

# Human Activity Recognition Using Optimized Deep Learning with Data Augmentation

Subna M P<sup>1\*</sup>, Dr. V. Kathiresan<sup>2</sup>

<sup>1\*</sup>Research Scholar (Full Time), Department of Computer Science, Park's College (Autonomous), Chinnakkarai, Tirupur-641605.

<sup>2</sup>Principal, A.V. P. College of Arts and Science, Tirupur.

Email: rehnasubi@gmail.com, vkathirphd@gmail.com

**Abstract:** Human Activity Recognition (HAR) involves classifying human movements and has become essential for assessing the frequency and length of different human acts. The development of intelligent assistive devices and the examination of manual operations both depend on this. HAR has recently made use of deep neural networks, especially when it comes to day-to-day tasks, utilizing multichannel time-series obtained from body-worn devices equipped with various sensors. To begin, the WISDM and MHEALTH datasets are utilized as input, and a Generative Adversarial Network (GAN) is employed to learn a generative model that produces time-series data exhibiting similar space and time dependencies as real data. Subsequently, an Optimized Long Short-Term Memory Neural Network with Improved Aquila Optimizer (IAO-ILSTM-NN) is applied to perform HAR. The random parameters impact on prediction accuracy is addressed by optimizing the ILSTM-NN structure parameters using the IAO. Based on the simulation outcomes, the recommended framework executes superior than alternative frameworks that use the same datasets and baseline models. This underscores the effectiveness of the model in enhancing human activity recognition, particularly on multimodal sensing devices.

**Keywords:** Generative Adversarial Network (GAN), (DA) Data Augmentation, Human Activity Recognition (HAR), Long Short-Term Memory Neural Network (LSTM-NN), Improved Aquila Optimizer (IAO) and Wearable devices.

---

## 1. Introduction

Identifying certain human behaviors and motions and then responding appropriately is named as Human Activity Recognition (HAR). To begin, sensors that can record gestures, movements, and other characteristics corresponding with human movement are used [1]. Following the translation of these motion signals into HAR code using the data gathered, computers can comprehend and carry out the associated orders. Internet of Things (IoT), cloud computing (CC), and edge computing are examples of networked sensing skills that have made substantial strides in the last decade [2]. In IoT applications, wearable sensors play a crucial role in swiftly recording diverse body movements to facilitate HAR. Wearable inertial measurement unit (IMU) sensors, composed of accelerometers and gyroscopes, have undergone rapid development, simplifying the process of detecting and tracking human movement [3].

This advancement has enabled the application of HAR across diverse industries, encompassing healthcare, biometrics, and other areas like human emotion recognition [4]. Its adaptability highlights the significance of HAR using wearable sensors, extending beyond exercise-related activities to encompass a variety of everyday activities, including eating, drinking, cleaning one's teeth, and recognizing anomalies in sleep, are categorized and recorded. The evolution of HAR research has progressed from traditional handcrafted feature-based methods to the utilization of sophisticated deep learning (DL) techniques [5]. Initially, HAR methods mostly relied on machine learning (ML) and manual feature extraction.

Low-level features were captured using handcrafted feature extraction approaches such as SURF, SIFT, HOG, and others [6]. To overcome the limitations of activity identification, statistical learning techniques have been widely used [7], [8]. Jumping, running, and walking were among the seven movements that were identified in research by Chavarriaga et al. [9] using Naïve Bayes (NB) and K-Nearest Neighbour (KNN) classifiers. They were unable to find discriminative characteristics for the accurate distinction of diverse activities, however, and their method depended on handcrafted features. When it comes to recognizing human activity, feature extraction techniques like transform coding [12], statistics of raw data [11], and symbolic representation [10] are often used. However, these techniques are heuristic and need specialized understanding of feature design [13].

Recently, the integration of DL (Deep Learning) technology into HAR applications has gained popularity. Unlike traditional statistical (ML) Machine Learning methods, DL offers increased ease in extracting and categorizing intricate data, especially when dealing with numerous sensor sources. For instance, Convolutional Neural Network (CNN) [14] not only automatically extracts features but also comprehensively learns complex high-dimensional nonlinear patterns [15]. Wearable sensor-based DL for HAR has been investigated by several researchers. However, much of the existing research has treated deep learning as a black box, with limited exploration into the underlying data.

In a recent study [16], for HAR, a multi-channel CNN with DA was presented; it was called AMC.-CNN. The methodology involves constructing the feature window using sliding windows in time series, followed by augmenting the feature window through data transformation and addition. However, in the realm of time series recognition, many datasets tend to be quite small, and any methodology to resolve this limitation is through the use of data augmentation with Generative Adversarial Network (GAN) methods. Recognizing the importance of hyperparameter tuning in enhancing the accuracy and efficiency of neural networks, this research proposes the IAO-ILSTM-NN model for HAR. The primary contribution of this work is outlined as follows:

- An optimized DL framework is introduced for multiple HAR, incorporating a combined LSTM-NN with IAO for feature extraction and leveraging temporal dependencies.
- This model proves to be adaptable across various sensor modalities, effectively identifying multiple human activities regardless of the variations in body movements.
- Through experimentation with publicly accessible datasets like WISDM and MHEALTH, when compared to other DL frameworks, the suggested model performs better, as evidenced by published results and comparisons with baseline deep learning models.

This study is segmented into the sections listed as follows: An outline of the relevant DL backdrop for DAR is provided in Section II. Section III provides an in-depth depiction of the proposed IAO-ILSTM-NN model's structure. The experiment setup is outlined, and section IV goes into more detail about the findings and analysis. Section V provides specifics on conclusions and possible further research.

## **2. Related Work**

Helmi et al. [17] created a robust HAR system using data from wearable sensors by fusing applications of swarm intelligence (SI) with DL. A residual CNN and a recurrent NN (RCNN-BiGRU) were used in the development of their lightweight feature extraction approach. Novel feature selection strategies were developed based on the marine predator algorithm (MPA) to get the best possible feature set. Three binary versions, MPA\_S10 and MPA\_V, were created for this purpose in addition to the standard MPA version. To guarantee the suggested MPA variations' high performance, thorough testing was done on them by comparing them with a range of optimization methods and using a variety of assessment metrics and statistical tests. When working with time-series data, difficulties have been faced despite the adoption of DL-based approaches in activity detection.

Athota & Sumathi [18] introduced Hybrid Learning Algorithms (HLA) as a means to construct an inclusive classification technique for human activity recognition (HAR) utilizing data from wearable sensors. The objective of this study is to employ the Convolution Memory Fusion Algorithm (CMFA) and Convolution Gated Fusion Algorithm (CGFA). The purpose of these techniques is to allow the model to learn gated terms in sequential data, as well as long-term relationships and local properties. By using a range of filter sizes, different local temporal dependencies are captured and the improvement is applied appropriately, improving feature extraction. After applying the Amalgam Learning Model to the WISDM dataset, accomplishment rates for CMFA and CGFA implementations on smartwatches and smartphones were 96.76% and 94.98%, respectively, and 96.91% and 84.35%, respectively. Neural networks with built-in self-attention mechanisms may be used to identify patterns in raw sensor data inputs.

Varshney et al., [19] Initially, data from accelerometers, magnetometers, and gyroscopes placed at the body's ankle are jointly learned to form the foundation for the lower arm. When the accelerometer, gyroscope, and magnetometer data are integrated, the results are better than when the data are taken from each sensor alone or in other combinations. The research makes use of 2 widely accessible datasets, mHEALTH and PAMAP2. The suggested framework's higher performance is shown by the experimental findings, which are benchmarked against the most advanced techniques. For balanced and unbalanced mHEALTH datasets, the model obtains test accuracy values of 98.48% and 98.63%, respectively. For balanced and unbalanced PAMAP2 datasets, the model gets test accuracy values of 92.00% and 94.19%.

