|\overrightarrow{BC}|} \quad (9)$$

For finding angle i.e. $\theta$ ,

$$\theta=\cos^{-1}\left(\frac{\overrightarrow{BA}.\overrightarrow{BC}}{|\overrightarrow{BA}|.|\overrightarrow{BC}|}\right) \quad (10)$$

Based on Eq. 10, the angular representation of salient points of a human body is modeled as shown in Fig. 3. The angle variations of different portion of the body are derived and represented as a vector:

$$\theta=\left\{\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6,\theta_7,\theta_8\right\} \quad (11)$$

Each of these $\theta_i$ can be represented as given below in Eq. 12:

$$\theta_1 = \cos^{-1}\left(\frac{\overrightarrow{LE_a\,LE_y}.\overrightarrow{LE_a\,N}}{|\overrightarrow{LE_a\,LE_y}|.|\overrightarrow{LE_a\,N}|}\right) \quad (12)$$

$$\theta_2 = \cos^{-1}\left(\frac{\overrightarrow{RE_a\,RE_y}.\overrightarrow{RE_a\,N}}{|\overrightarrow{RE_a\,RE_y}|.|\overrightarrow{RE_a\,N}|}\right) \quad (13)$$

$$\theta_3 = \cos^{-1}\left(\frac{\overrightarrow{LE\,LS}.\overrightarrow{LE\,LW}}{|\overrightarrow{LE\,LS}.|.|\overrightarrow{LE\,LW}|}\right) \quad (14)$$

$$\theta_4 = \cos^{-1}\left(\frac{\overrightarrow{RE\,RS}.\overrightarrow{RE\,RW}}{|\overrightarrow{RE\,RS}.|.|\overrightarrow{RE\,RW}|}\right) \quad (15)$$

$$\theta_5 = \cos^{-1}\left(\frac{\overrightarrow{LH\,LS}.\overrightarrow{LH\,LK}}{|\overrightarrow{LH\,LS}.|.|\overrightarrow{LH\,LK}|}\right) \quad (16)$$

$$\theta_6 = \cos^{-1}\left(\frac{\overrightarrow{RH\,RS}.\overrightarrow{RH\,RK}}{|\overrightarrow{RH\,RS}.|.|\overrightarrow{RH\,RS}|}\right) \quad (17)$$

$$\theta_7 = \cos^{-1}\left(\frac{\overrightarrow{LK\,LH}.\overrightarrow{LK\,LA}}{|\overrightarrow{LK\,LH}.|.|\overrightarrow{LK\,LA}|}\right) \quad (18)$$

$$\theta_8 = \cos^{-1}\left(\frac{\overrightarrow{RK\,RH}.\overrightarrow{RK\,RA}}{|\overrightarrow{RK\,RH}.|.|\overrightarrow{RK\,RA}|}\right) \quad (19)$$

The variants of the salient points in the skeleton structure of humans can be estimated by measuring the degree of displacement of these points in subsequent video frames. Since the salient points are modeled as triangles and its values are captured to estimate the localized and global displacements. Thus, the displaced angle values can be represented a:

$$\theta = \{\theta_1', \theta_2', \theta_3', \theta_4', \theta_5', \theta_6', \theta_7', \theta_8'\} \quad (20)$$

The difference between the original and displaced angle is measured by a similarity measure and in this paper, we use Vector Cosine Angle Distance (VCAD) and is shown as:

$$VCAD\,(\theta_i - \theta_i') \quad (21)$$

### *Key Angles and Associated Actions*

In this section, we present the key angles that are responsible for various actions. The human action is predicted by analyzing the angles between key points over a sequence of images. Each and every action has different pattern of changes in angles in local and global points. We have carried out extensive experiments and the observation on salient points that are associated with different portion of the body is presented in Tables 2-5.

In Table 2, the details are provided based on the portions and the angles to be considered for walking action. The waist to knee angle and knee to ankle angle are considered from the lower body. The degree of flexion and extension, type of pattern of change, etc are also highlighted. Similarly, shoulder to elbow and elbow to wrist angles are contributing from the upper portion of the body.

A slight oscillation from arms swing is expected. The key angle responsible for running action is presented in Table 3. The waist to knee and knee to ankle angles are considered from the lower body portion. There is a significant flexion and extension is noticed and repetitive synchronized action with waist to knee movement is observed.

