Attention-driven BI-LSTM for Robust Human Activity Recognition and Classification in Disaster Scenarios

Pagare Shreyas*, Kumar Rakesh and Gupta S.K.
Department of Computer Science and Engineering, Rabindranath Tagore University, Bhopal (Madhya Pradesh), INDIA
*shreyas_au211443@aisectuniversity.ac.in

Abstract
Accurate and robust human activity recognition are essential for surveillance, healthcare and smart environment applications. However, the unpredictability and complexity of human motions provide significant challenges in obtaining the desired levels of accuracy and robustness. Conventional machine learning models such as Decision Tree, Gaussian NB and K Neighbors, have shown limited efficacy with accuracy estimates ranging from 78.3% to 89.3%. Cutting-edge techniques like Random Forest, RBF SVC and XGB Classifier achieve a maximum accuracy of 93.8%. We introduce a BI-LSTM model that focuses on the most important features using bi-directional long short-term memory networks with a special attention mechanism to address these limitations.

The present model demonstrates exceptional performance, attaining an accuracy of 99.83%, a precision of 99.46%, a recall of 99.75% and an F1 score of 99.85%, thereby surpassing other approaches by a substantial margin. The obtained findings validate the model's resilience and effectiveness in precisely recognizing and categorizing human actions in different fields and situations.

Keywords: Human Activity Recognition (HAR), Attention-Driven BI-LSTM, Machine Learning Models, Deep Learning, Classification Accuracy, Temporal Sequence Analysis.

Introduction

Research in the fields of healthcare, surveillance, intelligent settings and sports centers on human activity recognition. The primary method of mechanically identifying and classifying human actions based on data from sensors or video recordings is known as human activity recognition (HAR). Anomaly detection, security, fitness monitoring and elder care are just a few of the many potential uses for this technology. The potential for quick data collection from wearables, cellphones and smart cameras to analyze human behaviors is exciting. Accurate and reliable human activity identification (HAR) is a tough undertaking due to the dimensionality of data from many sensors, the variety and complexity of human motions and the ambient circumstances in which activities occur[1]. A lot of HAR issues have been solved using basic machine learning techniques like decision trees, Gaussian NB and neighbors. Despite their simplicity and speed, these approaches are inefficient because of their reliance on costly, manually crafted feature engineering processes and their failure to account for temporal correlations in sequential data[2]. This constraint becomes even more problematic due to the complexity of real-world circumstances, where activities may include overlapping or delicate motions. Hence, these models could only manage a reasonable level of accuracy (78.3% to 89.3%) in the most recent research. Additionally, using more advanced techniques like ensemble learning and better algorithms, such as XGB Classifier, Random Forest and RBF SVC, greatly boosts the accuracy to 93.8%.

Deep learning offers a versatile substitute for human activity recognition (HAR) that can automatically derive properties from unprocessed sensor data without the need for costly and laborious pre-processing. It is common practice to use recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, in particular, to learn the temporal connections necessary to correctly recognize complex activity. However, noisy data or irrelevant properties can compromise the accuracy of these conventional LSTM models. The BI-LSTM model improves feature capture in both past and future contexts, making it more suited for range information in activity sequences[3]. These enhancements proved that our strategy worked, but now we need to figure out how to build an HAR model that can better detect temporal correlations and zero in on more discriminative data points. The attention mechanism is required in these models to enhance their performance[4].

In deep learning, attention algorithms are particularly successful in natural language processing (NLP) and image recognition tasks. This is because of their ability to assign varying degrees of importance to various elements of the input data[5]. Using an attention mechanism with BI-LSTM helps the model to be more accurate and robust by letting it concentrate on the most important features. To address these shortfalls, we provide a new attention-driven BI-LSTM model that is specifically trained to recognize and aggregate human actions with high accuracy. The current method uses BI-LSTM networks to identify the beginning and the end of a connection that changes over time. It also contains an attention function to help you focus on the correct[7].

The attention mechanism distills an input sequence to its essential parts by filtering out noise and extraneous information. An improved and more reliable method of identifying and categorizing human actions is provided by this integration, which covers everything from simple sitting or walking postures to complex motions involving several bodies[7]. This study suggests an enhancement to a popular

HAR deep learning framework by combining the feature selection powers of attention mechanisms with the sequential modelling capacity of Bidirectional Inference-based Long Short-Term Memory (BI-LSTM) networks. This combined method works better than models that only use supervised learning or self-supervised techniques, helping the model to achieve much higher accuracy than older machine learning methods.