Khelalef et al., [20] used Body Sensor Network (BSTN) data to extract features and then categorize activities using a convolutional neural network (CNN). We used three publicly accessible datasets (Weizmann, Keck Gesture, and KTH Database) to do several experiments to evaluate the efficacy of our technique. Based on the experimental outcomes, the methodology overtakes contemporary DL techniques and overtakes traditional state-of-the-art (SOTA) approaches in terms of recognition accuracy. Notably, our method is straightforward to implement, demands less computational power, and is applicable to multi-subject activity recognition. The imperative for HAR becomes apparent in the framework of constructing smart homes and intelligent healthcare environments, acknowledging the challenges inherent in HAR due to the complexity and heterogeneity of the sensors involved in its recognition.

Hassan et al., [21] approached the challenge of human activity recognition by framing it using information from wearable body sensors as a classification problem. For enhancing the precision of HAR, we specifically support the implementation of a Deep Belief Network (DBN) model. First, using the unprocessed body sensor data, we extract important properties. Then, to further improve these characteristics and increase their resilience for quick activity identification, we use linear discriminant analysis (LDA) and kernel principal component analysis (KPCA). Once these processed characteristics are used, the DBN is trained. To confirm the deep learning algorithm's effectiveness, many tests were run on a real-world dataset derived from wearable sensors. In activity recognition, the findings show that the suggested DBN performs well, outperforming competing algorithms. Nevertheless, since human emotions are so complex and varied, it is still difficult to reliably and automatically identify physical human activities using wearable sensors.

Kolkar & Geetha [22] used datasets such as UCI HAR, WISDM, KTH action, and PAMAP2 to train a deep neural network (DNN) and evaluate the suggested system using Spider Monkey Optimization (SMO). Walking, standing, lying down, running, climbing stairs, and descending stairs are among the actions covered by the datasets. The recurrent neural network's hidden layer is where the spider monkey model's fitness function is started in this method to improve precision and accuracy. In comparison to other cutting-edge techniques like DL-Q, End-to-end DNN, and SVM, experimental findings show performance increases. With a 98.92% accuracy, 98.12% precision, 98.9% recall, and 95.90% F1-score for the WISDM dataset, the suggested SMO-based approach is seen to perform better across many evaluations and tests. Comparing this execution to other SOTA methods, similar improvements of 2.8% in error rate are seen for different datasets. Given the challenge of collecting long-term interdependence and deriving effective characteristics from raw sensor data, identifying complex human behaviors is proving to be tough.

Choudhury & Soni [23] presented a simplified and resource-efficient hybrid deep learning model that uses sophisticated feature learning algorithms and physiological electromyography (EMG) sensors to recognize complex human actions. Multiple 1-D convolution layers are integrated into the proposed CNN-LSTM framework for the extraction of spatial features. To find long-term temporal relationships, the created feature maps are then fed into recurrent layers. Through training and testing with an unprocessed raw EMG dataset from physiological sensors, the suggested framework attained 84.12% accuracy at its peak and 83% on average. Our approach avoids the heavy data preprocessing and feature engineering overhead common in real-time activity recognition and instead prioritizes performance optimization, in contrast to much of the existing work on Human Activity Recognition (HAR), which frequently relies on heavily augmented and pre-processed data.

Chen et al., [24] suggested a semisupervised DL model which uses multimodal wearable sensory data to identify unbalanced activities. Our objective is to concurrently address limited labeled data, class-imbalance problems, and obstacles related to multimodal sensor data, such as interpersonal variability and interclass similarity. More specifically, we provide a semi-supervised framework that is pattern-balanced and capable of extracting and preserving a variety of latent patterns of activity. Furthermore, using our recurrent convolutional attention networks, we detect prominent areas suggestive of human activities by leveraging the independence of sensory input modalities. The results of our experiments demonstrate that, even with a mere 10% of labeled training data, our suggested model outperforms several cutting-edge techniques, including semi-supervised and supervised methods. Our method's resilience in handling tiny and unbalanced training datasets is further shown by its results. It's important to remember, however, that the importance of joints changes as the actions in a video go on and as they do differently.

Nikpour & Armanfard [25] Introduced is a method for spatial hard attention identification designed to eliminate uninformative or misleading joints in each frame. The spatial-attention-aware agent is trained via deep reinforcement learning, and the joint selection issue is formulated as a Markov decision process. Interestingly, agent training doesn't need any extra labelling. The agent generates a sequence of joint probabilities as output after processing a series of characteristics that are taken from skeletal videos. To improve performance, this technique may be easily combined with other skeleton-based activity identification methods. It functions as a

flexible framework. The conclusions show intense competition in activity identification amongst three popular datasets for HAR.

Ahmad et al., [26] recommended a technique for recognizing human activity that combines the use of a Bidirectional-Gated Recurrent Unit (Bi-GRU) and CNN to interpret visual input. First, we use CNN for deep feature extraction from frame sequences in videos of people doing things. Next, to improve performance and lower the model's computational complexity, we extract the most important characteristics from these deep representations. The second phase involves the introduction of Bi-GRU to capture the temporal dynamics of frame sequences. Bi-GRU learns both forward and backward temporal dynamics directions at every time step and is fed the deep-important characteristics that were derived from the frame sequence of human actions. Realistic films from datasets for human activity identification, such as YouTube11, HMDB51, and UCF101, are used in extensive tests. Lastly, we show the efficacy of our suggested procedure by comparing the findings with those from other approaches. It is important to keep in mind, nevertheless, that these current methods only record the local contents of human activities, that makes them appropriate for basic activity identification but less successful in situations when several people are engaged in different activities.

**Inference:** Each deep learning (DL) model possesses unique learning process configurations to assimilate data and enhance its execution. These configurations are linked to the hyperparameters within DL frameworks, significantly influencing training time, computational costs, and overall model performance. The primary challenge lies in selecting the optimal set of hyperparameters, as each hyperparameter exerts a distinct influence on the DL framework. A common technique to address this challenge is the trial-and-error methodology, wherein hyperparameters are chosen through empirical means. However, in DL architectures, there exists a trade-off among hyperparameters, meaning that adjusting one may impact others. Given these considerations, manually searching for the optimal set of hyperparameters can be a laborious task.

### 3. Proposed Methodology

In recent times, use of multichannel time-series data has allowed deep neural networks to be used to HAR in the framework of everyday actions. The wearable devices that are located on the body and have various kinds of sensors are the sources of these time series. Initially, the WISDM and MHEALTH datasets are utilized as inputs, and to train a generative model that can replicate the spatial and temporal correlations seen in actual data, a Generative Adversarial Network (GAN) is employed. Then, for HAR, a LSTM-NN is employed. To mitigate the effect of random parameters on prediction accuracy, the LSTM-NN's structural parameters are optimized using the Improved Aquila Optimizer (IAO).

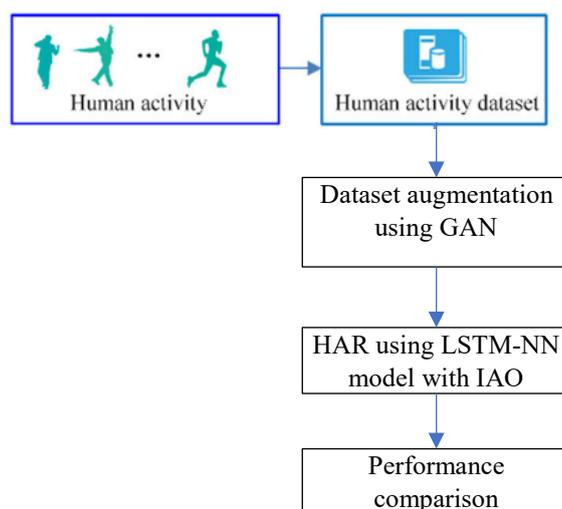


Figure 1. Proposed HAR model using LSTM-NN model with IAO

**Input dataset Description:** In this section, we present a novel LSTM-NN model enhanced by the Improved Aquila Optimizer (IAO) for the classification of Human Activity Recognition (HAR). This approach centers around the LSTM-NN model, which contains crucial hyperparameters significantly influencing its performance.

The IAO method is employed to optimize these hyperparameters, ensuring highly accurate predictions for various HAR activities across two datasets, namely WISDM and MHealth.

