Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.052771

ARTICLE

# Enhancing Human Action Recognition with Adaptive Hybrid Deep Attentive Networks and Archerfish Optimization

**Ahmad Yahiya Ahmad Bani Ahmad[1], Jafar Alzubi[2], Sophers James[3], Vincent Omollo Nyangaresi[4,5,*], Chanthirasekaran Kutralakani[6] and Anguraju Krishnan[7]**

[1]Department of Accounting and Finance, Faculty of Business, Middle East University, Amman, 11831, Jordan

[2]Faculty of Engineering, Al-Balqa Applied University, Salt, 19117, Jordan

[3]Department of Mathematics, Kongunadu College of Engineering and Technology (Autonomous), Tholurpatti, Trichy, 621215, India

[4]Department of Computer Science and Software Engineering, Jaramogi Oginga Odinga University of Science and Technology, Bondo, 210-40601, Kenya

[5]Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, 602105, India

[6]Department of Electronics and Communication Engineering, Saveetha Engineering College (Autonomous), Chennai, 602105, India

[7]Department of Computer Science and Engineering, Kongunadu College of Engineering and Technology (Autonomous), Tholurpatti, Trichy, 621215, India

*Corresponding Author: Vincent Omollo Nyangaresi. Email: vnyangaresi@jooust.ac.ke

Received: 15 April 2024 Accepted: 14 August 2024 Published: 12 September 2024

## ABSTRACT

In recent years, wearable devices-based Human Activity Recognition (HAR) models have received significant attention. Previously developed HAR models use hand-crafted features to recognize human activities, leading to the extraction of basic features. The images captured by wearable sensors contain advanced features, allowing them to be analyzed by deep learning algorithms to enhance the detection and recognition of human actions. Poor lighting and limited sensor capabilities can impact data quality, making the recognition of human actions a challenging task. The unimodal-based HAR approaches are not suitable in a real-time environment. Therefore, an updated HAR model is developed using multiple types of data and an advanced deep-learning approach. Firstly, the required signals and sensor data are accumulated from the standard databases. From these signals, the wave features are retrieved. Then the extracted wave features and sensor data are given as the input to recognize the human activity. An Adaptive Hybrid Deep Attentive Network (AHDAN) is developed by incorporating a "1D Convolutional Neural Network (1DCNN)" with a "Gated Recurrent Unit (GRU)" for the human activity recognition process. Additionally, the Enhanced Archerfish Hunting Optimizer (EAHO) is suggested to fine-tune the network parameters for enhancing the recognition process. An experimental evaluation is performed on various deep learning networks and heuristic algorithms to confirm the effectiveness of the proposed HAR model. The EAHO-based HAR model outperforms traditional deep learning networks with an accuracy of 95.36, 95.25 for recall, 95.48 for specificity, and 95.47 for precision, respectively. The result proved that the developed model is effective in recognizing human action by taking less time. Additionally, it reduces the computation complexity and overfitting issue through using an optimization approach.

**KEYWORDS**

Human action recognition; multi-modal sensor data and signals; adaptive hybrid deep attentive network; enhanced archerfish hunting optimizer; 1D convolutional neural network; gated recurrent units

## 1 Introduction

Over the last two decades, the HAR has progressed with bounds and leaps with various indirect and direct practical implications such as Artificial Intelligence (AI) mount cameras in robotics, wearable sensor devices, health and fitness apps in the healthcare industry, Automated surveillance systems in surveillance of traffic, and Interaction environment between the computers and computers are majorly influencing our daily lives [1]. Extensive research is required for the recognition of human activities and behaviors since these are the true motives behind these HAR over the recent years [2]. In wearable and mobile computing, the HAR becomes a wide area of research in which wearable devices are mostly used for the collection of data for recognition purposes. A deep understanding of the activity patterns of individuals is needed for the recognition of human actions. Furthermore, the long-term abilities and habits of individuals contribute to a wide range of user-centric applications. Video streaming, human-computer interaction, video surveillance, and healthcare systems are the wide of applications supported by the HAR [3]. In the field of computer vision, the recognition of human activities is extensively studied and limited to specific scenario-based applications. Hence, pre-installed cameras are equipped in the HAR models with guaranteed angle of view and sufficient resolution [4]. Yet, the use of wearable sensors in HAR approaches allows for continuous sensing without being constrained by spatio-temporal characteristics during daily activities [5]. Since the wearable does not need infrastructural support and is ubiquitous and hence special attention is needed for wearable devices during data acquisition. Wearable body sensors are utilized to collect specific body movements in HAR approaches, which are then transformed into various signal patterns for classification using both machine and deep structure models [6].

The 3D action data collected by the wearable sensors is considered to be multivariate time series. The sampling rate of the data collected by these sensors has a higher sampling rate, enabling them to function in challenging and dim conditions. The inertial sensors have some challenges similar to that of vision-based sensors including the awkwardness of wearing them all the time, inadequate onboard power, and sensor drift [7]. To enhance accessibility through smartwatches and commercial fitness trackers, there is no consensus on the optimal sensor position when adopting a wrist-mountable form during data acquisition [8]. Recently in [9], a comprehensive study was conducted to identify a suitable machine leaner for adapting wrist-mountable sensors in HAR. It studied the placement of the lower limb sensors for the HAR. Optimizing sensor placement for the HAR is also a research interest. Public datasets from Inertial Measurement Units (IMU) are available, with data collected from different parts of the human body such as the chest, knee, ear, wrist, arm, ankle, and waist using various parametric settings [10]. Depending on the type of activities, there is variation in data acquisition that needs to be investigated. Simple activities with coarse granularity, such as sitting and walking, are efficiently recognized using the low sampling rate-based sensor placed on the waist [11]. But, while detecting the combinatorial activities with finer granularity such as driving and eating, a satisfactory performance is not produced by the single sensor attached to the waist [12].

Effective machine learning techniques are necessary to accurately identify human activities using data from multiple sensors and signals. Support Vector Machine (SVM) and Hidden Markov Models (HMM) are machine learning models recently developed for the HAR [13]. However, the recognition performance over human activities is further improved by the adoption of deep learning methods and these models produce higher recognition accuracy. The recognition of human actions and activities is challenging in various situations due to the large variability in body movements [14]. On the other hand, it is difficult to recognize human behaviors and activities from the multimodal body sensor data and signals. Most of the conventional approaches mainly focused on single modalities for the identification of human activities that do not provide robust outcomes and it is not practical in healthcare applications [15]. Increasing the HAR accuracy over the multi-modal data has been a challenging task over the multi-modal data because it learns only fewer numbers of features from the sensor data. To improve the HAR performance, combinations of sensors like gyroscope sensors and accelerometer sensors are used for the acquisition of multimodal sensor data. Convolutional Neural Network (CNN) and Deep Belief Network (DBN) are the most widely used approaches for HAR, known for achieving higher recognition accuracy [16]. The most discriminative and relevant features from the data have been extracted to provide better recognition outcomes among these approaches [17]. However, recognizing HAR in real-world scenarios is challenging and time-consuming to produce accurate recognition outcomes. Moreover, the gradient vanishing issue is an important problem in the deep learning-aided HAR models [18]. Therefore, a new hybrid deep learning network with an attention mechanism is developed for the recognition of human action to get higher recognition accuracy.

The significant contributions of the proposed hybrid deep learning-based human action recognition models are given in the following points:

- To develop a human action recognition model using a hybrid deep learning network with an attention strategy. This model is designed to monitor human activities and interactions, benefiting applications such as patient monitoring, video surveillance, and suspicious activity detection.
- To implement an EAHO for updating the random parameters based on fitness values. This optimization technique fine-tunes the parameters of the 1DCNN and GRU to enhance the recognition performance of human actions.
- To develop an AHDAN model for monitoring human activities, the AHDAN model incorporates the 1DCNN and GRU with an attention mechanism. Parameter optimization enhances performance by maximizing accuracy and minimizing False Positive Rate (FPR).
- The efficacy of the recommended HAR model is validated with the conventional algorithms and techniques by various observation measures.

The remaining sections describe the proposed human action recognition model using a hybrid deep structure with an attention mechanism given in the following points. The previously developed human action recognition schemes with their advantages and demerits are provided in Section 2. The problem statement is also included in this section. The collection of signals and data using the sensors is provided in Section 3. Moreover, this section includes the architectural view of the recommended HAR framework. The extraction of wave features from the signals and the proposed EAHO algorithm are briefly elucidated in Section 4. The developed AHDAN model with the basic function of 1DCNN and GRU is given in Section 5. The experimental setup and the comparative analysis are provided in Section 6. The conclusion of the proposed HAR framework is elucidated in Section 7.

## 2 Related Works

In 2019, Gumaei et al. [19] implemented an intelligent HAR framework utilizing a multi-sensor-based hybrid deep structure mechanism. Here, the Gated Recurrent Unit (GRU) and Simple Recurrent Units (SRUs) have been integrated into the neural networks. The deep SRUs process the multimodal input data sequence by utilizing their ability to store internal memory states. After that, the amount of past information passed to the future state has been learned and stored using the deep GRUs for rectifying the instability or fluctuations in accuracy and gradient vanishing issues.