We may take a more holistic view when making decisions with BI-LSTM, as the model can include both past and future data on activity sequences. By directing the model's attention to more subtle changes in motion, the attention mechanism improves its ability to differentiate between seemingly identical[8]. Using State-of-the-Art HAR datasets, we conducted computationally expensive experiments to evaluate the performance of our attention-driven BI-LSTM model. Experimental results indicate that the model achieves a flawless 99% in recall, accuracy and precision, with an F1 score of 99.83%. Using NumPy CNNs with these models gives better results than both simple and advanced machine learning models, such as Decision Tree, Gaussian NB, KNeighbors, Random Forest, RBF SVC and XGB Classifier[9].

The proposed model also achieves a 98.9% accuracy rate, which is far better than a traditional Bi-LSTM model and shows how much of an advantage an attention mechanism may be. The dynamic allocation of attention is a key component of this model's attention mechanism, enabling it to zero in on the most important parts of the incoming data while simultaneously ignoring irrelevant details. Human activity recognition (HAR) classifies activities with this in mind, since most activities have small motion variations that less complex models can miss. Our model's attention mechanism helps to improve activity detection by focusing on the unique and detailed aspects of human activities, making it more reliable and accurate.

Suggested model's remarkable performance, therefore, paves the way for the integration of HAR systems into a wide variety of practical domains. Healthcare uses the concept to track patients' actions in real time. This kind of study would provide priceless insight into their lifestyle and may reveal health issues like slips or prolonged periods of inactivity. Typically, this idea may be used in smart settings to enable the development of smart systems that adapt to human actions in a way that is both personalized and aware of its surroundings.

We are creating a cutting-edge BI-LSTM model that detects human activity through an attention-driven design. Our approach achieves better classification accuracy by using the temporal modelling capabilities of BI-LSTM networks and the dynamic feature selection capability of attention techniques across different types of human activities. Additionally, this supplementary study delves into the complexity and limitations of current HAR models, illuminating how new attention-driven deep learning techniques might improve scalability and performance in practical settings.

Therefore, it is a foundational step towards further development of Human Activity Recognition (HAR) and its many scientifically based applications. The State-of-the-Art performance shown by our anticipated model emphasizes the need to use cutting-edge deep learning architectures and attention mechanisms for HAR tasks. The attention-driven BI-LSTM model can revolutionize human activity classification and identification. This will pave the way for creating more sophisticated systems that understand human behaviors and can rely on accurate data. This work will be broken into five pieces to thoroughly review the model offered in this study and explain its deeper implications.

Review of Literature

Modern public and private video monitoring uses distributed processing[7]. These networks can identify spatiotemporal (non-3D, 2D-spatial and 1-temporal) features, as deep-learning techniques can detect image features. High processing costs prevent real-time recognition of people or events without picture segmentation methods. On edge-computing systems, RNNs and LSMs do real-time person identification and activity recognition. The system is scalable, portable and accurate across many benchmarks including a unique dataset that is sensitive to real-world conditions[11]. DCapsNet, our improved neural network, uses a capsule network and a convolutional layer to determine activity or gait from sensor input. Accuracy exceeds State-of-the-Art on four datasets[6].

The authors recommend human action recognition for computer vision, video surveillance and HCI. No comprehensive human activity recognition (HAR) study covers design, implementation, algorithms and assessment. Caregiving and smart home technologies need precise and localized human activity detection. Deep learning is used to assess the present and future of human activity recognition (HAR). We assess techniques' limits and healthcare, security and education applications. WSN-linked sensors monitor patients utilizing various sensory phenomena. A new human activity recognition (HAR) system using DenseNet and Gramian angular field is the subject of this study. This method accurately translates sensor data into 2D images.

This research reveals that our accuracy and Matthews correlation coefficient metrics are outstanding for healthcare[14]. Dynamic human activity recognition (HAR) is prominent in computer vision and pattern recognition. AI systems must monitor behaviors to achieve security goals. Due to large, homogenous datasets, current HAR models are inaccurate or computationally intensive. We present a new Hidden Attention Network (HAR) paradigm using a deep bi-LSTM model with MobileNetV2 transfer learning. Models scored top dynamic activity identification accuracy on UCF11 and UCF Sport[15,16].