The WISDM dataset [34] includes sensor data from accelerometers in phones. The dataset, which is comprised of 1,098,209 records that capture 3-axis accelerometer sensor readings, was gathered via an application that was loaded on each user's phone. These readings originate from 29 users who carried a smartphone positioned in their front pants' pocket. Notably, the age, gender, and physical/behavioral features of the users are not included in the dataset. A rate of 20 Hz per second was used to sample the data. Figure 5 depicts the class distribution, with jogging and walking being the most prevalent activities.

The MHealth dataset [2] comprises sensor information obtained from ten individuals participating in twelve different activities. An electrocardiogram sensor, a 3-axis magnetometer, a 3-axis gyroscope, and a 3-axis accelerometer are the three devices from which the data is derived. These sensors are positioned at various body locations, including the chest, hand, and ankle. Similar to the WISDM dataset, the MHealth dataset lacks individual data of the users. There are 1,215,745 occurrences in all, and the distribution of classes across the different activities is well-balanced. Notably, the activity with the least amount of data is "Jump front & Back".

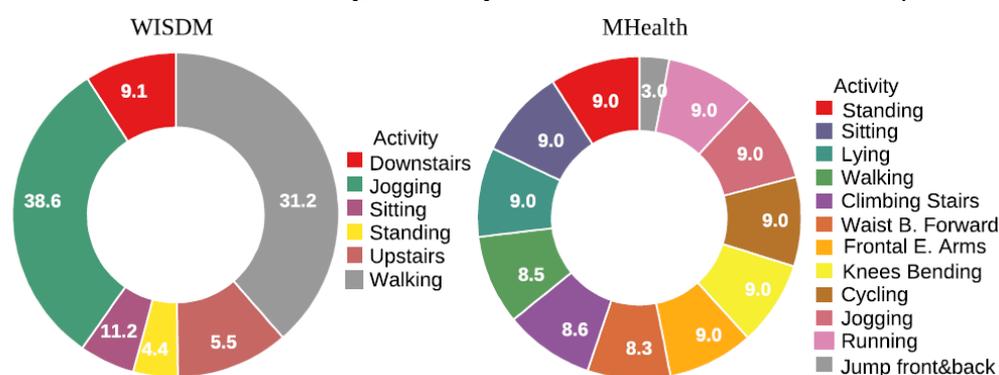


Fig. 2. Distribution of the Classes for each dataset [27].

### Construction of Feature Window

This research utilized the WISDM and MHEALTH datasets, focusing primarily on HAR for wearable devices. An accelerometer, gyroscope, magnetometer, or heart rate meter is one of the time-series data sources that primarily provide HAR data. Collecting human activity features from time-series data using sliding windows is the process of creating a feature window. The generated feature window should capture the properties unique to a given activity as time series features may cover a variety of human activities. As a result, the feature data in the time series is labeled before the feature window is constructed. This makes it possible for the sliding window to move within the independent time series segment in response to a certain coverage ratio. Sensor data are non-image data; examples of these are magnetometers, accelerometers, gyroscopes, and heart rate monitors, possessing a relatively simple structure typically with a few dimensions. An accelerometer, for example, provides data in the format (x, y, z), which is a sequence of three-dimensional coordinates at various places. Figure 3 shows the steps involved in building the feature window.

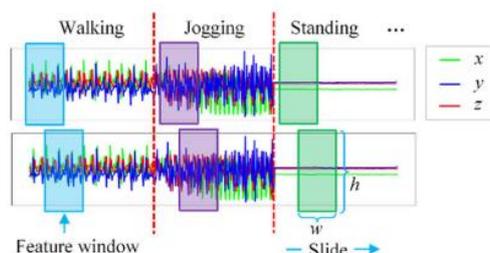


Fig.3. Sliding window-based feature construction [16]

Given a sampling rate of "f," "w" in the time series of acquired data indicates the amount of data points gathered in "t" seconds, and "h" indicates the cumulative dimensions of information obtained by every sensor. A certain coverage ratio determines how the window glides. To preserve the feature window's independence, a new sliding

window opens when a new sequence action begins. The basic mapping relationship between features and activity categorization is established by the creation of a feature window. Here is a summary of the matching formula:

$$w = f \times t \tag{1}$$

$$h = \sum_{i=1}^n \dim(S_i) \tag{2}$$

$$\mathit{fw} = \begin{bmatrix} S_{11} & S_{21} & \dots & S_{w1} \\ S_{12} & S_{22} & \dots & S_{w2} \\ \dots & \dots & \ddots & \dots \\ S_{1h} & S_{2h} & \dots & S_{wh} \end{bmatrix} \tag{3}$$

In the given expression, the data dimension of the  $i$ th sensor can be denoted as  $\dim(S_i)$ , the feature window can be represented as  $fw$ , and  $S_{wh}$  corresponds to the data of dimension  $h$  sampled at the  $w$ th time within the  $fw$ .

### Data Augmentation for Feature Window

In Computer Vision (CV), DA aims to incorporate prior knowledge for optimizing model performance through straightforward and efficient solutions. By employing data augmentation, the network model learns specific data transformation methods, enhancing its overall performance [16]. Meaningful data alterations or additions may be made to the feature window that is retrieved from wearable devices and represents multidimensional user activity characteristics from numerous sensors across time. Because of these changes, the original feature window's distribution space is widened and activity characteristics are highlighted. Consequently, the model can gather more feature information with ease. Thus, enhancing the model's capacity for generalization and maintaining a strong correlation are guaranteed when data augmentation is applied to the feature window.

Within a predetermined window, the feature window collects sensor sample data, capturing distinct characteristics of human activity. For instance, distinguishing jogging from walking reveals more pronounced changes in data from the  $x$ ,  $y$ , and  $z$  axes. Furthermore, it also reveals key differences when a detailed examination of the  $x$ -axis,  $y$ -axis, and  $z$ -axis is done. Metrics like average value, standard deviation, and feature dimension comparisons are used to improve the feature window's representation capacity. A simplified technique for enhancing data is described in this paper, and the enhanced feature window may be computed using the following formula:

$$avg_i = \sum_{k=1}^w S_{ki} / w, 1 \leq i \leq h \tag{4}$$

$$std_i = \sqrt{\frac{\sum_{k=1}^w (S_{ki} - avg_i)^2}{w}}, 1 \leq i \leq h \tag{5}$$

The variables  $avg_i$  and  $std_i$  in the provided equations stand for the average value and standard deviation of the  $i$ th row, respectively, in the  $fw$ . Throughout the given time,  $avg_i$  represents the mean value of the sensor feature dimension, and  $std_i$  denotes the degree of variance in that dimension. The average deviation among feature dimensions'  $x$  and  $y$  for a particular sensor is stated in Eq. 6, assuming the  $fw$  has 3 feature dimensions,  $x$ ,  $y$ , and  $z$ . Similar calculations may be made for  $\mathit{delta}_{xz}$  and  $\mathit{delta}_{yz}$  for the same sensor. By using equations 4, 5, and 6, equation 7 defines the enhanced portion of the  $fw$ . Interestingly, the feature dimension sizes of each sensor may vary; for this reason, the default value of the feature dimension's average deviation in Eq. 7 is 0. The augmented  $fw$  is generated by transposing  $fw$  to represent  $fw^T$ , as described in Eq. 8, through addition of the DA to the original  $fw$ . Vertical connections between the transposed  $fw$  and transposed augmented data are essentially what create the augmented feature window  $afw$ .

$$\mathit{delta}_{xy} = \frac{\sum_{i=1}^w |x_i - y_i|}{w} \tag{6}$$

$$\mathit{agt} = \begin{bmatrix} avg_1 & std_1 & \mathit{delta}_{xy} \\ avg_2 & std_2 & \mathit{delta}_{xy} \\ \dots & \dots & \dots \\ avg_h & std_h & \mathit{delta}_{yz} \end{bmatrix} \tag{7}$$

$$\mathit{afw} = \begin{bmatrix} fw^T \\ \mathit{agt}^T \end{bmatrix} \tag{8}$$

Improved data sample quality for HAR, reduced redundancy in the dataset, and increased model capacity to learn more varied human activity aspects are the objectives of this study. The research suggests connecting many augmented feature windows that correspond to the same activity horizontally to accomplish this. A reconstructed feature sample is produced when three enhanced feature windows are connected horizontally [16].