**Table 2:** Key angles for walking action

| Body portion | Angles to be considered |
|---|---|
| Lower Body | Waist to Knee Angle:<br>Moderate flexion and extension, alternating between legs<br>Less extreme than running, with a steady and rhythmic pattern<br>Knee to Ankle Angle:<br>Regular flexion, synchronised with the waist-to-knee movement<br>A smooth cyclical pattern |
| Upper Body | Shoulder to Elbow Angle:<br>Minimal movement, slight oscillation as arms swings naturally<br>Elbow to Wrist Angle:<br>Maintained mostly constant, with minor swings corresponding to the natural arm swing |

**Table 3:** Key angles for running action

| Gesture | Angles to be considered |
|---|---|
| Lower Body | Waist to Knee Angle:<br>Repetitive and significant flexion and extension<br>More pronounced and faster than walking<br>Knee to Ankle Angle:<br>Repetitive and synchronized with the waist-to-knee movement<br>Greater extension compared to walking |
| Upper Body | Shoulder to Elbow Angle:<br>Less pronounced movement, slight flexion<br>Elbow to Wrist Angle:<br>Maintained mostly constant, with minor oscillations |

**Table 4:** Key angles for standing action

| Gesture | Angles to be considered |
|---|---|
| Lower Body | Waist to Knee Angle:<br>Minimal flexion, nearly straight<br>Stable, with no significant movement<br>Knee to Ankle Angle:<br>Minimal flexion, nearly straight |
| Upper Body | Shoulder to Elbow Angle:<br>Maintained constant, no significant movement<br>Elbow to Wrist Angle:<br>Maintained constant |

**Table 5:** Key angles for sitting action

| Gesture | Angles to be considered |
|---|---|
| Lower Body | Waist to Knee Angle:<br>Significant flexion, typically around 90 degrees<br>Static, with little to no movement<br>Knee to Ankle Angle:<br>Flexed, typically around 90 degrees, with little to no movement<br>Shoulder to Elbow Angle: |
| Upper Body | May vary depending on posture, often slightly flexed<br>Elbow to Wrist Angle:<br>Maintained constant or slightly flexed if arms are resting |

Clinicians evaluate the shoulder to elbow followed by elbow wrist joint angles for the upper human body. It is observed that there are low movement and slight flexion from shoulder to elbow angle and there is a minor oscillation from elbow to wrist angle.

In Table 4, the angle responsible for standing action is presented and Table 5 presents the angle details of sitting action. As explained above in Table 2 and 3, the information is presented and explained in Tables 4-5 for standing and sitting actions respectively.

## Experimental Results

The video is decomposed as frames and human objects are located and bounded. The reference points of human objects are localized and the angle between the key points are calculated and labeled as action/ activity. The video is segmented into 30 frames / sec and the angle-based feature is extracted and represented as time series data such that it can be trained effectively on LSTM model. The first LSTM layer has 128 units and takes the features as input. The LSTM layer generates the full sequence feature points. The subsequent LSTM layers have 128 and 64 units, respectively, and feature points are passed in a sequence to the next LSTM layer. The time distributed dense layer serves as a fully connected layer for time sequences and produces class probability distributions using the softmax activation function. The Adam is an adaptive learning rate optimizer and is popular due to its efficiency in handling sparse gradients on noisy problems and is used with 0.001 as learning rate. In this work, the categorical cross entropy is used, since we resolve a classification with multiple classes. The model uses accuracy as the performance metric during both training and evaluation. Summary () function gives the summary of the model architecture, details of layers, output shapes, and the number of parameters. Fig. 5 depicts the snippet of code of the model.

```
model = Sequential()
model.add(LSTM(128, return_sequences=True, input_shape=(time_steps, num_features)))
model.add(LSTM(128, return_sequences=True))
model.add(LSTM(64, return_sequences=True))
model.add(LSTM(64, return_sequences=True))
model.add(TimeDistributed(Dense(num_classes, activation='softmax')))
# Compile the model
optimizer = Adam(learning_rate=0.001)
model.compile(optimizer=optimizer, loss='categorical_crossentropy', metrics=['accuracy'])
# Display model summary
model.summary()
# Train the model
history = model.fit(X_train, y_train, epochs=100, batch_size=32, validation_split=0.2)
# Evaluate the model
loss, accuracy = model.evaluate(X_test, y_test)
print(f'Test Loss: {loss}')
print(f'Test Accuracy: {accuracy}')
```