In 2022, Roche et al. [20] introduced a HAR framework for leveraging the benefits of multimodal machine learning and sensor fusion. Subjects performed activities using RGB and point cloud data initially described using a 3-D modified Fisher vector model and Regions-based-CNN (R-CNN). The outputs of the human activity classification were accurate when evaluated through custom-accustomed multimodal data.

In 2021, Buffelli et al. [21] proposed a purely attention-based strategy for the recognition of human activities. The analysis showed that the performance of the developed attention-based human action recognition model was significantly superior to previous methods. The personalizing human activity recognition approach attained greater importance in terms of F1 score.

In 2023, Wang et al. [22] offered a HAR approach using the "Multidimensional Parallel Convolutional Connected (MPCC)-based deep learning" method based on multi-dimensional data. Multi-dimensional convolutional kernels have been fully used to recognize human activities. The diversity of feature information was improved by incorporating "Multi-scale Residual Convolutional Squeeze-and-Excitation (MRCSE)" blocks. The developed model's performance was confirmed through tenfold cross-validation.

In 2021, Ahmad et al. [23] implemented a "Multistage Gated Average Fusion (MGAF)"-based HAR framework that extracted and fused the features of CNN from all layers. The MGAF, known as Signal Images (SI) and Sequential Front View Images (SFI), has transformed the inertial sensor data into depth images.

In 2023, Hu et al. [24] presented an innovative data fusion methodology based on multimodal data, which was skeleton-guided and it modified the RGB, depth, and optical flow information into images. The transformation into depth images has been accomplished concerning key point sequences. The multi-modal fusion network has comprehensively extracted the actions of the pattern, significantly increasing recognition effectiveness for rapid inference speed. Finally, extensive experiments were made to show its efficacy on two large-scale datasets likely and the results achieved exciting recognition outcomes.

In 2020, Yudistira et al. [25] presented a multimodal CNN that captured the multimodal correlations over arbitrary timestamps to recognize human actions. By using the deep CNN, the temporal and spatial features were needed for recognizing the actions, a fusion of these two streams and decreasing overfitting were the open problems. After, the pre-trained CNN was learned through the Shannon fusion-based correlation network. The simple fully connected layers of the correlation network captured spatiotemporal correlations from long-duration videos over arbitrary times. The effectiveness of the multi-modal correlation was confirmed by comparing it to conventional fusion methods using the HMDB-51 and UCF-101 datasets.

In 2019, Chung et al. [26] introduced a recognition scheme for human actions using deep learning. The data collection process was done through an Android mobile device and eight body-worn IMU sensors. The human activity data has been trained using Long Short-Term Memory (LSTM) was taken in both controlled and real-world scenarios. The experimental results demonstrated that the model effectively identified daily living actions such as driving and eating activities.

## 3 Structural Demonstration of Proposed Human Action Recognition Model over Multi-Modal Data Using Adaptive and Network

### *3.1 Architectural View of Developed Human Action Recognition Model*

The recognition of fine-grained activities of human action is the challenging task of conventional human action recognition models. Fine-grained activities play a crucial role in identifying the action sequence and sub-actions, which offer more detailed information about the context. As a result, deep learning strategies are employed for human action recognition, yielding average accuracy levels. Furthermore, the ability of the current deep structure-based HAR approach to generalize is limited. Hence, advanced methodologies are needed to provide a better balance between the accuracy of HAR and the resources utilized for the recognition process. Additional data is required for accurate human action recognition. In contrast, traditional deep learning techniques have low activity detection efficiency and require high computation time for human action recognition. Additionally, identifying everyday human activities poses a significant challenge. These challenges are addressed and solved by the newly recommended advanced deep structure-aided HAR framework. This model effectively identifies descriptive functions from wave signals, leading to accurate human action recognition. The advanced deep learning strategy used in the human action recognition model demonstrates high scalability. The architectural view of the implemented advanced deep learning-based HAR is shown in Fig. 1.

A new human action recognition approach is implemented via an advanced deep structure strategy for classifying manually trimmed actions with higher recognition accuracy. The input signals wanted for the recognition of human action are collected using traditional databases like Kaggle. Gathered sensor signals are given to the wave feature extraction stage, where wave features like P, T, U, and QRS complex waves are attained for the recognition of human action. Then, the collected bio-signals and the sensor data are passed as the single combined input to the proposed attention-aided hybrid deep structure for the recognition of human actions. The feature extraction steps involve extracting both the spatial and temporal features. The spatial and temporal features are obtained from the human motion signals from Dataset 2. Combining features from sensor signals and data enhances action recognition by providing detailed information for training the deep learning model. Incorporating ECG wave features, both spatial and temporal, helps the model distinguish between different human actions such as jogging, walking, and running. These make an additional source and increase the effectiveness of the HAR technique.

This network is developed using 1DCNN and GRU networks to attain better recognition outcomes. The recognition accuracy of the implemented AHDAN is further increased by optimizing hyperparameters of the 1DCNN and GRU. Here, the hyperparameters such as “hidden neuron count and epoch size” from both 1DCNN and GRU are optimized by the proposed EAHO. This optimal tuning of parameters maximizes the accuracy and minimizes the FPR over the identification of human action. The performance of the suggested HAR model is validated with the conventional heuristic algorithms and previously developed human action recognition models to show effectiveness.

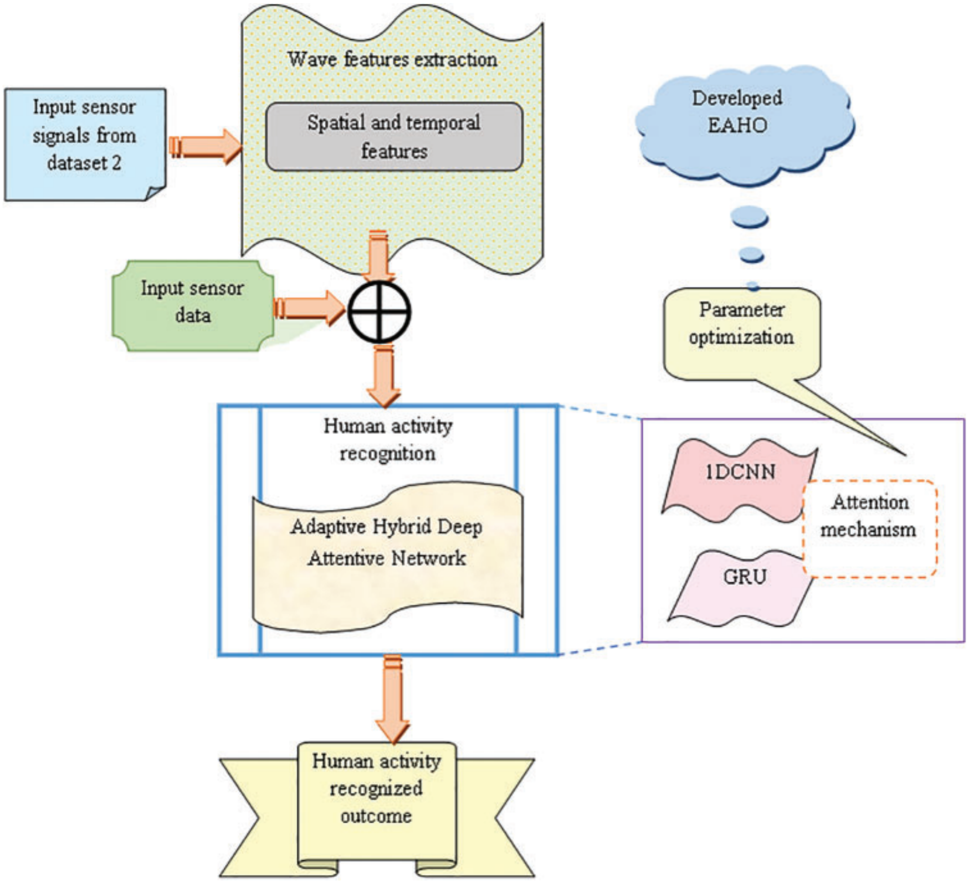

Input sensor signals from dataset 2
Wave features extraction
Spatial and temporal features
Developed EAHO
Input sensor data
Parameter optimization
Human activity recognition
Adaptive Hybrid Deep Attentive Network
1DCNN
Attention mechanism
GRU
Human activity recognized outcome

**Figure 1:** Block schematic demonstration of recommended advanced deep learning-based HAR model

### *3.2 Multi-Modal Dataset Collection*

Sensors are used for the collection of bio-signals and data to recognize human action. This information is acquired from traditional sources for processing the proposed HAR model. For the experimentation of the proposed model, 75% of the entire data has been used for training purposes, and the remaining 25% is used for testing purposes. This can ensure the validation of the proposed methodology.

**Dataset 1 (Sensor Data):** The wanted sensor data are collected from the "Human Activity Recognition with Smartphones" Dataset from the source of https://www.kaggle.com/datasets/uciml/human-activity-recognition-with-smartphones (accessed 01 February 2024). This HAR database is generated with the recordings of a total of 30 study participants with the help of embedded initial sensors mounted on the smartphone. This dataset contains the "estimated body acceleration and Triaxial acceleration from the accelerometer", 561 feature vectors with the "frequency and time domain attributes, Triaxial angular velocity from the gyroscope, identifier of the subject and its activity label". With the support of the embedded gyrometer and accelerometer, "3-axial linear acceleration and 3-axial angular velocity" are captured at the constant frequency of 50 Hz.