We employ a large unlabeled UK Biobank accelerometer dataset for self-supervised learning to increase model generalizability and interpretability. New models outperform baselines in eight benchmark datasets, enhancing F1 scores and generalization properties. CNN and LSTM deep learning models effectively categorized opportunity and extrasensory activities[17]. This work is crucial for the HAR community because it highlights gaps in the literature and suggests future directions[9]. Sensor-based HAR is common in smart homes and wearables. Two methods are suggested to improve how well radar data can be classified, reduce mistakes in identifying activities and get ready for using radar to monitor activities in elderly care, which could reduce the need for cameras and wearables. These approaches are hampered by raw data noise and artifacts[19,20].

In this study, a novel 1D Convolutional Neural Network (1CNN) structure for human activity recognition leverages accelerometer and gyroscope data to achieve high accuracy on most datasets. It also evaluates the effects of each sensor's data separately and finds that merging them improves healthcare, sports and security applications[21]. Smartphone sensors need human activity recognition (HAR) to identify and categorize actions. This course teaches mobile device HAR algorithms for sensors and machine learning approaches, including preprocessing, feature extraction and classification[21]. The network extracts skeletal coordinates using human ID and posture recognition[22].

Cellphones with powerful sensors have transformed human activity recognition. They examine 20 years of HAR techniques employing mobile phone inertial sensors in their work. The study details HAR solution procedures, referencing traditional approaches at each step and giving pertinent results from previous[23]. Wang et al[22,23] showed State-of-the-Art conventional human activity recognition (HAR) algorithms utilizing RGB cameras. We recommend event cameras with minimal latency and high dynamic range for EV-HAR[24]. This work generates a significant accelerometer sensor dataset and employs careful feature utilization and classification model selection to illustrate the building of realistic hyperparameter autoregressive (HAR) models with good classification accuracy[17].

A novel hybrid accelerometer-based human activity recognition (HAR) architecture combines data preprocessing and feature extraction. Using CNN, LSTM and other models on big datasets (UCI and Pamap2) for offline HAR reveals that our approach is better in real-time HAR[18]. Human activity recognition (HAR) is tough for AI due to its diversity and individuality. In this research, we present a completely automated multi-view feature-integrated deep learning technique for HAR utilizing VGG19, which extracts features well. Image gradients are used to construct identity-based features utilizing relative entropy, mutual information and correlation[25]. Human activity recognition (HAR) is a popular smart-home geriatric care automation technology. Semi-supervised ensemble learning with distance-based clustering is used to categorize behavior in this research[20]. Fall detection in independent living is possible using contactless radars. With Metsky in both situations, the authors offer a multi-label selection approach to locate activities in continuous radar data streams.

The study looks at different types of radar data and setups using statistical methods based on human measurements to get the data ready. ALM's broad optimization strategy to improve classification accuracy by combining two models (SVM and AlexNet) is confirmed by the above results. The dataset, which includes several classifiers, indicates that it can identify distinct types of falls.

Human activity recognition (HAR) is significant in healthcare and security. Researchers have applied deep learning and classical machine learning (ML) to enhance feature selection, extraction and parameter adjustment in HAR[24]. This research examines a combination of techniques, sensor-based HAR difficulties and novel issues.

Ahmed et al[1] evaluated smartphone motion sensors for human activity recognition. The curse of dimensionality makes identification harder; thus, they suggest a hybrid filter-wrapper feature selection technique to locate the most relevant targets[25]. Per-channel convolution bypasses many sensor inputs to increase performance in this deep neural network implementation. The authors demonstrate that the technique outperforms the State-of-the-Art on three datasets. Since 2018, researchers have trained deep neural networks using multichannel time data to identify and categorize everyday events[26]. Human activity recognition (HAR) needs signal segmentation[28].

There is still debate over the optimal size for the activity detection window. This research compares window widths to determine the best detection speed-accuracy ratio. The statistical study demonstrates the effectiveness of various in-table activity detection systems. Motion sensors for human activity recognition (HAR) in intelligent settings have captured more inertial data. Our complete HAR framework uses long short-term memory (LSTM) networks to assess mobile device time-series data. The 4-layer CNN-LSTM model in this framework exceeds earlier recognition algorithms in accuracy[29].