**Fig. 5:** Snippet of the code of the model

### *Hardware Details of the Edge Device*

The Raspberry Pi reaches customers through its affordable design as the credit-card-sized computer operates with System-on-a-Chip (SoC) technology. This device integrates a 32-bit ARM processor with its processing capabilities spread between 700 MHz to 1000 MHz and the GPU operates as a Video Core 4. The Raspberry Pi 4 presents itself with a 64-bit quad-core processor which delivers high-end performance while providing dual Micro HDMI ports that support 4K resolution along with hardware 4Kp60 video decoding and 4GB of RAM and dual-band wireless LAN and Bluetooth 5.0 and Gigabit Ethernet and USB 3.0 and Power over Ethernet (PoE) functionality. A Raspberry Pi camera incorporates a 5-megapixel image sensor alongside a designed add-on board which works with 720p and 1080p video. Fig. 6 shows the internal configuration of the Raspberry Pi board, and its specifications are detailed in Table 6.

### *Experimental Configuration and Outcomes*

### *Calculating Appropriate Training Epoch and Hyper Parameters*

The subsequent part contains a thorough examination of experimental test outcomes. We have adjusted the number of epochs to find the most suitable value to have encouraging accuracy for training and validation datasets. The optimization process has demonstrated essential value in boosting the performance of the model since it has provided effective data generalization abilities for new information. The experiments employed the UCI-HAR standard dataset that is generated with thirty adults having their age between 19 to 48 years. All these subjects wore smart watches on their waist during six different activities that included walking, sitting, standing, running and walking upstairs. Generators handled 70% of the split dataset through training while testing is performed using the remaining 30% by testers because of random participant division. Two distinct segments of data set are allocated for training activities alongside testing activities respectively.

**DOI**: 10.3844/jcssp.2026.787.799

Various testing epochs from 10 to 300 with increments of 10 are conducted to examine accuracy and loss in both validation and testing datasets. The best training epochs are determined through selecting periods where validation loss reached its lowest point while accuracy is at its highest. Figs. 6-8 show performance results of the models.

In Fig. 7 (a), the training and validation accuracy using RNN with CNN is presented and it is observed that the accuracy is high after 100 epochs and reached its saturation. Similar observations in Fig. 7(b) shows that the training and validation loss is low after 100 epochs of training and getting sustained.


**Fig. 6:** Illustration of Raspberry Pi board architecture

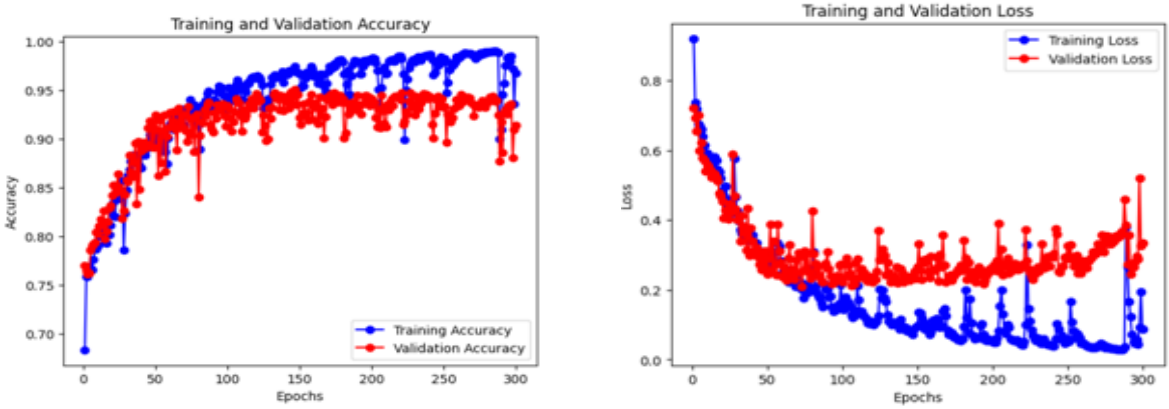
**Table 6:** Specification of Raspberry pi 4