**Dataset 2 (Sensor Signals):** The sensor signals required for the recognition of human action are taken from the dataset of "UTD Multimodal Human Action Dataset (UTD-MHAD)", which is attained from the source of https://personal.utdallas.edu/~kehtar/UTD-MHAD.html (accessed 01 February 2024). Low-cost wearable IMUs are used for the collection of the HAR dataset. The sampling rate of the IMU sensor is about 50 Hz. This UTD-MHAD database contains 27 different actions including crossing arms in the chest, walking in place, squatting, and so on. These 27 actions are performed by a total of 8 subjects included in this dataset. After removing all corrupted sequences, 861 data sequences are presented. Four types of data, including skeleton, colour, inertial, and depth information, are used for the recognition of human actions. The sample action image and the sensor image for swipe left are given in Figs. 2 and 3, sample action image and the sensor image for swipe right are given in Figs. 4 and 5 and for wave are given in Figs. 6 and 7. Data is often in signal or raw format. To simplify representation, sensor signals are converted into images to extract spatial and temporal features easily.

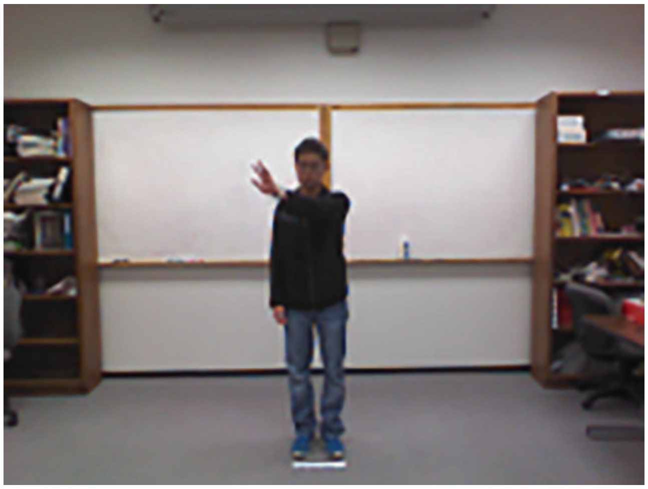


**Figure 2:** Sample image for swipe left

## 4 Description of Proposed Optimization Strategy for Parameter Tuning

### *4.1 Proposed EAHO*

A new EAHO algorithm based on AHO is developed to get the optimum values of the design variables to improve the recognition performance over human action. The hyperparameters are optimized from the developed AHDAN to achieve the maximum performance during human action recognition. The hyperparameters like "hidden neuron counts and epoch count" are tuned from the 1DCNN model and GRU approach for improving the HAR efficacy. This optimal tuning of parameters improves the effectiveness by maximizing accuracy and minimizing FPR.

The average performance metric, known as mean fitness, is also utilized in developing the adaptive concept. The conventional AHO algorithm uses a randomly assigned parameter for controlling the optimization procedure. In some cases, related to data scarcity, having an unequal distribution of data can lead to finding the optimal solution prematurely or prolonging the convergence time. Both cases are ineffective in finding an optimal solution. One way of overcoming this problem is adjusting the random parameter using a cost function.

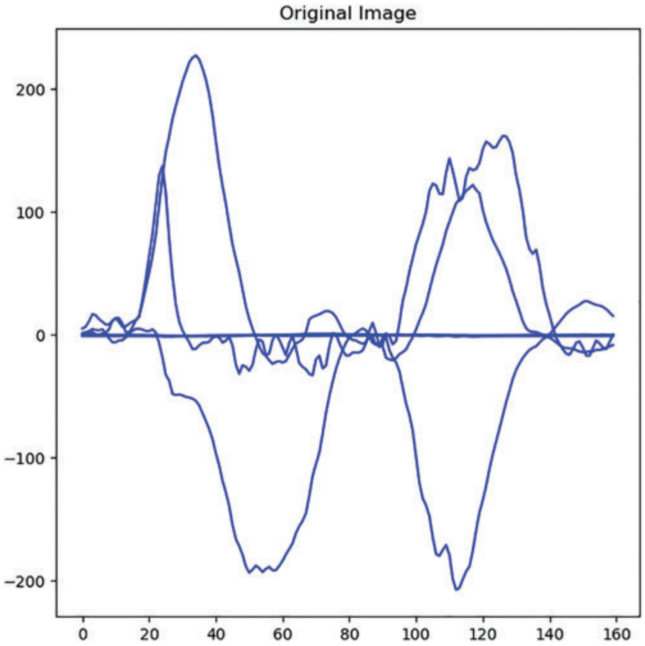


**Figure 3:** Sensor image for swipe left

**Figure 4:** Sample image for swipe right

In the proposed EAHO, the random parameter $\vartheta_1$ is improved based on the best, worst, and the mean fittest solutions. The updated adaptive concept $\vartheta_1$ based on fitness is provided in below Eq. (1):

$$\vartheta_1 = \frac{(Fit_{Best} * Mean_{Fit})}{(Fit_{Worst} * Mean_{Fit})} \tag{1}$$

The best fitness determined from the algorithm is denoted by $Fit_{Best}$, the worst fitness calculated from the algorithm is indicated by $Fit_{Worst}$ and the mean fitness attained from the algorithm is denoted by $Mean_{Fit}$. The proposed updated random parameter-based EAHO provides better searching ability by incorporating data retention and higher accuracy abilities.

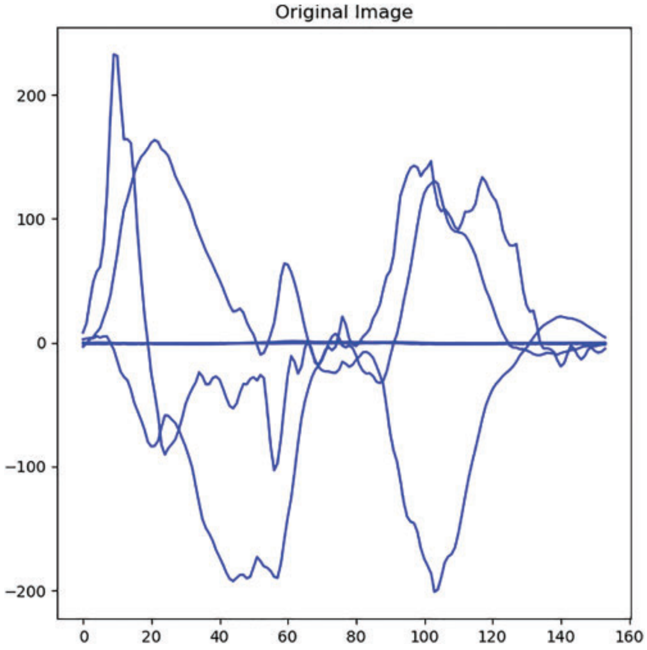


**Figure 5:** Sensor image for swipe right

**Figure 6:** Sample image for wave

The algorithmic steps involved in the proposed EAHO are briefed as the step-by-step procedure below:

Step 1: First, the random solution is populated, and each candidate in the solution refers to the sensor signals and the data.

Step 2: It involves defining the parameters, attributes, and fitness functions for the optimization process.

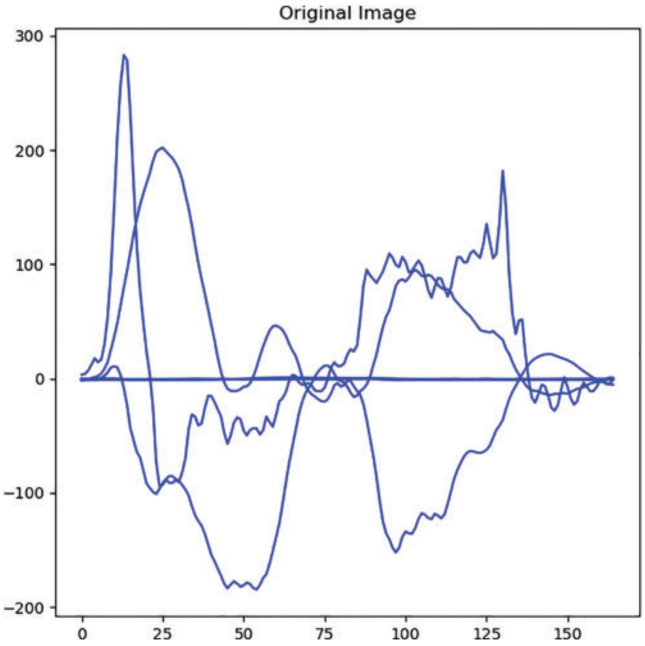


**Figure 7:** Sensor image for wave

Step 3: The elite or the optimal solution candidate is found using the fitness function, and now the solution with the top fitness candidate, worst fitness candidate, and the mean fitness candidate are represented as $Fit_{Best}$, $Fit_{Worst}$ and $Mean_{Fit}$, respectively.

Step 4: Find the adaptable attribute $\vartheta_1$ for regulating various update procedures.

Step 5: Update the positions of the candidate according to the shooting and jumping behavior of the Archerfish as defined by the existing AHO algorithm.

Step 6: After the set of the iteration, find the optimal solution representing the suitably hidden neurons and epoch for the HAR recognition.