**Human Activity Recognition Using IAO-ILSTM-NN**

In this section, the proposed IAO-based ILSTM-NN is used for performing human activity classification and prediction using features obtained from data augmentation phase.

**Improved Long Short-Term Memory (LSTM) Neural Network:** An improvement over RNN is LSTM [28], which adds more interactions per module (or cell) to get over some of the drawbacks of RNN. Long-term dependencies may be understood by LSTMs, a particular kind of RNN, which can also remember information for a long time. The LSTM model [29] has a chain structure in its architectural design. The repeating module, however, has a unique setup. It contains 4 interacting layers using a different communication mechanism as opposed to a single neural network as in a typical RNN. Figure 4 shows how the LSTM neural network is organized.

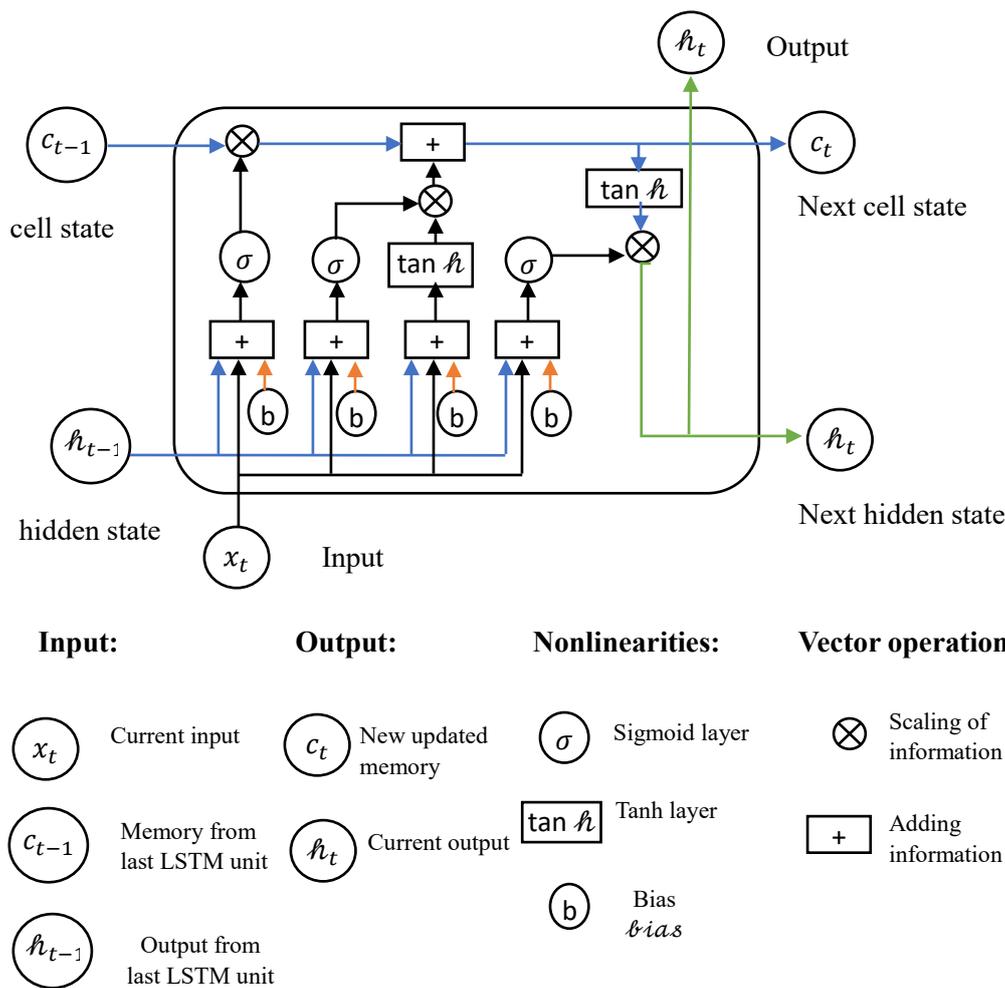


Fig.4. Structure of LSTM model

Two states the cell state and the hidden state are transferred from one memory block, or cell, to the next in an ordinary LSTM network. Data may flow forward with little changes, while certain linear transformations could take place, as the cell state is the main conduit for data flow. A series of matrix operations with unique individual weights, or a layer, is how sigmoid gates control the addition or removal of data from the cell state. LSTMs are engineered to tackle long-term dependency issues by utilizing gates to control the memorization process. Constructing an LSTM network begins by identifying unnecessary information to be excluded from the cell at that step. The output of the previous LSTM unit ( $h_{t-1}$ ) at time t-1 and the current input  $x_t$  at time t are taken into account by the sigmoid function, which controls this selection process. The forget gate, also known as  $f_t$ , is the gate that the sigmoid function uses to identify the part of the previous output that should be deleted. Each element in the cell state,  $C_{t-1}$  is represented by a vector called  $f_t$ , whose values range from 0 to 1.

$$f_t = \sigma(\mathcal{W}_f[h_{t-1}, X_t] + bias_f) \tag{9}$$

In this context,  $\sigma$  represents the sigmoid function, while  $\mathcal{W}_f$  and  $bias_f$  denote the bias and weight matrices connected to the forget gate, respectively. The next step is to determine and update the cell state by adding the data from the incoming input ( $X_t$ ). This process consists of two components: the first being the sigmoid layer, which evaluates the new information can be retained or disregarded (assigned values of 0 or 1), and the second being the  $\tanh$  layer. By allocating weights to the given numbers, the  $\tanh$  function establishes their significance level between -1 and 1. This updated memory is then combined with the current memory  $C_{t-1}$  to create the updated memory  $C_t$ . The cell state is modified by the product of these two values.

$$i_t = \sigma(\mathcal{W}_i[h_{t-1}, X_t] + bias_i), \tag{10}$$

$$N_t = \tanh(\mathcal{W}_n[h_{t-1}, X_t] + bias_n), \tag{11}$$

$$C_t = C_{t-1}f_t + N_t i_t. \tag{12}$$

$W$  and  $bias$  indicate the weight matrices and bias connected to the cell state, respectively, in this context, whereas  $C_{t-1}$  and  $C_t$  stand for the cell states at times  $t-1$  and  $t$ . The output values ( $h_t$ ) are eventually generated in a refined form from the output cell state ( $O_t$ ). First, the components of the cell state that contribute to the output are identified via a sigmoid layer. The freshly produced values from the  $\tanh$  layer applied to the cell state ( $C_t$ ) are then increased by the output of the sigmoid gate ( $O_t$ ), where these values fall within the range of -1 to 1.

$$O_t = \sigma(\mathcal{W}_o[h_{t-1}, X_t] + bias_o), \tag{13}$$

$$h_t = O_t \tanh(C_t). \tag{14}$$

In this instance, the output gate's weight matrices and bias are denoted by  $\mathcal{W}_o$  and  $bias_o$ , respectively. To reduce prediction oscillation, cell propagation incorporates a hyperbolic tangent activation function that keeps values between (-1, 1). With this alteration, oscillations are no longer visible, and the cell divergence phenomena seen in the normal model are largely eliminated. Furthermore, the hyperbolic tangent formula is recognized as the hyperbolic function and is defined as follows:

$$\tanh(x_t) = \frac{\sin h(x_t)}{\cos h(x_t)} \tag{15}$$

In order to diminish inherent simplicity and accommodate the growing volume of information, the LSTM's cell state dimension was decoupled from the input size, enabling the flexibility to expand state sizes and retain additional temporal information. To establish this independence in cell dimensions, the output gating was followed by the introduction of a single-layer NN with a Rectified Linear Unit (ReLU) activation function. The utilization of this activation function effectively removed negative outputs, as illustrated in Figure 5.