| Sl. No | Specification | Description |
|---|---|---|
| 1 | Size | Around 25x25x9 mm |
| 2 | Weight | 3 g |
| 3 | Still resolution | 5 Megapixels |
| 4 | Video modes | 1080 p30, 720 p60 and 640x480 p60/90 |
| 5 | Linux Integration | V4L2 driver available |
| 6 | Sensor | Omni Vision OV5647 |
| 7 | Sensor Resolution | 2592x1944 pixels |
| 8 | Sensor Image Area | 3.76x2.74 mm |
| 9 | Pixel Size | 1.4x1.4 µm |
| 10 | Optical Size | ¼ |
| 11 | Full-frame SLR lens equivalent | 35 mm |
| 12 | S/N Ratio | 36 B |


a) Accuracy


b) Loss

**Fig. 7:** Training and validation using CNN with RNN




a) Accuracy b) Loss

**Fig. 8:** Training and validation using CNN with LSTM (a and b)

In Fig. 8 (a), the training and validation accuracy using LSTM with CNN is presented and it is observed that the accuracy is high after 100 epochs and reached its saturation. Similar observation in Fig. 8(b) that the training and validation loss low after 100 epochs of training and getting sustained.

In Fig. 9 (a), the training and validation accuracy using GRU with CNN is presented and it is observed that the accuracy is high after 100 epochs and reached its saturation. Similar observation in Fig. 9 (b) that the training and validation loss low after 100 epochs of training and getting sustained.

From Figs. 7-8 and 9 it is concluded that the accuracy reaches the maximum value after 100 epochs and reached its saturation for both training and testing. Hence, the remaining experiment in this research work is carried out by fixing the epoch as 100 and the extensive experimental results are presented in the subsequent sections.

## *Performance Analysis With RNN, LSTM and GRU*

Evaluation of the proposed model utilizes multiple performance metrics which include confusion matrix alongside support and precision and recall and the F1 score and accuracy. Different evaluation metrics combined together provide an extensive evaluation of model effectiveness ensuring reliable generalizable results. The confusion matrix presents specific data about predicted/actual outcome matches to show strengths and weaknesses in model performance. The F1 score acts as a valuable assessment method for multiple-classification models because it creates an equilibrium between precision and recall metrics. The model's positive prediction accuracy is measured by precision and overall accuracy comprises the total proportion of accurate predictions throughout all categories. Support provides meaningful information about the true instances per class which helps analyze other metrics. Based on this overall analysis 100 epochs turned out to give the best results for the simple RNN model. The formulas to determine accuracy and F1 score and precision together with recall appear in Equations 22-25 respectively:

$$Accuracy = \frac{(True\ positive + True\ Negative)}{(True\ Positive + True\ Negative + False\ Positive + \ False\ Negative)} \quad (22)$$

$$\text{F-Measure} = \frac{(2 * \text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (23)$$

$$\text{Precision} = \frac{True\ Positive}{(True\ Positve + False\ Positive)} \quad (24)$$

$$Recall = \frac{True\ Positive}{(True\ Positive + False\ Negative)} \quad (25)$$

From Table 7, it is observed that RNN with CNN model achieves better performance in prediction of all activities such as walking, sitting, standing and running. The increase in validation loss indicates specific opportunities for model improvement that will result in stronger predictions. The training loss computation uses categorical cross-entropy to determine the accuracy between predicted and actual activity labels. The proposed approach gives the average accuracy of 91.5% for the various activities considered for experimental analysis.

The results of Table 8 illustrates that the LSTM model is trained for 100 epochs and the training loss is logged at each epoch. The test loss high lights how the model is responding to unseen data. The final test accuracy is 99.8aftertraining the LSTM model for 100 epochs.

From Table 9, it is observed that GRU with CNN model yields better performance in the prediction of sitting compared to other activates considered for investigation. The higher validation loss suggests areas for potential improvement, which could enhance the model's predictive capabilities. The training loss is computed using categorical cross-entropy, which measures how well the predicted activity label matches the ground-truth labels. The proposed approach gives the average accuracy of 91.8% for the various activities considered for experimental analysis.