## 5 Classification of Spatial and Temporal Features for the Recognition of Human Action Using Adaptive and Attentive Deep Learning Model

### *5.1 1DCNN Model*

1DCNNs [27] offer an efficient solution for recognizing human actions. The convolution filters are trained to adaptively extract meaningful spatiotemporal parameters from the input channels. As a result of this process, the input sequence is transformed into a feature space, enhancing recognition capabilities. Mostly, the one-dimensional convolution network has significantly reduced the number of trainable parameters, making it computationally efficient for HAR.

In the one-dimensional CNN, the $m^{th}$ layer associated with the $\mathrm{X}^m$ hidden state is represented in Eq. (2).

$$\mathrm{X}^m = \begin{cases} \mathrm{P}_{\mathrm{X}}^m * \ell + Bs_X^m, & \mathit{if}\ m = 1 \\ \mathrm{P}_{\mathrm{X}}^m * \mathrm{X}^{m-1} + Bs_X^m, & m = 2, 3, \ldots, M \end{cases} \tag{2}$$

The convolution operation is indicated by $*$, the biases are represented by the term $Bs_X^m$ and the convolution filter of the 1DCNN is denoted by $\mathrm{P}_{\mathrm{X}}^m$, which has the dimensionality of $g^m \times e^{m-1} \times e^l$. Here, the term $\ell$ is the feature matrix of the fine-tuned network. The filter size is indicated by the terms $g^m$ and $e^m$, correspondingly. The matrix for the hidden state for the $m^{th}$ layer is indicated by the term $\mathrm{X}^m$, which has the dimensionality of $e^m \times \ell^m$. Here, the term $\ell^m$ indicates the length of the filter. The hidden state outcomes are passed to the activation function of ReLU, which is depicted in Eq. (3).

$$\mathrm{X}^m = \begin{cases} 0, & if\ \mathrm{X}^m \leq 0 \\ \mathrm{X}^m, & otherwise \end{cases} \tag{3}$$

The last hidden layer's hidden state matrix is flattened by the vector $\vec{j}^{M}$ for fully connecting with the output layer, which is indicated by Eq. (4).

$$\vec{k} = \mathrm{P}_k \cdot \vec{j}^{M} + \vec{Bs}_k, \quad \vec{k} \in \Re^{\mathrm{H}} \tag{4}$$

The affine transformation of $\vec{j}^{M}$ is defined by the term $\vec{k}$ that has been associated with the bias matrix $\vec{Bs}_k$ and weight matrix $\mathrm{P}_k$. The outcome of the fully connected layer is given to the softmax layer and that layer produces the predicted probability score. The structural demonstration of the 1DCNN approach is provided in Fig. 8.

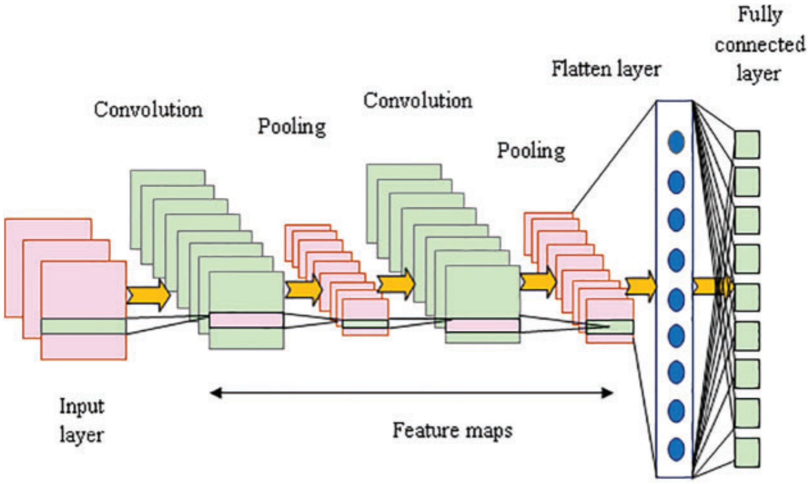


**Figure 8:** Sample schematic illustration of 1DCNN

### 5.2 *GRU Model*

The GRU model [28] is used for the recognition of human action, where all the hidden layers are interconnected multiple times. There are only two gates presented, "reset gate and update gate". Here, the reset gate is denoted by the term $S_g$ and $U_g$. The update gate primarily controls how much information from the previous hidden state influences the current state. When the update gate has a higher value, it gathers more details from the previous state. The reset gate determines which information from the previous state should be disregarded. When the reset gate has a lower value, it results in more information being ignored. Reset gates are used to handle short-term dependencies, while update gates are utilized for long-term dependencies.

The function to be followed in the reset gate is represented in Eq. (5).

$$S_g = \vartheta \left( T_S \cdot \left\lfloor hi_{g-1}, Ex_g \right\rfloor \right) \tag{5}$$

The weight matrix of the reset gate is indicated by $T_s$, the logistic sigmoid function is indicated by the term $\vartheta$ and the information of the previous hidden state is represented by $hi_{g\text{-}1}$. Moreover, the term $Ex_g$ denotes the input of wave features applied to the input gate.

The function to be performed in the update gate is indicated in Eq. (6).

$$U_g = \vartheta \left( T_U \cdot \left\lfloor hi_{g-1}, Ex_g \right\rfloor \right) \tag{6}$$

The prior hidden and current hidden state functions are expressed in Eqs. (7) and (8), respectively.

$$\tilde{hi}_g = \tan\text{h} \left( T_{\tilde{hi}} \cdot \left[ S_g * hi_{g-1}, Ex_g \right] \right) \tag{7}$$

$$hi_g = \left( 1 - U_g \right) * hi_{g-1} + U_g * \tilde{hi}_g \tag{8}$$

The function of the output gate is expressed in the following Eq. (9):

$$Out_g = \vartheta \left( T_{Out} \cdot hi_g \right) \tag{9}$$

From these expressions, the long-term dependencies are effectively learned using the GRU model. This model provides better human action recognition outcomes.

### 5.3 AHDAN-Based Human Action Recognition

A new HAR approach is implemented using an intelligent deep structure strategy for monitoring the movement and activities of humans, which is helpful in fields like healthcare, human/computer interface, gaming, intelligent monitoring, and sports performance analysis sector. Sensors are utilized to collect necessary data and signals, enhancing the efficiency of HAR through advanced feature extraction capabilities. The spatial and temporal features extracted from the data, along with the sensor data, are inputted into the proposed AHDAN model for the recognition of human activities. The proposed AHDAN model is developed with the assistance of the attentive-based 1DCNN and GRU models. The attention mechanism in the hybrid deep learning network replicates the way biological systems internally monitor activities. Detailed information from the data is acquired to improve the recognition of human actions. Significant data is captured and emphasized by assigning weights in the attention mechanism. After the weights are distributed, feature vectors are derived. Subsequently, the accumulated features are aggregated to generate the ultimate feature matrix.

The hyperparameters like "hidden neuron count, and epoch count" from the 1DCNN and GRU model are optimized for further improving the local feature extraction capability of the proposed AHDAN model. Through parameter optimization, the proposed AHDAN model enhances accuracy and reduces $FPR$ in human action recognition. The objective function of the proposed AHDAN method with the optimization process is given in Eq. (10).

$$Obj_{fun} = \underset{\left\{ Hid_{x*}^{CNN}, Ep_{y*}^{CNN}, Hid_{u*}^{GRU}, Ep_{v*}^{GRU} \right\}}{\arg\min} \left( \frac{1}{ARY} + FPR \right) \tag{10}$$

The term $Obj_{fun}$ denotes the fitness function of the proposed AHDAN, the maximized accuracy is indicated as $ARY$, and minimized $FPR$ is signified as $FPR$. The optimized hidden neuron is signified as $Hid_{x*}^{CNN}$, which is present in the interval of $[5, 255]$ and the optimally tuned epoch count from 1DCNN is denoted as $Ep_{v*}^{CNN}$, lies in the range of [5, 50]. The optimized hidden neuron from the GRU present in between [5, 255] is represented as $Hid_{u*}^{GRU}$ and the optimized epoch from the GRU is signified as $Ep_{v*}^{GRU}$, which is tuned in the range between [5, 50]. The accuracy and FPR formula is given as below in Eqs. (11) and (12), respectively:

$$ARY = \frac{X^{pos} + X^{ngv}}{X^{pos} + X^{ngv} + Y^{pos} + Y^{ngv}} \tag{11}$$

$$FPR = \frac{Y^{pos}}{X^{ngv} + Y^{pos}} \tag{12}$$

Here, the terms "$X^{pos}$, $X^{ngv}$, $Y^{pos}$ and $Y^{ngv}$ represent the true positives, true negatives, false positives and false negatives".

The human action recognition process using the proposed AHDAN-based HAR is shown in Fig. 9.