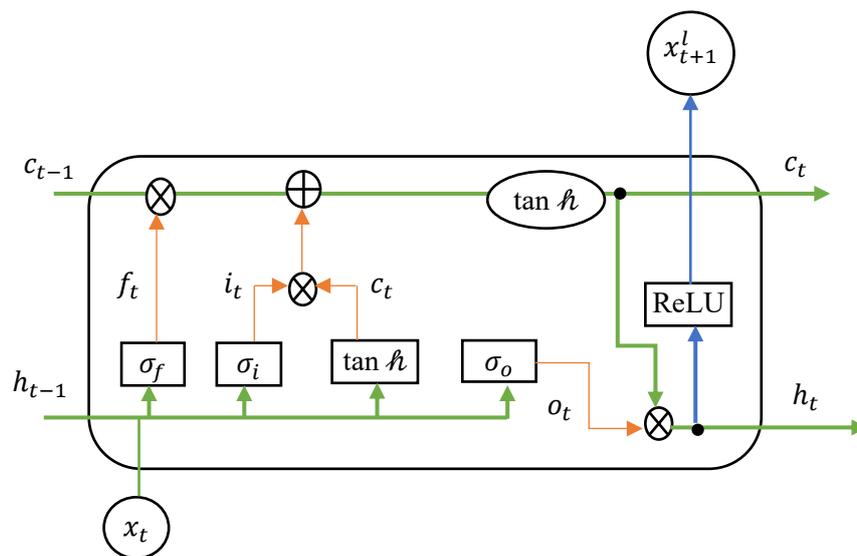


Fig.4. Structure of Improved LSTM model

In order to diminish the simplistic nature and accommodate the growing influx of information, the input size had no effect on the LSTM's cell state dimension. This independence facilitated the expansion of state sizes, enabling the retention of more temporal information. These adjustments markedly enhanced prediction accuracy, leading to the adoption of this structure as a foundational model, labeled ILSTM. Effective training on continuous data was hampered by the hyperbolic tangent's inclusion, which brought up the vanishing gradient issue once again. Discrete blocks of feature data were used for model training in a hybrid backpropagation technique to modify the issue. Simultaneously, without requiring a lengthy backpropagation, some long-term trend learning was made possible by the initialization of the cell and concealed states from the previous iteration.

**Hyperparameters Tuning Using IAO Algorithm:** In this study, the IAO algorithm is employed for tuning hyperparameters in the ILSTM-NN. The AO parameters and hyperparameters of the LSTM network, depending on the findings shown in Fig. 5, variables like the batch size, time step, and amount of hidden layers are initialized. The introduction of IAO, inspired by modifications to the SCF from IAO in a prior study [30], led to further adjustments to the AO. Nevertheless, it was discovered that the SCF convergence features slowed down the precision of epochs in IAO, which could make it more difficult to get the best result. A revised search control factor (RSCF) that was specially designed for the second and third search procedures was included in a modified version of IAO that was released in response to these difficulties. The next section presents a thorough synopsis of the IAO methodology, emphasizing certain adjustments and how they affect the optimization process. Aquila's mobility in terms of epochs is minimized by using the ISCF to control the search range. As such, the search space is far less than with the prior SCF, and the best solution is found considerably faster than with the previous method. The following is how the updated RSCF is displayed:

$$RSCF(t) = 2 \times \exp\left(1 - \left(\frac{t \times (t \times 0.1)}{\max T}\right)\right) \times dcf \quad (16)$$

$$dcf = \begin{cases} 1 & \text{if } ri < 0.5, \\ -1 & \text{else.} \end{cases} \quad (17)$$

Here, the variable  $t$  signifies the current iteration, and  $T$  represents the maximum iteration. A random number between 0 and 1 is represented by the parameter  $ri$ , while the direction control factor is indicated by the symbol  $dcf$ . The direction in which the Aquila fly is determined in large part by these variables. By restricting the Aquila's movement, the RSCF function lowers optimization latency and achieves fast convergence. Compared to the original AO, the improved method finds the optimum solution set in less time. A total of 250 epochs were used to run both optimization strategies. The approach that is being provided incorporates four unique phases of search, which are further discussed below. This is made possible by the inclusion of the RSCF function.

Step 1: Vertical Dive Attack ( $S_1$ ): The Aquila starts its hunting operation by using a high-altitude swoop to locate the target area and choose the optimal hunting location. The following is a description of these moves, sometimes referred to as vertical dive attacks:

$$S_1(t + 1) = S_{best}(t) \times \left(1 - \frac{t}{T}\right) + (S_M(t) - S_{best}(t) \times r) \quad (18)$$

In Equation (18),  $S_1(t + 1)$  represents the candidate solution for  $(t + 1)$  epochs, where  $S_{best}(t)$  is the best answer found up to the  $i$ th generation, and " $r$ " stands for a random number within the range  $[0, 1]$ . To control the search area, the word  $\left(1 - \frac{t}{T}\right)$  is used. The mean value of the present solution up to the  $i$ th epoch is also shown by  $S(t)$ .

Step 2: Updated Comprehensive Search with a Brief Glide Strike (RS): The Aquila thoroughly investigates the solution space at a variety of angles and speeds before attacking its target, conducting a thorough search characterized by shorter glide attacks, as illustrated below:

$$RS_2(t + 1) = S_R(t) + RSCF(t) \times (S_{best}(t) - S(t)) \times r \times (y - x) \quad (19)$$

In Equation (19), the locations or coordinates of the points that formed the spiral shape throughout the search phase are represented by the variables  $x$  and  $y$ . The search control factor is denoted by  $RSCF(t)$ , and the word  $r$  denotes a random integer in the range  $[0, 1]$ . To deal with the problem of becoming stuck in a locally optimum solution, we used  $RSCF$  in place of the Levy flight (LF) distribution.

Step 3: Modified Search Around Prey and Attack (MS): After the  $MS_2$  search stage, the prey's position is pinpointed exactly. The search around prey and attack is the term for the Aquila's thorough reconnaissance of the area around the target. It uses simulated assaults to ascertain the prey's reaction.

$$RS_3(i, j) = lb_j + r \times (ub_j - lb_j) + r(S_R(j) - S_{best}(j)) \times RSCF(t) \times \left(1 - \frac{t}{T}\right) \quad (20)$$

The random collection of solutions is represented by  $S_R(j)$  in Equation (20), while the current solution for  $t$  epochs is indicated by  $MS_3(i, j)$ .

Step 4: Walk and Grab Attack (S): Atlast, using the motions of the prey as a guide in the fourth search approach, the Aquila launches an assault from an elevated posture. The term "Walk and Grab Prey" describes this kind of search approach,

$$S_4(t + 1) = Q_F \times S_{best}(t) - (G_1 \times S(t) \times ri) - G_2 \times lev(D), \quad (21)$$

$$Q_F = t^{\frac{2 \times r - 1}{(1-T)^2}}, \quad (22)$$

$$G_1 = 2 \times ri - 1, \quad (23)$$

$$G_2 = 2 \times \left(1 - \frac{t}{T}\right). \quad (24)$$

In this context,  $S_4(t + 1)$  signifies the current solution achieved, and  $lev(D)$  represents the distribution of Levy across the  $D$ -dimensional space. The quality function (QF) in the search process helps maintain balance.  $G_1$  encompasses the various movements of the Aquila during the hunt, while  $G_2$  denotes the gradient of the hunting process. The selection of fitness is a pivotal aspect of the IAO method, wherein the encoder performance is utilized as a measure for a superior solution candidate. The performance value is now the primary criterion employed to construct *Fitness*.

$$Fitness = max(P) \quad (25)$$

$$P = \frac{TP}{TP + FP} \quad (26)$$

where the true and false positive values are denoted by  $TP$  and  $FP$ .