**Table 7:** Performance of RNN with CNN

| | Predicted standing | Predicted Walking | Predicted Sitting | Predicted Running | Support | Precision | Recall | F1-Score | Accuracy (%) |
|---|---|---|---|---|---|---|---|---|---|
| Actual standing | 1082 | 94 | 24 | 0 | 1200 | 0.81 | 0.90 | 0.85 | 90.16 |
| Actual Walking | 103 | 8877 | 375 | 125 | 9480 | 0.92 | 0.93 | 0.92 | 93.63 |
| Actual Sitting | 143 | 381 | 6649 | 27 | 7200 | 0.93 | 0.92 | 0.92 | 92.34 |
| Actual Running | 0 | 237 | 56 | 2677 | 2970 | 0.94 | 0.90 | 0.91 | 90.13 |
| Average accuracy | | | | | | | | | 91.5% |

**Table 8:** Performance of LSTM with CNN

| | Predicted standing | Predicted Walking | Predicted Sitting | Predicted Running | Support | Precision | Recall | F1-Score | Accuracy (%) |
|---|---|---|---|---|---|---|---|---|---|
| Actual standing | 1180 | 15 | 5 | 0 | 1200 | 0.98 | 0.98 | 0.98 | 99.8 |
| Actual Walking | 20 | 9450 | 5 | 5 | 9480 | 0.99 | 0.99 | 0.99 | 99.83 |
| Actual Sitting | 10 | 5 | 7185 | 0 | 7200 | 0.99 | 0.99 | 0.99 | 99.83 |
| Actual Running | 0 | 6 | 4 | 2960 | 2970 | 0.99 | 0.98 | 0.98 | 99.8 |
| Average accuracy | | | | | | | | | 99.8% |

**Table 9:** Performance of GRU with CNN

| | Predicted standing | Predicted Walking | Predicted Sitting | Predicted Running | Support | Precision | Recall | F1-Score | Accuracy (%) |
|---|---|---|---|---|---|---|---|---|---|
| Actual standing | 1096 | 39 | 65 | 0 | 1200 | 0.95 | 0.91 | 0.92 | 91.33 |
| Actual Walking | 27 | 8556 | 258 | 639 | 9480 | 0.93 | 0.90 | 0.91 | 90.25 |
| Actual Sitting | 26 | 331 | 6767 | 76 | 7200 | 0.95 | 0.93 | 0.94 | 93.98 |
| Actual Running | 0 | 209 | 31 | 2730 | 2970 | 0.79 | 91 | 0.85 | 91.91 |
| Average accuracy | | | | | | | | | 91.8% |

It is observed from the results of Tables 7-9, the proposed frame work of LSTM with CNN yields better classification accuracy for all the activities.

### *HAR Using LSTM With CNN on Real-Time Dataset*

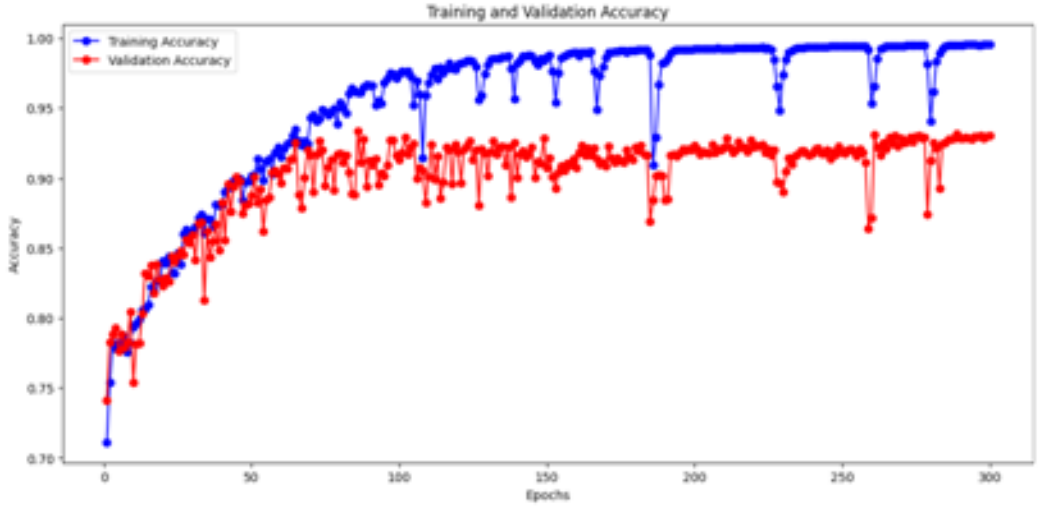
The performance of the proposed method on the real-time dataset is presented in this subsection in addition to the evaluation on the standard dataset. The real-time dataset is created from ten different objects for walking, standing, sitting and running actions. The input sequence is collected from the surveillance/ edge device and fed into YOLO model to identify the salient points in the human body. The various angles are extracted from the salient points and are given as input to LSTM and CNN model for the human activity recognition. Fig. 10 illustrates the classification of various activities using LSTM with CNN.