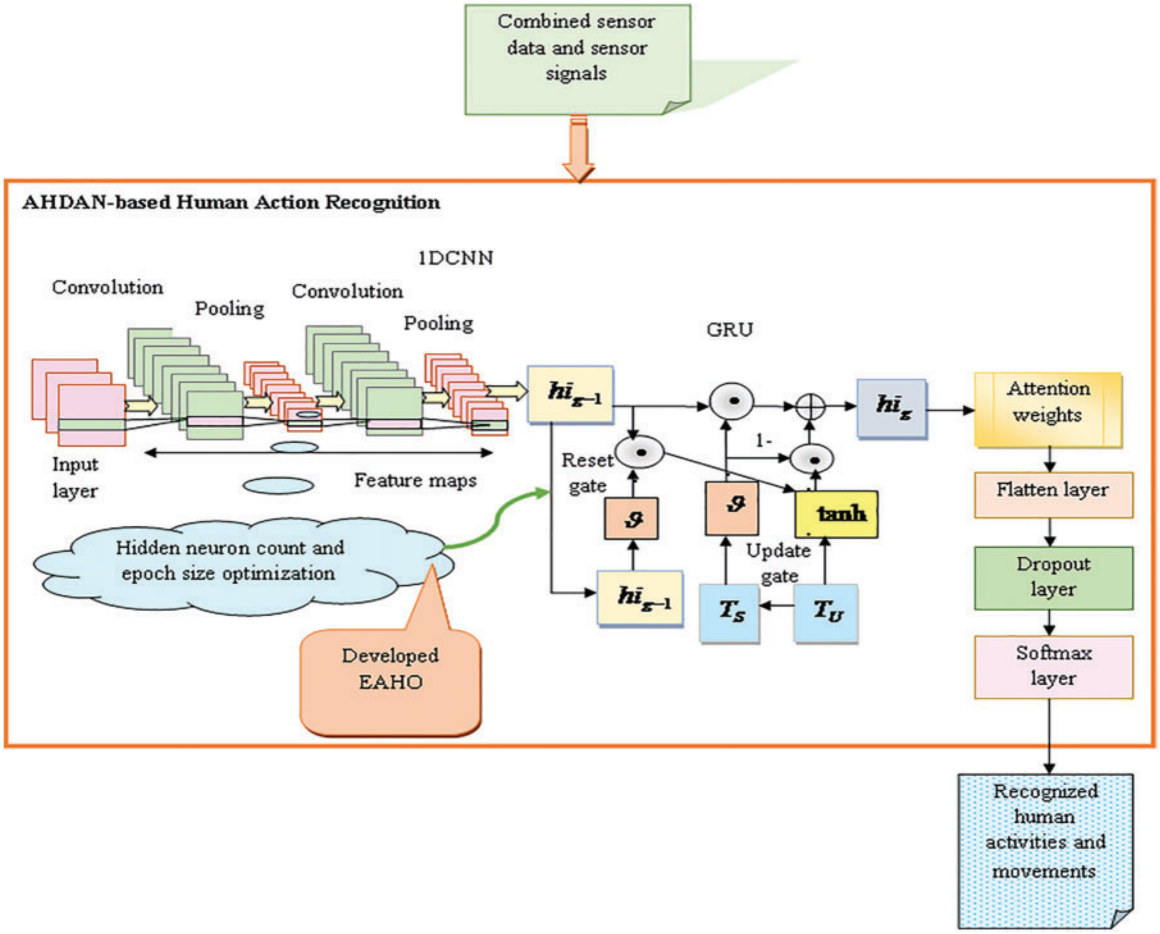


**Figure 9:** Human action recognition process using proposed AHDAN

## 6 Results and Discussion

### 6.1 Experimental Setup

A new model for recognizing human movements and activities has been developed using a combination of deep learning techniques and optimization strategies in Python, aiming for improved accuracy in action recognition. The recognition results of the new approach have been compared with traditional strategies and existing models to assess its performance in human action recognition. In the experiments, a population of 10, a chromosome length of 4, and a maximum of 50 iterations were used. The conventional heuristic algorithms like "Golden Eagle Optimizer (GEO) [29], Humboldt Squid Optimization Algorithm (HSOA) [30], Gorilla Troops Optimizer (GTO) [31], and Archerfish Hunting Optimizer (AHO) [32]" were considered for the evaluation of recognition performance. The previously developed human action recognition models using LSTM [33], 1DCNN [34], GRU [35], and 1DCNN-GRU [36] were considered for performing the experimental analysis. The cost function analysis, convergence evaluation, and positive as well as negative measure analysis were conducted for this experiment. Table 1 below provides an overview of the hardware and software details relevant to the study implementation.

**Table 1:** Performance validation of the proposed human action recognition model using hybrid deep learning

| Hardware/Software | Specification |
|---|---|
| Processor | Intel(R) Core(TM) i3-1005G1 CPU @ 1.20 GHz 1.19 GHz |
| RAM | 16.0 GB |
| System type | 64-bit operating system, x64-based processor |
| OS | Windows |
| Edition | Windows 11 pro |
| Software development environment | Python |
| Interpreter | Python 3.11 |

### 6.2 Performance Validation in Terms of K-Fold

Several positive measures such as F1-Score, accuracy, sensitivity, precision, and specificity are considered for the experimentation and also negative measures such as FPR, FNR, and FPR are taken for analyzing the performance. The effectiveness among the traditional models is provided in Fig. 10 and the performance validation among the heuristic strategies is depicted in Fig. 11. Experimental results indicate that the F1-Score of the presented framework outperforms LSTM by 6.35%, 1DCNN by 5.5%, GRU by 3.83%, and 1DCNN-GRU by 3.25% when considering a K-fold value of 3. The graphical results demonstrate that the effectiveness of human action recognition in the presented scheme is significantly improved through the use of conventional optimization strategies and previous approaches. As depicted below, the HAR performance by the proposed EAHO has been more effective than other comparative techniques. The comparative heuristics have yielded lower performance outcomes, highlighting the superior effectiveness of the proposed EAHO-based HAR strategy. The proposed EAHO demonstrates superior accuracy at 95.36043, validating its effectiveness in inaccurate activity recognition. Furthermore, performance metrics such as Recall, Specificity, Precision, FPR, FNR, NPV, FDR, F1-Score, and MCC show enhancements in the EAHO compared to the values of 95.2454, 95.47546, 95.46503, 4.52454, 4.754601, 95.47546, 4.534973, and 95.35509 in comparative techniques.

### *6.3 Numerical Analysis among Traditional Algorithms*

The performance validation of the proposed human action recognition model among the conventional heuristic algorithms and traditional models is illustrated in Table 2. The analysis results show that the precision of the presented human action recognition model is improved by 4.02% to GEO-AHDAN, 4.6% to GTO-AHDAN, 3.68% to HSOA-AHDAN, and 2.52% to AHOO-AHDAN. All the analysis results show that the efficacy of the recognition model is greatly enhanced than the conventional techniques while performing the recognition of human actions among the sensor data.