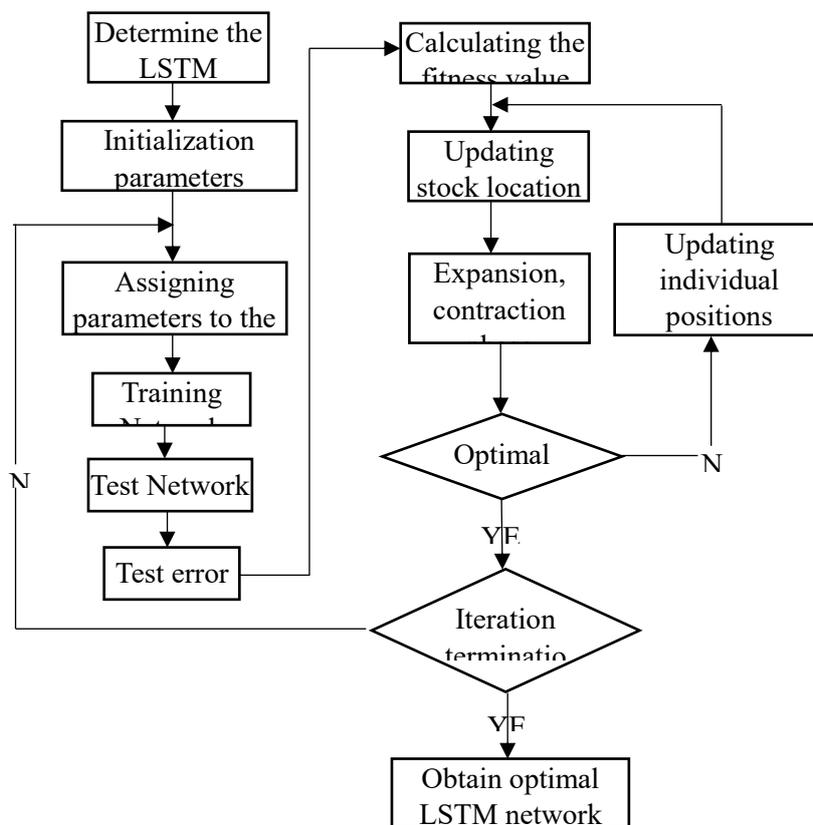


Fig.5. Flowchart of IAO based hyperparameter tuning of ILSTM-NN

### Experimental Results and Discussion

In this study, we presented a technique using a multi-channel CNN with feature window DA, with the objective of achieving precise Human Activity Recognition (HAR) through a relatively lightweight network structure. We

performed a comparison examination of the framework's recognition execution with and without augmentation to evaluate the efficacy of this strategy, using the WISDM and MHEALTH datasets. The proposed IAO-ILSTM-NN was compared against existing methods, including AMC-CNN [16], HLA [18], and SMO-DNN [22]. The experimental setup utilized a Windows 10 Professional 64-bit OS (Operating System), an Intel i9-9900k CPU, 32GB RAM, NVIDIA RTX 2080Ti GPU, and the PyTorch DL structure. Additionally, the evaluation metrics involved true negative value (TN) for accurately predicting the negative class, while false negative value (FN) and false positive (FP) represented misclassified samples. Using these parameters, the equations for accuracy, precision, F1-score, Mathew Correlation-coefficient (MCC), specificity, and sensitivity are expressed as follows: The classification model's accuracy indicates its overall performance and can be computed using the provided formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100 \tag{27}$$

**AUC (Area under the ROC Curve).** All possible categorization thresholds, AUC provides a thorough evaluation of performance. The chance that the model gives preference to a randomly selected positive example over a randomly selected negative example is one way to evaluate AUC calculations.

$$AUC = \frac{Recall+sensitivity}{2} * 100 \tag{28}$$

The F1-score serves as a weighted metric, considering both recall precision and sensitivity. Its scale spans from 0 to 1, with a value of 1 representing excellent performance by the classification algorithm, and a value of 0 signifying poor performance.

$$F1\ score = \frac{2*Precision*Recall}{Precision+Recall} \tag{29}$$

MCC, or Matthew Correlation Coefficient, is a correlation coefficient that assesses how well the expected and actual results correlate. The resulting MCC values fall within the range of -1 to +1. An MCC of -1 indicates a completely incorrect prediction by the classifier, 0 suggests a classifier making random predictions, and +1 signifies an ideal prediction by the classification models. The formula for computing MCC values is provided below:

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{30}$$

Specificity is described as the proportion of the total amount of persons to the ones who were accurately categorized as inactive. This suggests projecting a bad scenario while the subject is, in fact, passive. The following formula may be used to calculate specificity:

$$Specificity = \frac{TN}{TN+FP} * 100 \tag{31}$$

The ratio of newly identified activities to all heart disease cases is known as sensitivity. This indicates that the model predicts a positive outcome when the person is indeed active. The formula for computing sensitivity is provided below:

$$Sensitivity = \frac{TP}{TP+FN} * 100 \tag{32}$$

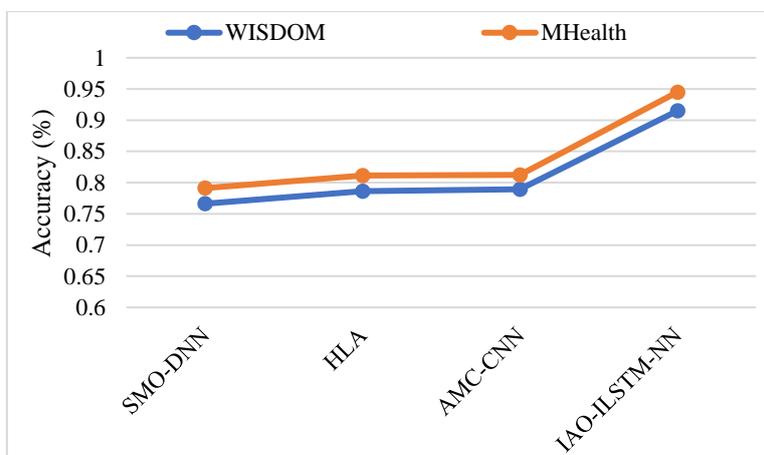


Fig.6. Comparison of the distribution of Accuracy validation values.

The suggested IAO-ILSTM-NN strategy was compared to three other machine-learning techniques to assess its efficacy: HLA, SMO-DNN, and AMC-CNN, along with their respective strategies. Figure 6 illustrates the corresponding accuracy distributions for these four approaches. Considering the average accuracy values, it is evident that IAO-ILSTM-NN surpasses the performance of the other approaches. Notably, the highest accuracy values for the proposed method were achieved through the TL-based strategy. Specifically, IAO-ILSTM-NN achieves an accuracy of 91.5% (WISDOM) and 94.5% (MHealth) compared to all other models. Results for Human Activity Recognition (HAR) might be improved by the higher quality produced by the suggested IAO-ILSTM-NN. In contrast, existing methods such as HLA, SMO-DNN, and AMC-CNN yield lower accuracy values. Consequently, the proposed algorithm demonstrates superiority over existing algorithms in terms of achieving robust validation results for activity prediction.

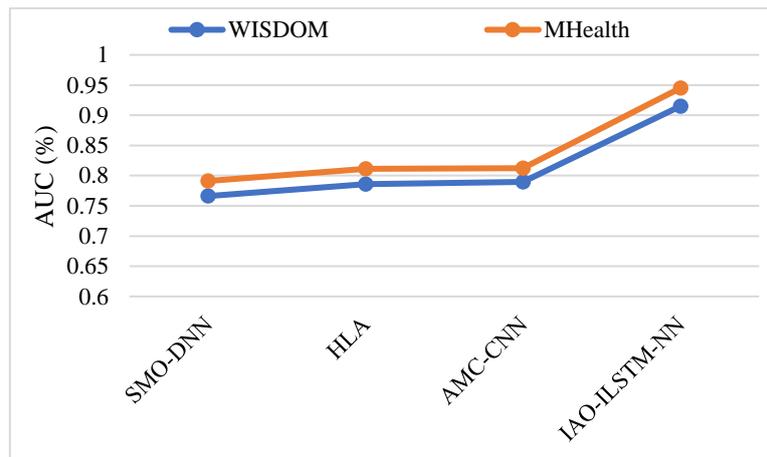


Fig.7. The distribution of AUC validation values

The proposed IAO-ILSTM-NN approach's distribution of 50 validation AUC values is contrasted with each of the three machine learning methodologies' distribution of 50 validation AUC values in Figure. 7 (SMO-DNN, HLA, and AMC-CNN). Notably, the TL-based technique produced the best results for the proposed method. Given the excellent quality derived from the IAO-ILSTM-NN, it has the potential to enhance disease identification. The IAO-ILSTM-NN achieves AUC values of 88% (WISDOM), 89% (MHealth), and 91.4% (Leukemia) in comparison to all other models. Results from established approaches include SMO-DNN (74.15% for WISDOM cancer, 73.15% for MHealth cancer, and 75.15% for leukemia), HLA (74.21% for WISDOM cancer, 72.2% for MHealth cancer, and 76.21% for leukemia), and AMC-CNN (81% for WISDOM cancer, 81% for MHealth cancer, and 82% for leukemia). Consequently, in terms of reliable validation results, the suggested approach outperforms current cancer prediction algorithms. Existing approaches fall short in providing accurate detection results in the presence of denoised features. The suggested method outperforms the other three in overcoming the impact of pre-processing features and accurately classifying Human Activity Recognition (HAR).