The proposed approach has achieved the maximum accuracy of 99.1% for walking, 99.13% for sitting and running and 99.3% for standing respectively.

### *Performance Comparison of ECEHAR*

In order to show the efficiency of ECEHAR, the performance is compared with the PRISM-HAR approach and the results are presented in Table 10.

The results in Table 10 indicate the ECEHAR method surpasses the performance of PRISM HAR for every HAR class. The proposed method achieves good results on both standard UCI HAR and real-time dataset.




a) Accuracy


b) loss

**Fig. 9:** Training and validation using CNN with GRU (a and b)

**Table 10:** Performance evaluation of ECEHAR with existing approach

| Dataset | Accuracy (%) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | ECEHAR | | | | PRISM HAR | | | |
| | Walking | Standing | Sitting | Running | Walking | Standing | Sitting | Running |
| UCI-HAR | 99.3 | 99.8 | 99.8 | 99.8 | 97.8 | 98 | 97.8 | 98.3 |
| Real- time dataset | 99.1 | 99.3 | 99.13 | 99.13 | 96.8 | 97 | 98.5 | 97.3 |

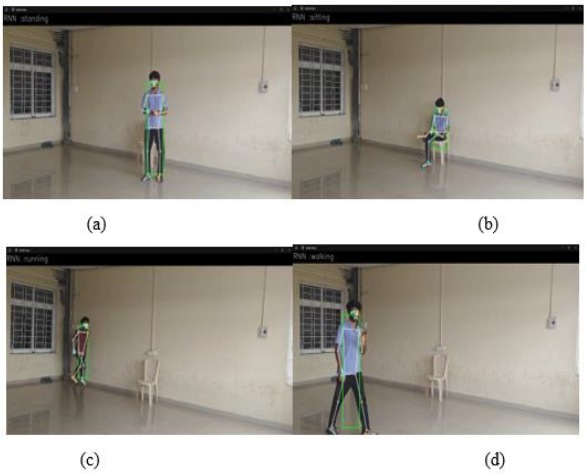



**Fig. 10:** Prediction of human action (a) standing (b) sitting (c) walking and (d) running

## Conclusion

### *Technical Development*

The proposed hybrid approach integrates the YOLO v3 model with deep learning architectures such as CNN and LSTM to deliver a robust solution for human action recognition. YOLO v3 ensures efficient real-time pose estimation and hand tracking, enabling immediate and accurate analysis of human actions. When combined with CNN and LSTM, the framework has provided deeper insights compared to traditional machine learning models. Extensive experiments are conducted with RNN, LSTM, and GRU by varying the number of epochs from 10 to 300 with optimal training configuration. The proposed approach has achieved high performance on both benchmark and real-time datasets, with maximum accuracy of 99.83% for walking, sitting, and running, and 99.8% for standing.

### *Applied Deployment*

The proposed approach has demonstrated a strong potential for various real-time applications across domains, including healthcare, sports, etc. The future work will evaluate the impact of activity speed (slow, normal, fast) for walking, running, and sitting. Broader deployment scenarios will also be considered with larger and more diverse populations, expanded activity sets, and varied environments to improve generalizability and scalability.In addition, limitations on the size of the dataset and real-world applicability will be addressed. Future research directions include exploring multimodal HAR, implementing federated learning, and deploying the system over 5G/6G networks for enhanced edge intelligence.