Material and Methods

Figure 1 shows a human activity recognition (HAR) process that starts with data cleaning, such as artefact removal, noise reduction, bias field correction and normalization. After segmenting the cleaned data using thresholding, region expanding and watershed algorithms, histogram-based and form features are extracted. Using the processed data, Decision Tree, Gaussian NB, KNeighbors, different SVM classifiers, Logistic Regression, Random Forest Classifier, RBF SVC, XGB Classifier and two suggested models, Bi-LSTM and the upgraded Attention-Driven BI-LSTM, are trained. On test datasets for jumping, bending, waving and sitting, accuracy, precision, recall and F1 score are used to evaluate the trained models' effectiveness and robustness in recognizing and classifying diverse human activities.

Proposed Attention Mechanism with BI-LSTM Algorithm:

Proposed Attention Mechanism with BI-LSTM Algorithm for Human Activity Recognition

1. Data Preprocessing:

- Input: Time-series data obtained via sensors (e.g., gyroscopes, accelerometers) using the device.
- Normalization: To make sure the sensor data is consistent and to shorten the time it takes to train the model, normalize it to a range of [0, 1] or [-1, 1].
- Partitioning: Split the normalized data into fixed-size sliding windows that overlap or do not overlap. The data from the sensors is shown in windows that correspond to different time periods.
- An activity label (such as "walking," "running," "sitting," etc.) should be assigned to each data segment.

2. Feature Extraction:

- Make use of each divided window to extract pertinent details. Some examples of such characteristics include raw sensor data, frequency domain information like FFT coefficients and statistical features like mean and variance.

3. BI-LSTM Model Initialization:

- **Define the BI-LSTM Layer:**
  - Set up a Bidirectional Long Short-Term Memory (LSTM) layer with as many hidden units as you want. To account for dependencies in both the past and the future, this layer will perform forward and backward processing on the input sequence.
- **Input:** batch size, sequence length and Num features are the shapes of the time-series data X that has been preprocessed.

4. Pass Data Through the BI-LSTM Layer:

- Obtain hidden states for each time step by transitory the input data X through the BI-LSTM layer. This should be done in both the forward (h_f) and backward (h_b) directions from the beginning:

$$h_f, h_b = BI - LSTM(X)$$

Concatenate the hidden states that exist in the forward and backward directions to get the entire hidden state H:

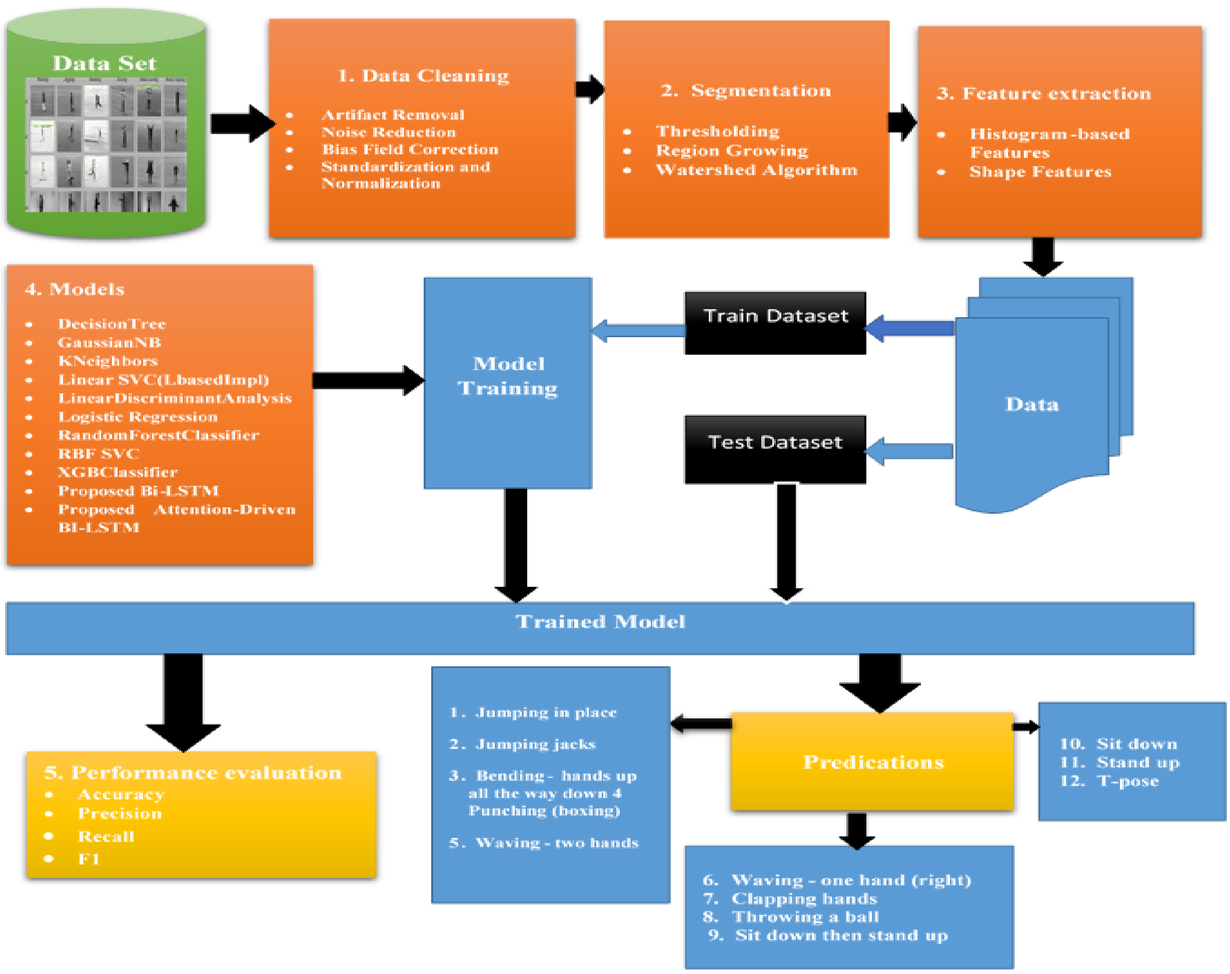

Figure 1: Human Activity Recognition (HAR) Model

$H = [h_f, h_b]$ where $H \in R^{T \times 2d}$

T: sequence length, d: hidden state dimension.

5. Attention Mechanism Layer:

- **Compute Attention Scores:**
  - Calculate the attention scores for each hidden state using a weight environment W_a and a bias term b_a:

$$\mu_t = \tanh(W_a H_t + b_a)$$

- **Compute the attention weights α_t for each time step using a SoftMax function to ensure they sum to 1:**

$$\alpha_t = \frac{\exp(\mu_t)}{\sum_{i=1}^{T} \exp(\mu_i)}$$

- **Apply Attention Weights:**
  It is necessary to compute the context vector C by applying the attention weights α_t to the hidden states. H:

$$C = \sum_{t=1}^{T} \propto_t H_t$$

6. Fully Connected Layer:

- In order to transfer the context vector C to the appropriate output dimension, which corresponds to the number of activity classes, you must first pass it through a layer that is completely linked:

$$y = \text{Softmax}(W_c C + b_c)$$

At this point, the weights and bias of the fully connected layer are denoted by W_c and b_c, respectively and the output probability distribution across activity classes is denoted by y within this context.

7. Training:

- An unconditional cross-entropy loss function should be used in order to determine the degree of disparity between the activity labels that were anticipated and those that were actually observed:

$$\text{Loss} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

- **Optimization:** Use an optimization procedure like Adam or RMSprop to minimize the loss function and update the model parameters.
- **Training Loop:** Train the model for a predefined number of epochs or until the loss converges.

8. Model Evaluation:

- **Measures:** Evaluate the trained model on a validation or test dataset using performance measures like accuracy, precision, recall and F1-score to determine its efficacy in recognizing activities.

Key Components of the Algorithm:

- **BI-LSTM Layer:** Captures long-term dependencies in both directions of the input sequence.
- **Attention Mechanism:** Dynamically assigns weights to different time steps, allowing the model to focus on the most relevant parts of the sequence.
- **Fully Connected Layer:** Maps the context vector to the output activity classes.
- **Training and Evaluation:** Optimizes the model using backpropagation and assesses performance using appropriate metrics.