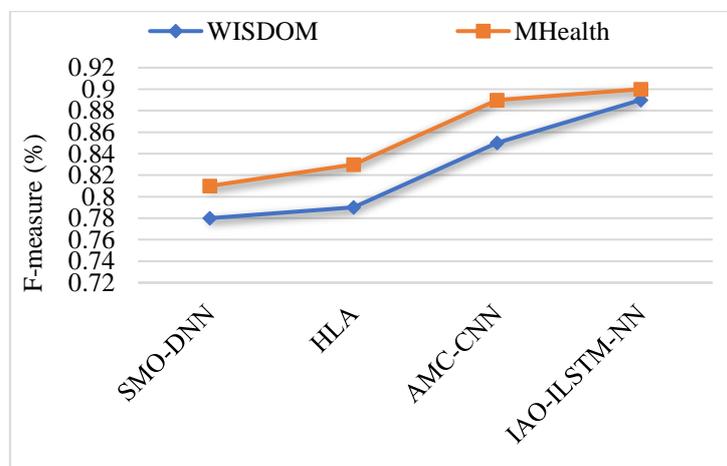


Fig.8. The distribution of F-Measure validation values

Figure 8 illustrates that the IAO-ILSTM-NN surpasses other approaches, including HLA, SMO-DNN, and AMC-CNN. Notably, the TL-based strategy yields the highest values for the proposed method. The IAO-ILSTM-NN achieves an f-measure of 89% (WISDOM), 90% (MHealth), and 91% (Leukemia) compared to all other models. IAO-ILSTM-NN helps improve illness diagnosis due to its high-quality result. HLA and SMO-DNN results, among other approaches now in use, and AMC-CNN, are (78% - WISDOM, and 81% - MHealth, 85%), (79% - WISDOM, and 83% - MHealth, 89%), and (85% - WISDOM, and 89% - MHealth, 90%), respectively. Thus, the proposed algorithm outperforms existing algorithms, yielding superior validation results for predicting cancer diseases. The AMC-CNN method is sensitive to weak edges and provides suboptimal HAR results. The HLA method tends to identify the brightest part in the inactivity region, resulting in inaccurate detection. Conversely, the proposed method, benefiting from deep features extracted from the pretrained IAO-ILSTM-NN, is less affected by noisy data and, with optimal hyperparameters, achieves desirable HAR results.

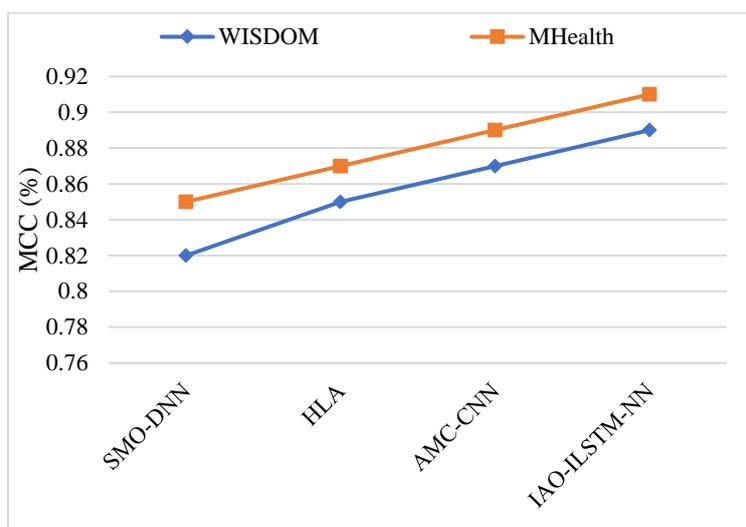


Fig.9. The distribution of MCC validation values

The four similar MCC distributions are shown in Figure 9. Taking into account the mean precision metrics, IAO-ILSTM-NN outperforms the other approaches. Notably, the TL-based strategy yields the highest values for the proposed method. The IAO-ILSTM-NN achieves an MCC of 89% (WISDOM), 91% (MHealth), and 92% (Leukemia) compared to all other frameworks. The outcomes of other approaches comparison, such as HLA, SMO-DNN, and AMC-CNN, the high quality produced by the suggested IAO-ILSTM-NN may improve illness identification. To forecast cancer illnesses, the suggested algorithm performs better than current algorithms in terms of validation findings. Through the analysis of experimental results, here is how well the current approaches perform. In comparison with the mentioned methods, the proposed approach can, to a certain extent, overcome the influence of activity and accurately extract activity boundaries through the efficient IAO method.

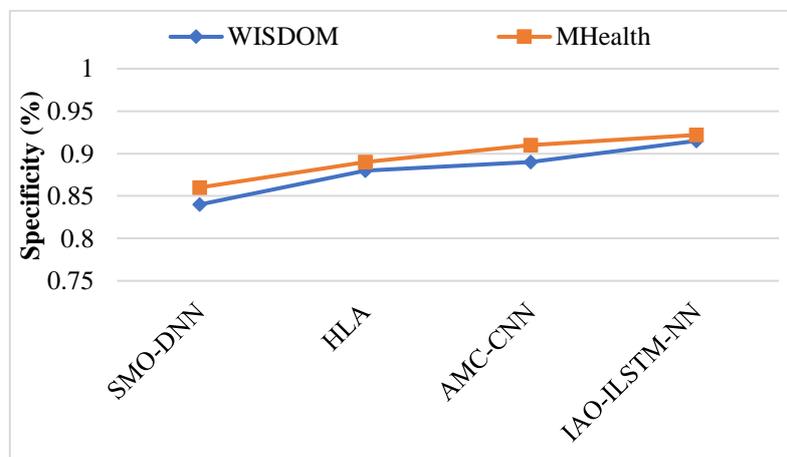


Fig.10. The distribution of Specificity validation values

In Figure 10, the graph illustrates that IAO-ILSTM-NN achieves specificity values of 91.5% (WISDOM), 92.2% (MHealth), and 92.5% (Leukemia) compared to all other models. The high quality generated by the proposed IAO-ILSTM-NN can significantly enhance disease detection. In contrast, existing methods such as SMO-DNN (84% - WISDOM, and 86% - MHealth), HLA (88% - WISDOM, and 89% - MHealth), and AMC-CNN (89% - WISDOM, 91% - and MHealth 92%) yield lower specificity values. Therefore, the proposed algorithm excels over existing algorithms in terms of achieving superior validation results for predicting human activity. The observed outcomes specify that the suggested technique has attained similar or better outcomes compared to other approaches, demonstrating its effectiveness with the IAO method in achieving accurate Human Activity Recognition (HAR).