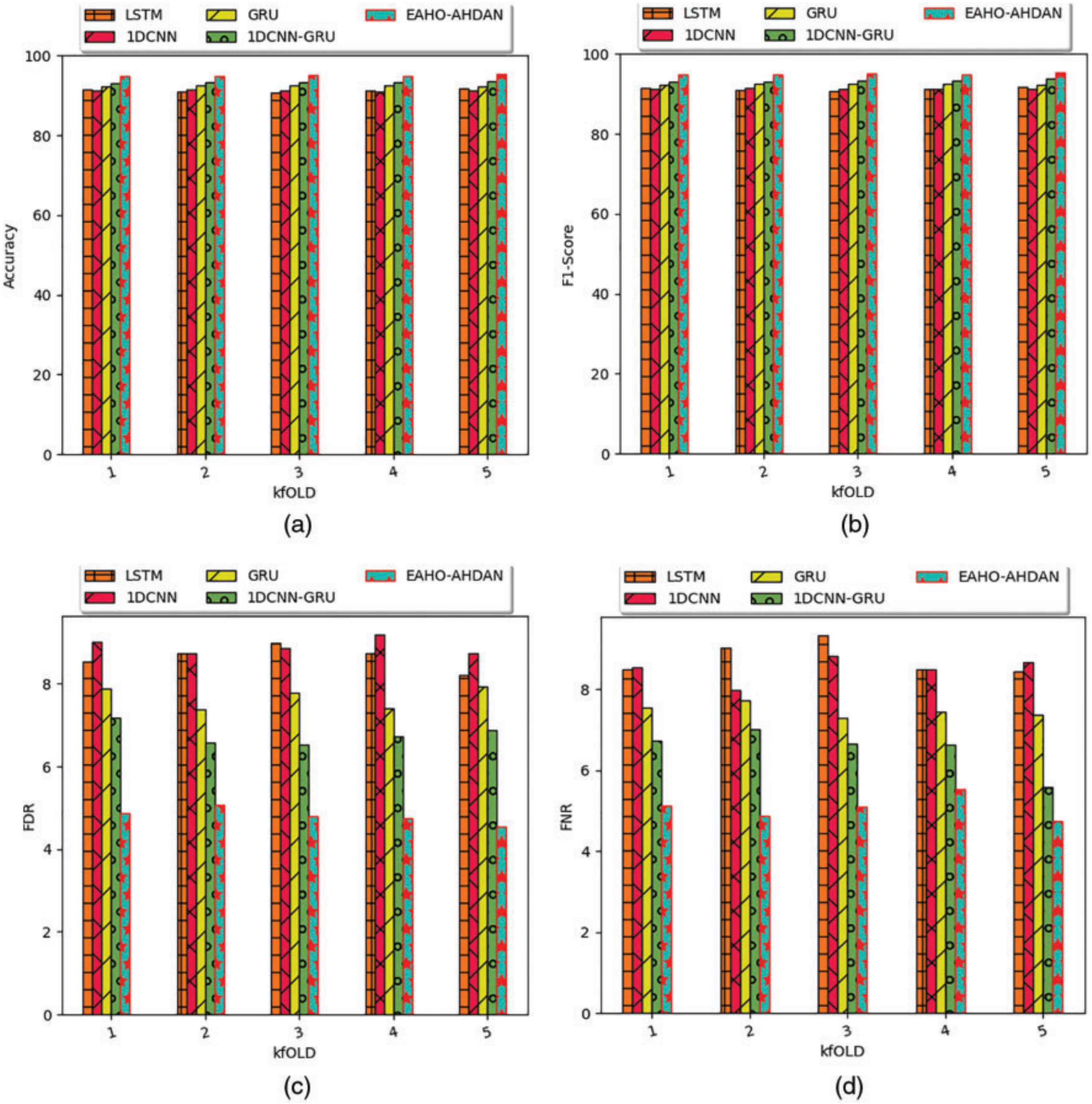
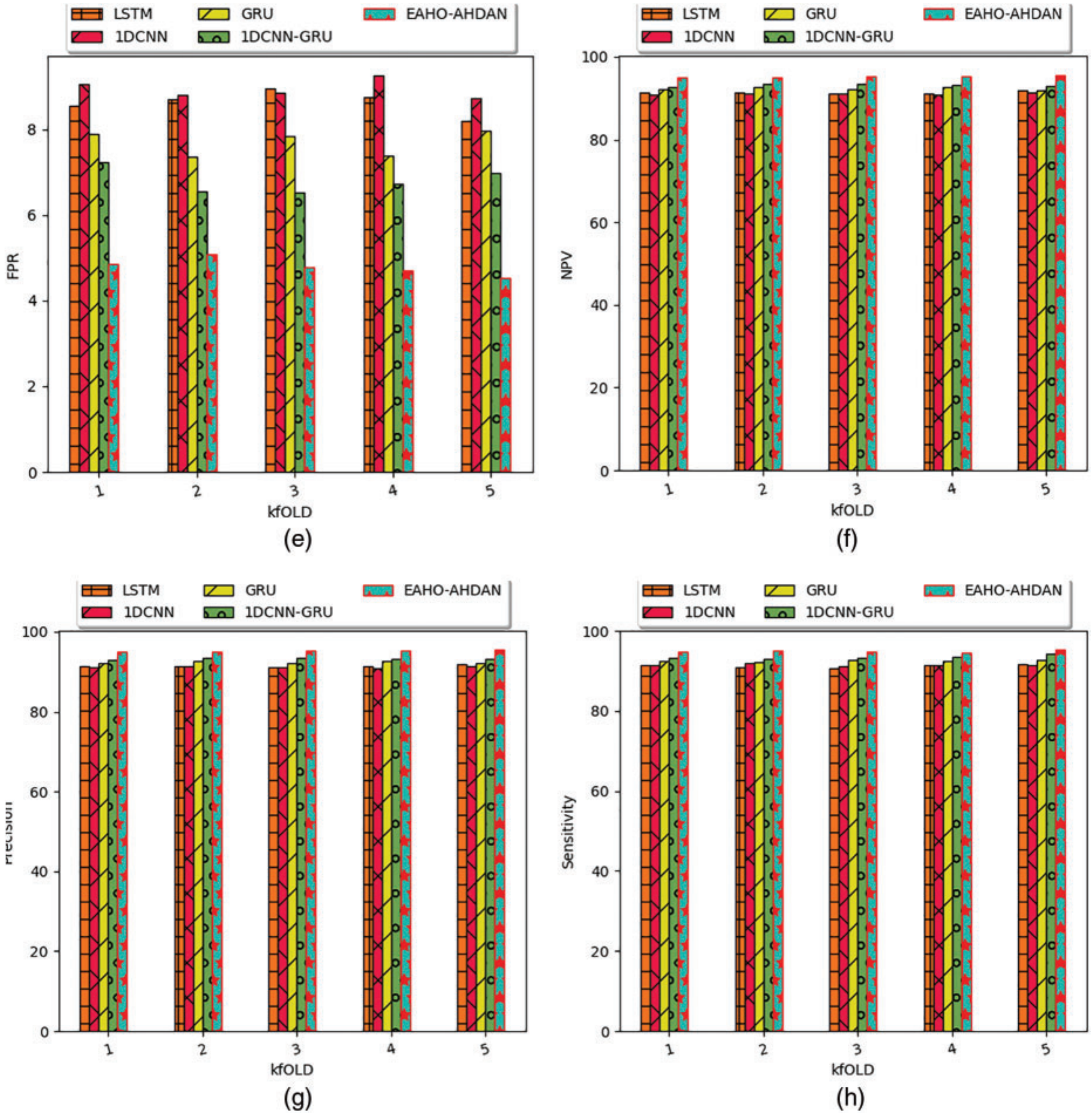


**Figure 10:** (Continued)



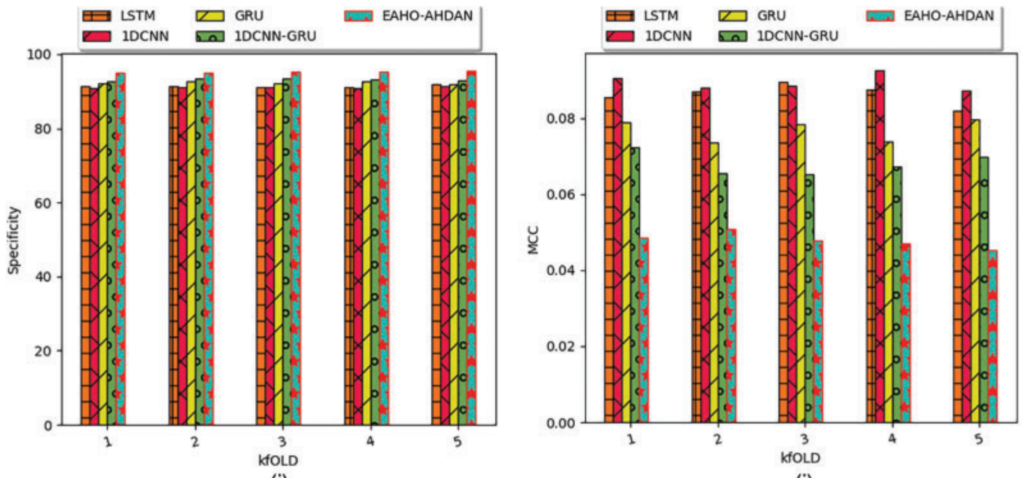
(e)

(f)

(g)

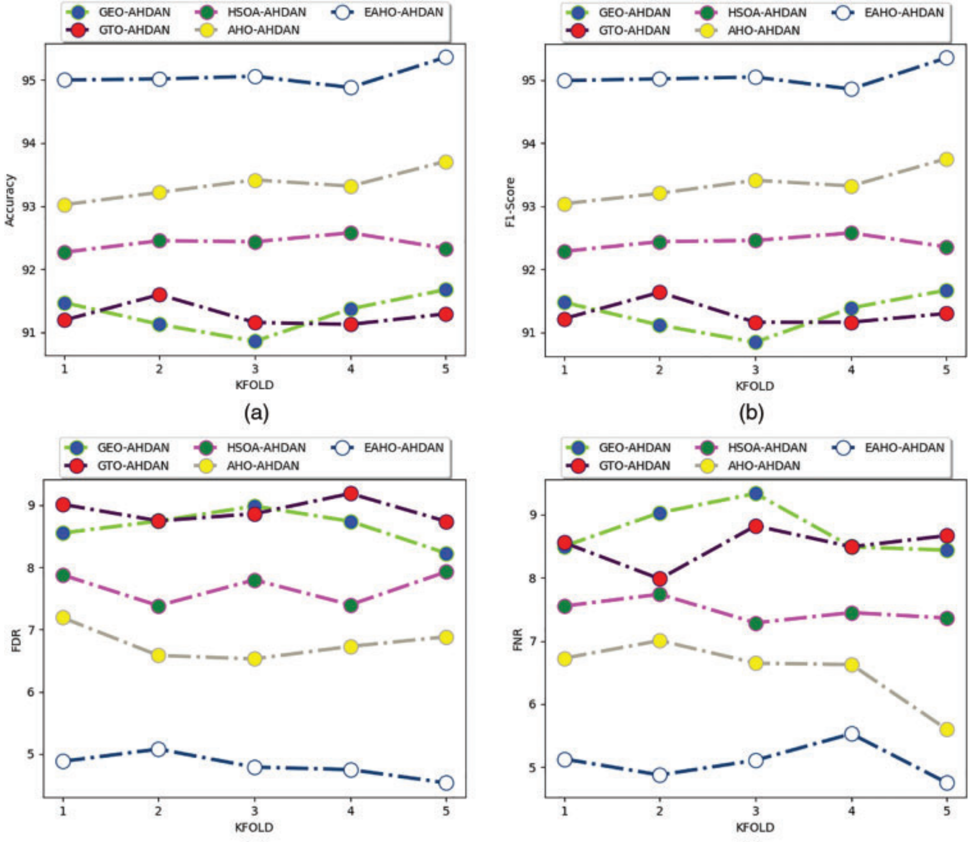
(h)