## Acknowledgment

Thank you to the publisher for their support in the publication of this research article. We are grateful for the resources and platform provided by the publisher, which have enabled us to share our findings with a wider audience. We appreciate the efforts of the editorial team in reviewing and editing our work, and we are thankful for the opportunity to contribute to the field of research through this publication.

## Funding Information

On Behalf of all authors the corresponding author states that they did not receive any funds for this project.

## Author's Contributions

**Suresh Kumar:** Performed the Analysis the overall concept, writing and edited.

**M Y. Mohamed Parvees:** Participated in the methodology, Conceptualization, Data collection and writing the study.

## Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

### *Data Availability Statement*

All the data is collected from the simulation reports of the software and tools used by the authors. Authors are working on implementing the same using real world data with appropriate permissions.

## References

Alam, G., McChesney, I., Nicholl, P., & Rafferty, J. (2023). Open Datasets in Human Activity Recognition Research Issues and Challenges: A Review. *IEEE Sensors Journal*, *23*(22), 26952–26980. https://doi.org/10.1109/jsen.2023.3317645

Aran, O., Sanchez-Cortes, D., Do, M.-T., & Gatica-Perez, D. (2016). Anomaly Detection in Elderly Daily Behavior in Ambient Sensing Environments. *Human Behavior Understanding*, *9997*, 51–67. https://doi.org/10.1007/978-3-319-46843-3_4

Awotunde, J. B., Jimoh, R. G., Folorunso, S. O., Adeniyi, E. A., Abiodun, K. M., & Banjo, O. O. (2021). Privacy and Security Concerns in IoT-Based Healthcare Systems. *The Fusion of Internet of Things, Artificial Intelligence, and Cloud Computing in Health Care*, 105–134. https://doi.org/10.1007/978-3-030-75220-0_6

Chen, C., Li, X., Chen, W., Zhang, Y., & Wang, Z. (2021). A comparative study of deep learning methods for human activity recognition using wearable sensors. *Sensors*, *21*(16), 1–22.

Demrozi, F., Bacchin, R., Tamburin, S., Cristani, M., & Pravadelli, G. (2020a). Toward a Wearable System for Predicting Freezing of Gait in People Affected by Parkinson's Disease. *IEEE Journal of Biomedical and Health Informatics*, *24*(9), 2444–2451. https://doi.org/10.1109/jbhi.2019.2952618

Demrozi, F., Pravadelli, G., Bihorac, A., & Rashidi, P. (2020b). Human Activity Recognition Using Inertial, Physiological and Environmental Sensors: A Comprehensive Survey. *IEEE Access*, *8*, 210816–210836. https://doi.org/10.1109/access.2020.3037715

Demrozi, F., Pravadelli, G., Tighe, P. J., Bihorac, A., & Rashidi, P. (2020c). Joint Distribution and Transitions of Pain and Activity in Critically Ill Patients. *Proceeding of the Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 4534–4538. https://doi.org/10.1109/embc44109.2020.9176453

Firouzi, F., Farahani, B., Panahi, E., & Barzegari, M. (2021). Task Offloading for Edge-Fog-Cloud Interplay in the Healthcare Internet of Things (IoT). *Poceeding of the IEEE International Conference on Omni-Layer Intelligent Systems (COINS)*, 1–8. https://doi.org/10.1109/coins51742.2021.9524098

Ghibellini, A., Bononi, L., & Di Felice, M. (2022). Intelligence at the IoT Edge: Activity Recognition with Low-Power Microcontrollers and Convolutional Neural Networks. *Proceeding of the IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*, 707–710. https://doi.org/10.1109/ccnc49033.2022.9700665

Gia, T., Sarker, V. K., Tcarenko, I., Rahmani, A. M., Westerlund, T., Liljeberg, P., & Tenhunen, H. (2018). Energy efficient wearable sensor node for IoT-based fall detection systems. *Microprocessors and Microsystems*, *56*, 34–46. https://doi.org/10.1016/j.micpro.2017.10.014

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning the MIT Press*.