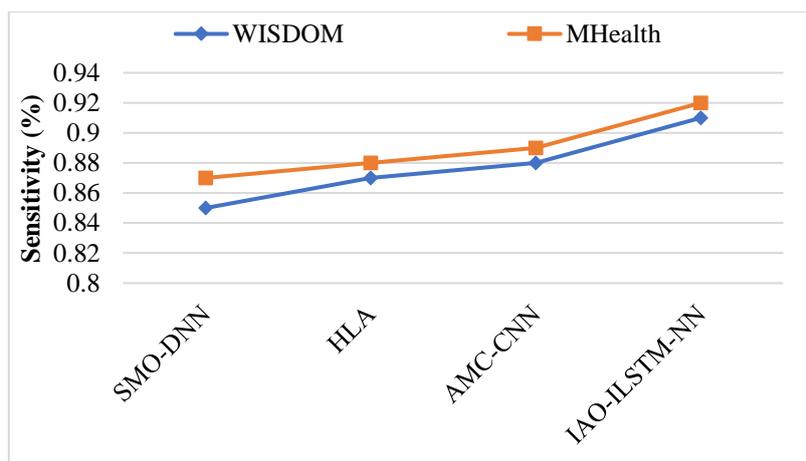


Fig.11. The distribution of Sensitivity validation values

In Figure 11, the graph illustrates that IAO-ILSTM-NN achieves an accuracy of 91% (WISDOM) and 92% (MHealth) compared to all other models. The high-quality outcomes produced by the proposed IAO-ILSTM-NN can significantly enhance Human Activity Recognition (HAR) classification. Consequently, the offered methods are surpassed by the recommended strategy in terms of superior validation results for human behavior prediction. Focusing on detection efficiency, this suggested approach takes into account the benefits of the hyperparameters selection module as well as the precision of detection targets.

#### 4. Conclusion and Future Work

The AMC-CNN model introduced in this study excels in capturing sample characteristics within time series data, optimizing the deep learning network structure, and enhancing Human Activity Recognition (HAR) capabilities. This approach has substantial application value in a variety of disciplines, including human-computer interaction, sports, and healthcare, by fusing deep learning with DA in sensor-based human activity perception. Data augmentation plays a crucial role in elevating the feature representation capacity of HAR, leading to heightened discrimination among various activities in time series and improving the overall data quality of samples. The comparatively lightweight design of the IAO-ILSTM-NN facilitates effective mapping between feature data and human behaviors, thereby augmenting the efficiency of HAR. The inclusion of LSTM-NN with cell regulation and ReLU units contributes to increased accuracy. In this study, the IAO system is employed for hyperparameter tuning, leading to improved detection results for ILSTM-NN. Through the use of the WISDM and MHEALTH datasets, experimental findings illustrate how well IAO-ILSTM-NN performs in recognizing human activity in single-sensor and multi-sensor settings. While the proposed technique focuses on sensor-based HAR, there remains a need for further research on lightweight real-time HAR methods depends on CV and WiFi. Future endeavors will involve applying IAO-ILSTM-NN to HAR in specific scenarios. We also want to expand the scope of our research to include more deep learning models and datasets, such as deep Boltzmann machines and DBNs.

#### 5. References

1. Krishnan, N. C., & Cook, D. J. (2014). Activity recognition on streaming sensor data. *Pervasive and mobile computing*, 10, 138-154.

2. Daga, Y., & Meena, S. (2022, December). Applications of Human Activity Recognition in Different Fields: A Review. In 2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) (pp. 1-6). IEEE.
3. Al Mamun, M. A., & Yuce, M. R. (2019). Sensors and systems for wearable environmental monitoring toward IoT-enabled applications: A review. *IEEE Sensors Journal*, 19(18), 7771-7788.
4. Yadav, S. K., Tiwari, K., Pandey, H. M., & Akbar, S. A. (2021). A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. *Knowledge-Based Systems*, 223, 106970.
5. Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z., & Liu, Y. (2021). Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)*, 54(4), 1-40.
6. Routray, S., Ray, A. K., & Mishra, C. (2017, February). Analysis of various image feature extraction methods against noisy image: SIFT, SURF and HOG. In 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT) (pp. 1-5). IEEE.
7. Z. Wang, D. Wu, R. Gravina, G. Fortino, Y. Jiang and K. Tang, "Kernel fusion based extreme learning machine for cross-location activity recognition", *Inf. Fusion*, vol. 37, pp. 1-9, Sep. 2017.
8. W. Liu, J. Yang, L. Wang, C. Wu and R. Zhang, "Movement behavior recognition based on statistical mobility sensing", *Adhoc Sensor Wireless Netw.*, vol. 25, no. 3, pp. 323-340, 2015.
9. R. Chavarriaga et al., "The opportunity challenge: A benchmark database for on-body sensor-based activity recognition", *Pattern Recognit. Lett.*, vol. 34, no. 15, pp. 2033-2042, 2013.
10. C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks", *Expert Syst. Appl.*, vol. 59, pp. 235-244, Oct. 2016.
11. M. Zeng et al., "Convolutional neural networks for human activity recognition using mobile sensors", *Proc. MobiCASE*, pp. 197-205, Nov. 2014.
12. O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors", *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192-1209, 3rd Quart. 2013.
13. Z. Chen, L. Zhang, Z. Cao and J. Guo, "Distilling the knowledge from handcrafted features for human activity recognition", *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4334-4342, Oct. 2018.
14. C. Szegedy et al., "Going deeper with convolutions", *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1-9, Jun. 2015.
15. Z. Chen, L. Zhang, Z. Cao and J. Guo, "Distilling the knowledge from handcrafted features for human activity recognition", *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4334-4342, Oct. 2018.
16. Shi, W., Fang, X., Yang, G., & Huang, J. (2022). Human activity recognition based on multichannel convolutional neural network with data augmentation. *IEEE Access*, 10, 76596-76606.
17. Helmi, A. M., Al-qaness, M. A., Dahou, A., & Abd Elaziz, M. (2023). Human activity recognition using marine predators algorithm with deep learning. *Future Generation Computer Systems*, 142, 340-350.
18. Athota, R. K., & Sumathi, D. (2022). Human activity recognition based on hybrid learning algorithm for wearable sensor data. *Measurement: Sensors*, 24, 100512.
19. Varshney, N., Bakariya, B., & Kushwaha, A. K. S. (2022). Human activity recognition using deep transfer learning of cross position sensor based on vertical distribution of data. *Multimedia Tools and Applications*, 81(16), 22307-22322.
20. Khelalef, A., Ababsa, F., & Benoudjit, N. (2019). An efficient human activity recognition technique based on deep learning. *Pattern Recognition and Image Analysis*, 29, 702-715.
21. Hassan, M. M., Huda, S., Uddin, M. Z., Almogren, A., & Alrubaian, M. (2018). Human activity recognition from body sensor data using deep learning. *Journal of medical systems*, 42, 1-8.
22. Kolkar, R., & Geetha, V. (2023). Human activity recognition using deep learning techniques with spider monkey optimization. *Multimedia Tools and Applications*, 1-18.
23. Choudhury, N. A., & Soni, B. (2023). Enhanced Complex Human Activity Recognition System: A Proficient Deep Learning Framework Exploiting Physiological Sensors and Feature Learning. *IEEE Sensors Letters*.
24. Chen, K., Yao, L., Zhang, D., Wang, X., Chang, X., & Nie, F. (2019). A semisupervised recurrent convolutional attention model for human activity recognition. *IEEE transactions on neural networks and learning systems*, 31(5), 1747-1756.
25. Nikpour, B., & Armanfard, N. (2023). Spatial Hard Attention Modeling via Deep Reinforcement Learning for Skeleton-Based Human Activity Recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.

26. Ahmad, T., Wu, J., Alwageed, H. S., Khan, F., Khan, J., & Lee, Y. (2023). Human Activity Recognition Based on Deep-Temporal Learning Using Convolution Neural Networks Features and Bidirectional Gated Recurrent Unit With Features Selection. *IEEE Access*, 11, 33148-33159.
27. Garcia, K. D., de Sá, C. R., Poel, M., Carvalho, T., Mendes-Moreira, J., Cardoso, J. M., ... & Kok, J. N. (2021). An ensemble of autonomous auto-encoders for human activity recognition. *Neurocomputing*, 439, 271-280.
28. Schmidt, R. M. (2019). Recurrent neural networks (rnns): A gentle introduction and overview. arXiv preprint arXiv:1912.05911.
29. Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM--a tutorial into long short-term memory recurrent neural networks. arXiv preprint arXiv:1909.09586.
30. Mumenin, K. M., Biswas, P., Khan, M. A. M., Alammery, A. S., & Nahid, A. A. (2023). A Modified Aquila-Based Optimized XGBoost Framework for Detecting Probable Seizure Status in Neonates. *Sensors*, 23(16), 7037.