**Figure 10:** (Continued)



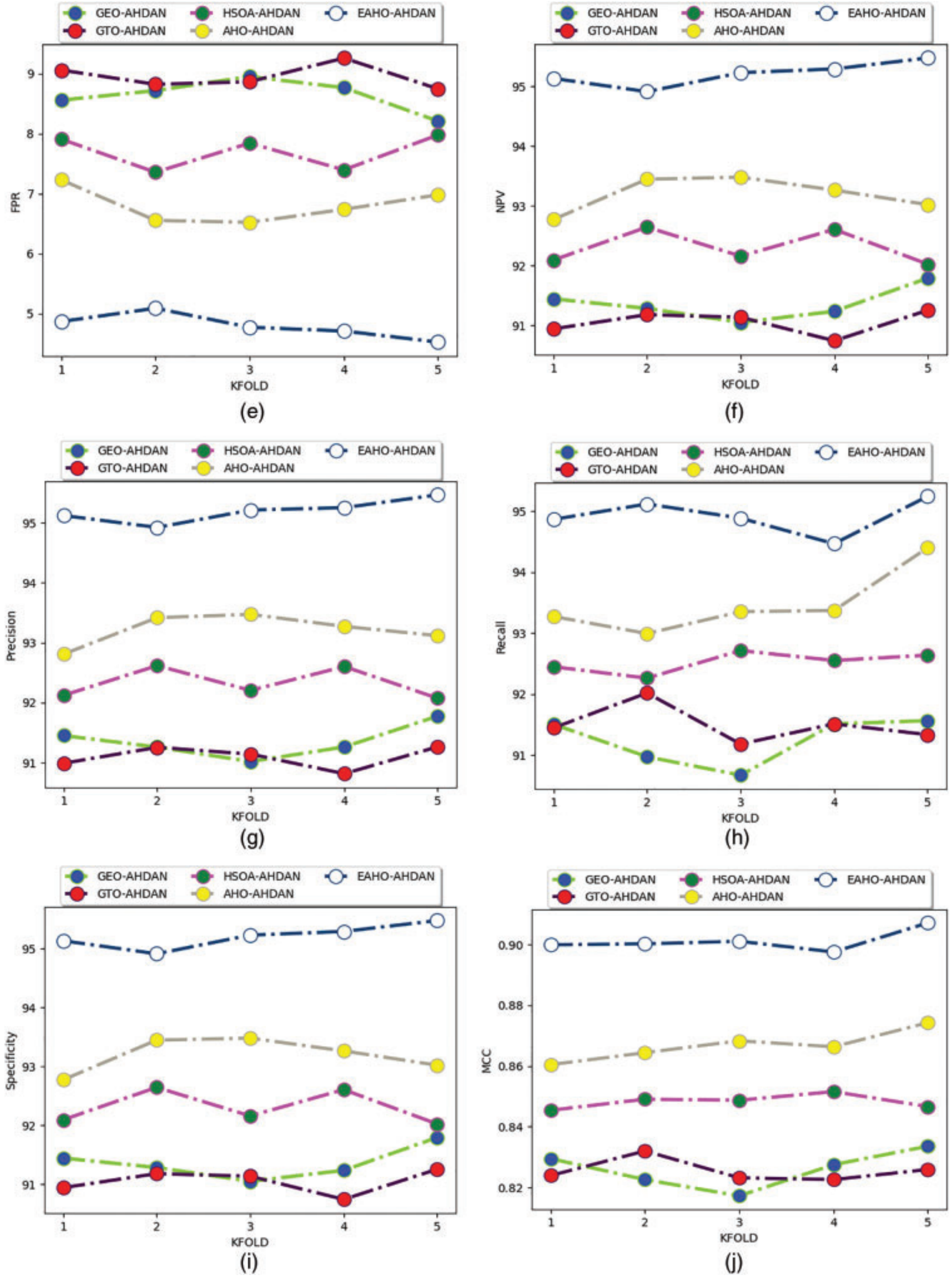
(i) (j)

**Figure 10:** Human action recognition performance of the presented approach among the prior works in regards to "(a) accuracy, (b) F1-Score, (c) FDR, (d) FNR, (e) FPR, (f) NPV, (g) precision, (h) sensitivity, (i) specificity, and (j) MCC" by varying K-fold



(a) (b) (c) (d)

**Figure 11:** (Continued)



(e) (f)

(g) (h)

(i) (j)

**Figure 11:** Human action recognition efficiency of the proposed model among the heuristic algorithms in regards to "(a) accuracy, (b) F1-Score, (c) FDR, (d) FNR, (e) FPR, (f) NPV, (g) precision, (h) recall, (i) specificity, and (j) MCC" by varying K-fold

**Table 2:** Performance validation of the proposed human action recognition model using hybrid deep learning

**Among algorithms**

| Performance measures | GEO-AHDAN [29] | GTO-AHDAN [31] | HSOA-AHDAN [30] | AHO-AHDAN [32] | EAHO-AHDAN |
|---|---|---|---|---|---|
| Accuracy | 91.67945 | 91.29601 | 92.33129 | 93.71166 | 95.36043 |
| Recall | 91.56442 | 91.33436 | 92.63804 | 94.40184 | 95.2454 |
| Specificity | 91.79448 | 91.25767 | 92.02454 | 93.02147 | 95.47546 |
| Precision | 91.77556 | 91.26437 | 92.07317 | 93.11649 | 95.46503 |
| FPR | 8.205521 | 8.742331 | 7.97546 | 6.978528 | 4.52454 |
| FNR | 8.435583 | 8.665644 | 7.361963 | 5.59816 | 4.754601 |
| NPV | 91.79448 | 91.25767 | 92.02454 | 93.02147 | 95.47546 |
| FDR | 8.224443 | 8.735632 | 7.926829 | 6.88351 | 4.534973 |
| F1-Score | 91.66987 | 91.29935 | 92.35474 | 93.75476 | 95.35509 |
| MCC | 0.833591 | 0.82592 | 0.846642 | 0.874316 | 0.907211 |

**Among techniques**

| Performance measures | LSTM [33] | 1DCNN [34] | GRU [35] | 1DCNN-GRU [36] | EAHO-AHDAN |
|---|---|---|---|---|---|
| Accuracy | 89.68558 | 91.71779 | 92.06288 | 92.52301 | 95.36043 |
| Recall | 89.11043 | 90.87423 | 92.10123 | 93.32822 | 95.2454 |
| Specificity | 90.26074 | 92.56135 | 92.02454 | 91.71779 | 95.47546 |
| Precision | 90.1474 | 92.4337 | 92.03065 | 91.84906 | 95.46503 |
| FPR | 9.739264 | 7.43865 | 7.97546 | 8.282209 | 4.52454 |
| FNR | 10.88957 | 9.125767 | 7.898773 | 6.671779 | 4.754601 |
| NPV | 90.26074 | 92.56135 | 92.02454 | 91.71779 | 95.47546 |
| FDR | 9.852599 | 7.566303 | 7.969349 | 8.150943 | 4.534973 |
| F1-Score | 89.62592 | 91.64733 | 92.06593 | 92.58273 | 95.35509 |

## 7 Conclusion

A new HAR model has been developed to monitor human activities and movements for various applications. Sensor data and signals needed for recognizing human activities were collected from online sources. Spatial and temporal features were then extracted from these signals. The collected sensor data and signals were integrated and provided to the implemented AHDAN for human action recognition. This network was built using 1DCNN and GRU models with an attention mechanism. Optimizing parameters improves recognition accuracy and reduces FPR. The proposed human action recognition model plays a crucial role in computer vision applications. The effectiveness of the implemented HAR scheme with hybrid deep learning was validated against conventional algorithms and previous models. The analysis results demonstrate that the presented framework achieved an accuracy of 95.36% in recognizing human actions. The recognition efficacy of the implemented HAR framework has significantly surpassed that of previous works.

**Acknowledgement:** All authors sincerely thank all organizations and institutions that have provided data and resources. Thanks to all the members of our research group, their suggestions and support have provided important help and profound impact on our research work.

**Funding Statement:** There is no funding received for this research work.

**Author Contributions:** Study conception and design: Ahmad Yahiya Ahmad Bani Ahmad; Methodology: Jafar Alzubi; Software and validation: Anguraju Krishnan; Formal analysis and investigation: Chanthirasekaran Kutralakani; Resources and data curation: Vincent Omollo Nyangaresi; Writing—original draft preparation, review and editing, visualization: Sophers James; Supervision: Chanthirasekaran Kutralakani. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** Data sharing not applicable–no new data generated.

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

[1] Z. Liu, Q. Cheng, C. Song, and J. Cheng, "Cross-scale cascade transformer for multimodal human action recognition," *Pattern Recognit. Lett.*, vol. 168, no. 10, pp. 17–23, 2023. doi: 10.1016/j.patrec.2023.02.024.

[2] X. Li, Q. Huang, and Z. Wang, "Spatial and temporal information fusion for human action recognition via center boundary balancing multimodal classifier," *J. Vis. Commun. Image Represent.*, vol. 90, 2023, Art. no. 103716. doi: 10.1016/j.jvcir.2022.103716.

[3] L. Chen, X. Liu, L. Peng, and M. Wu, "Deep learning based multimodal complex human activity recognition using wearable devices," *Appl. Intell.*, vol. 51, pp. 4029–4042, 2021.