Jin, X., Li, L., Dang, F., Chen, X., & Liu, Y. (2022). A survey on edge computing for wearable technology. *Digital Signal Processing*, *125*, 103146. https://doi.org/10.1016/j.dsp.2021.103146

Khatun, Mst. A., Yousuf, M. A., Ahmed, S., Uddin, Md. Z., Alyami, S. A., Al-Ashhab, S., Akhdar, H. F., Khan, A., Azad, A., & Moni, M. A. (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. https://doi.org/10.1109/jtehm.2022.3177710

Koch, M., Schlenke, F., Kohlmorgen, F., Kuller, M., Bauer, J., & Woehrle, H. (2022). Detection and Classification of Human Activities using Distributed Sensing of Environmental Vibrations. *Proceeding of the IEEE International Conference on Omni-Layer Intelligent Systems (COINS)*, 1–6. https://doi.org/10.1109/coins54846.2022.9854970

Li, X., Wang, J., & Gao, Y. (2022). Exploring temporal modeling with recurrent neural networks for human activity recognition. *Neurocomputing*, *469*, 479–489.

Luo, F., Khan, S., Huang, Y., & Wu, K. (2021). Binarized neural network for edge intelligence of sensor-based human activity recognition. *IEEE Transactions on Mobile Computing*, *22*(3), 1356–1368. https://doi.org/10.1109/tmc.2021.3109940

Mantovani, E., Demrozi, F., Hertz, D. L., Turetta, C., Ferro, O., Argyriou, A. A., Pravadelli, G., & Tamburin, S. (2022). Wearables, sensors, and smart devices for the detection and monitoring of chemotherapy-induced peripheral neurotoxicity: Systematic review and directions for future research. *Journal of the Peripheral Nervous System*, *27*(4), 238–258. https://doi.org/10.1111/jns.12518

Mukherjee, D., Misra, S., Saha, N., & Obaidat, M. S. (2018a). Energy-efficient decision fusion for human activity recognition in fog computing. *IEEE Transactions on Cloud*, *6*(2), 468–478.

Mukherjee, M., Shu, L., & Wang, D. (2018b). Survey of Fog Computing: Fundamental, Network Applications, and Research Challenges. *IEEE Communications Surveys & Tutorials*, *20*(3), 1826–1857. https://doi.org/10.1109/comst.2018.2814571

Presotto, R., Civitarese, G., & Bettini, C. (2023). Federated Clustering and Semi-Supervised learning: A new partnership for personalized Human Activity Recognition. *Pervasive and Mobile Computing*, *88*, 101726. https://doi.org/10.1016/j.pmcj.2022.101726

Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *ArXiv Preprint*, 1–6.

Sharma, A., & Wang, J. (2019). Towards Real-Time Human Activity Recognition in IoT-Enabled Smart Home Environments. *Proceeding of the International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 1–1. https://doi.org/10.1109/percomw.2019.8730719

Sharma, S. K., & Wang, X. (2017). Live Data Analytics With Collaborative Edge and Cloud Processing in Wireless IoT Networks. *IEEE Access*, *5*, 4621–4635. https://doi.org/10.1109/access.2017.2682640

Trotta, A., Montori, F., Ciabattini, L., Billi, G., Bononi, L., & Di Felice, M. (2024). Edge human activity recognition using federated learning on constrained devices. *Pervasive and Mobile Computing*, *104*, 101972. https://doi.org/10.1016/j.pmcj.2024.101972

Trotta, A., Montori, F., Vallasciani, G., Bononi, L., & Di Felice, M. (2023a). Optimizing IoT-based Human Activity Recognition on Extreme Edge Devices. *Procedding of the IEEE International Conference on Smart Computing (SMARTCOMP)*, 41–48. https://doi.org/10.1109/smartcomp58114.2023.00023

Trotta, R., Aceto, G., M. Rodrigues, A. C., Persic, V., & Pescapé, A. M. (2023b). Edge computing for human activity recognition: Survey and research challenges. *IEEE Internet of Things*, *10*(5), 4119–4142.

Yahaya, S. W., Lotfi, A., & Mahmud, M. (2019). A Consensus Novelty Detection Ensemble Approach for Anomaly Detection in Activities of Daily Living. *Applied Soft Computing*, *83*, 105613. https://doi.org/10.1016/j.asoc.2019.105613

Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., & Alshurafa, N. (2022). Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances. *Sensors*, *22*(4), 1476. https://doi.org/10.3390/s22041476