[4] T. Singh and D. K. Vishwakarma, "A deep multimodal network based on bottleneck layer features fusion for action recognition," *Multimed. Tools Appl.*, vol. 80, no. 24, pp. 33505–33525, 2021.

[5] N. A. Choudhury and B. Soni, "Enhanced complex human activity recognition system: A proficient deep learning framework exploiting physiological sensors and feature learning," *IEEE Sens. Lett.*, vol. 7, no. 11, pp. 1–4, 2023. doi: 10.1109/LSENS.2023.3326126.

[6] A. Rezaei, M. C. Stevens, A. Argha, A. Mascheroni, A. Puiatti and N. H. Lovell, "An unobtrusive human activity recognition system using low resolution thermal sensors, machine and deep learning," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 1, pp. 115–124, Jan. 2023. doi: 10.1109/TBME.2022.3186313.

[7] N. A. Choudhury and B. Soni, "An efficient and lightweight deep learning model for human activity recognition on raw sensor data in uncontrolled environment," *IEEE Sens. J.*, vol. 23, no. 20, pp. 25579–25586, Oct. 2023. doi: 10.1109/JSEN.2023.3312478.

[8] H. Bi, M. Perello-Nieto, R. Santos-Rodriguez, and P. Flach, "Human activity recognition based on dynamic active learning," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 4, pp. 922–934, Apr. 2021. doi: 10.1109/JBHI.2020.3013403.

[9] Y. Liu, X. Liu, Z. Wang, X. Yang, and X. Wang, "Improving performance of human action intent recognition: Analysis of gait recognition machine learning algorithms and optimal combination with inertial measurement units," *Comput. Biol. Med.*, vol. 163, no. 9, 2023, Art. no. 107192. doi: 10.1016/j.compbiomed.2023.107192.

[10] X. Zhou, W. Liang, K. I. -K. Wang, H. Wang, L. T. Yang and Q. Jin, "Deep-learning-enhanced human activity recognition for internet of healthcare things," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6429–6438, Jul. 2020. doi: 10.1109/JIOT.2020.2985082.

[11] N. T. Hoai Thu and D. S. Han, "HiHAR: A hierarchical hybrid deep learning architecture for wearable sensor-based human activity recognition," *IEEE Access*, vol. 9, pp. 145271–145281, 2021. doi: 10.1109/ACCESS.2021.3122298.

[12] S. Mekruksavanich, A. Jitpattanakul, K. Sitthithakerngkiet, P. Youplao, and P. Yupapin, "ResNet-SE: Channel attention-based deep residual network for complex activity recognition using wrist-worn wearable sensors," *IEEE Access*, vol. 10, pp. 51142–51154, 2022. doi: 10.1109/ACCESS.2022.3174124.

[13] M. Ronald, A. Poulose, and D. S. Han, "iSPLInception: An inception-ResNet deep learning architecture for human activity recognition," *IEEE Access*, vol. 9, pp. 68985–69001, 2021. doi: 10.1109/ACCESS.2021.3078184.

[14] T. Mahmud, A. Q. M. S. Sayyed, S. A. Fattah, and S. -Y. Kung, "A novel multi-stage training approach for human activity recognition from multimodal wearable sensor data using deep neural network," *IEEE Sens. J.*, vol. 21, no. 2, pp. 1715–1726, 2020. doi: 10.1109/JSEN.2020.3015781.

[15] I. K. Ihianle, A. O. Nwajana, S. H. Ebenuwa, R. I. Otuka, K. Owa and M. O. Orisatoki, "A deep learning approach for human activities recognition from multimodal sensing devices," *IEEE Access*, vol. 8, pp. 179028–179038, 2020. doi: 10.1109/ACCESS.2020.3027979.

[16] K. K. Verma and B. M. Singh, "Deep multi-model fusion for human activity recognition using evolutionary algorithms," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 7, no. 2, pp. 44–58, 2021. doi: 10.9781/ijimai.2021.08.008.

[17] M. Moencks, V. D. Silva, J. Roche, and A. Kondoz, "Adaptive feature processing for robust human activity recognition on a novel multi-modal dataset," 2019, *arXiv:1901.02858*.

[18] M. Muaaz, A. Chelli, A. A. Abdelgawwad, A. C. Mallofré, and M. Pätzold, "WiWeHAR: Multimodal human activity recognition using Wi-Fi and wearable sensing modalities," *IEEE Access*, vol. 8, pp. 164453–164470, 2020. doi: 10.1109/ACCESS.2020.3022287.

[19] A. Gumaei, M. M. Hassan, A. Alelaiwi, and H. Alsalman, "A hybrid deep learning model for human activity recognition using multimodal body sensing data," *IEEE Access*, vol. 7, pp. 99152–99160, 2019. doi: 10.1109/ACCESS.2019.2927134.

[20] J. Roche, V. De-Silva, J. Hook, M. Moencks, and A. Kondoz, "A multimodal data processing system for LiDAR-based human activity recognition," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 10027–10040, Oct. 2022.

[21] D. Buffelli and F. Vandin, "Attention-based deep learning framework for human activity recognition with user adaptation," *IEEE Sens. J.*, vol. 21, no. 12, pp. 13474–13483, Jun. 2021. doi: 10.1109/JSEN.2021.3067690.

[22] Y. Wang *et al.*, "A multidimensional parallel convolutional connected network based on multisource and multimodal sensor data for human activity recognition," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14873–14885, Aug. 2023. doi: 10.1109/JIOT.2023.3265937.

[23] Z. Ahmad and N. Khan, "CNN-based multistage gated average fusion (MGAF) for human action recognition using depth and inertial sensors," *IEEE Sens. J.*, vol. 21, no. 3, pp. 3623–3634, Feb. 2021. doi: 10.1109/JSEN.2020.3028561.

[24] Z. Hu, J. Xiao, L. Li, C. Liu, and G. Ji, "Human-centric multimodal fusion network for robust action recognition," *Expert. Syst. Appl.*, vol. 239, no. 6, 2023, Art. no. 122314. doi: 10.1016/j.eswa.2023.122314.

[25] N. Yudistira and T. Kurita, "Correlation net: Spatiotemporal multimodal deep learning for action recognition," *Signal Process.: Image Commun.*, vol. 82, 2020, Art. no. 115731. doi: 10.1016/j.image.2019.115731.

[26] S. Chung, J. Lim, K. J. Noh, G. Kim, and H. Jeong, "Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning," *Sensors*, vol. 19, no. 7, 2019, Art. no. 1716. doi: 10.3390/s19071716.

[27] K. M. Lim, C. P. Lee, K. S. Tan, A. Alqahtani, and M. Ali, "Fine-tuned temporal dense sampling with 1D convolutional neural network for human action recognition," *Sensors*, vol. 23, no. 11, 2023, Art. no. 5271. doi: 10.3390/s23115276.

[28] R. Fajar, N. Suciati, and D. A. Navastara, "Real time human activity recognition using convolutional neural network and deep gated recurrent unit," in *2020 Int. Conf. Electr. Eng. Inform. (ICELTICs)*, Aceh, Indonesia, 2020, pp. 1–6.

[29] A. Mohammadi-Balani, M. D. Nayeri, A. Azar, and M. T. Yazdi, "Golden eagle optimizer: A nature-inspired metaheuristic algorithm," *Comput. & Ind. Eng.*, vol. 152, Feb. 2021, Art. no. 107050. doi: 10.1016/j.cie.2020.107050.

[30] M. V. Anaraki and S. Farzin, "Humboldt squid optimization algorithm (HSOA): A novel nature-inspired technique for solving optimization problems," *IEEE Access*, vol. 11, pp. 122069–122115, 2023. doi: 10.1109/ACCESS.2023.3328248.

[31] R. R. Mostafa, M. A. Gaheen, M. A. E. Abd ElAziz, M. A. Al-Betar, and A. A. Ewees, "An improved gorilla troops optimizer for global optimization problems and feature selection," *Knowl.-Based Syst.*, vol. 269, no. 5, Jun. 2023, Art. no. 110462. doi: 10.1016/j.knosys.2023.110462.

[32] F. Zitouni, S. Harous, A. Belkeram, and L. E. B. Hammou, "The archerfish hunting optimizer: A novel metaheuristic algorithm for global optimization," *Arab. J. Sci. Eng.*, vol. 47, no. 2, pp. 2513–2553, 2022. doi: 10.1007/s13369-021-06208-z.

[33] M. Majd and R. Safabakhsh, "Correlational convolutional LSTM for human action recognition," *Neurocomputing*, vol. 396, pp. 224–225, Jul. 2020.

[34] M. G. Ragab, S. J. Abdulkadir, and N. Aziz, "Random search one dimensional CNN for human activity recognition," in *2020 Int. Conf. Comput. Intell. (ICCI)*, Bandar Seri Iskandar, Malaysia, 2020, pp. 86–91.

[35] T. R. Mim *et al.*, "GRU-INC: An inception-attention based approach using GRU for human activity recognition," *Expert Syst. Appl.*, vol. 216, Apr. 2023, Art. no. 119419. doi: 10.1016/j.eswa.2022.119419.

[36] N. Dua, S. N. Singh, and V. B. Semwal, "Multi-input CNN-GRU based human activity recognition using wearable sensors," *Computing*, vol. 103, no. 3, pp. 1–18, 2021.