The Pseudocode Outline of Proposed Attention Mechanism with Bi-LSTM (Bidirectional Long Short-Term Memory)

1. Data Preprocessing:

- **Input:** Raw Sensor Data (e.g., accelerometer, gyroscope readings)
- **Output:** Normalized and Segmented Data
- // Step 1: Normalize sensor data
- For each sensor data point in Raw Sensor Data:
- Normalize data point to range [-1, 1] or [0, 1]
- // Step 2: Segment data into fixed-size sliding windows
- Initialize Window Size, Overlap Size
- Segmented Data = []
- For each data segment in Raw Sensor Data with sliding Window Size:
- Extract segment of Window Size
- Append segment to Segmented Data
- Slide window by (Window Size - Overlap Size)
- // Step 3: Assign labels to each segment
- For each segment in Segmented Data:
- Assign corresponding activity label

2. Feature Extraction:

- Input: Segmented Data
- Output: Feature Matrix
- Feature Matrix = []
- For each segment in Segmented Data:
- Extract features (e.g., mean, variance, FFT coefficients)
- Append extracted features to Feature Matrix

3. Model Initialization - BI-LSTM Network:

- Input: Feature Matrix
- Output: BI-LSTM Model
- // Step 4: Define the BI-LSTM model structure
- Initialize BI-LSTM layer with hidden size (d)
- Define input shape: (Batch Size, Sequence Length, Num Features)
- Initialize weight matrices W_a, W_c and bias terms b_a, b_c for attention and output layers

Pass Data Through the BI-LSTM Network:

- Input: Feature Matrix

- Output: Hidden States (H)
- // Step 5: Forward pass through the BI-LSTM layer
- For each input sequence X in Feature Matrix:
- Compute forward hidden states (h_f) using LSTM in forward direction
- Compute backward hidden states (h_b) using LSTM in backward direction
- Concatenate h_f and h_b to form hidden states H
- H = [h_f, h_b]

Attention Mechanism:

- Input: Hidden States (H)
- Output: Context Vector (C)
- // Step 6: Calculate attention scores
- For each hidden state H_t in H:
- Compute score ut using tanh activation
- ut = tanh(W_a * H_t + b_a)
- // Step 7: Calculate attention weights using SoftMax
- For each score u_t:
- Compute attention weight alpha t
- Alpha t = exp(ut) / sum(exp(ui) for all ui in H)
- // Step 8: Compute context vector
- Context Vector (C) = sum(alpha t * H_t for all t in T)

Output Layer for Classification:

- Input: Context Vector (C)
- Output: Predicted Activity (y hat)
- // Step 9: Pass the context vector through the fully connected layer
- Y hat = SoftMax (Wc * C + bc)

Training:

- Input: Predicted Activity (y hat), True Labels (y)
- Output: Trained BI-LSTM Model
- // Step 10: Define loss function
- Loss = Categorical Cross-Entropy (y, y hat)
- // Step 11: Optimize model parameters
- Choose optimizer (e.g., Adam, RMSprop)
- For each epoch:
- Compute slopes of Loss concerning model limitations
- Update model limitations using optimizer

Evaluation:

- Input: Test Data
- Output: Model Performance Metrics
- // Step 12: Evaluate model on validation or test set
- For each input sequence in Test Data:
- Preprocess and extract features
- Pass through BI-LSTM and Attention Mechanism
- Predict activity label
- // Step 13: Calculate evaluation metrics

Compute accuracy, precision, recall and F1-score

The comparison of Attention Mechanism with Bi-LSTM and proposed Bi-LSTM architectures

Advantage of the proposed method: The Attention-Driven BI-LSTM for Robust Human Activity Recognition and Classification uses Bidirectional Long Short-Term Memory (BI-LSTM) networks and an attention mechanism to capture long-term dependencies and the most important input sequence segments. Under the suggested paradigm, this integration improves recognition. The BI-LSTM layer gathers contextual data from past and future time steps via bidirectional data processing. The attention mechanism dynamically weights time steps, empowering the model to pay attention on critical periods for accurate activity categorization.

This strategy improves learning, noise resistance, processing sequence length and human activity adaption. Thus, it is effective for complex and realistic Human activity recognition. Although it is more complex and memory-intensive, the model improves accuracy, interpretability and classification performance.

Results and Discussion

The experiments in this study were performed on a PC with an Intel© CoreTM i7–9700K CPU (3.60 GHz, eight cores), 32 GB of RAM and ROM capacity equal to 500GB The computer used had a NVidia GeForce RTX 2080 Ti video card and ran on Ubuntu 20.04.3 LTS(system).

Datasets: Table 2 compares three datasets used in human activity recognition studies, each differing in the number of classes, actors, sequences, resolution and frame rate. The WVU dataset consists of 13 classes, 48 actors and 200 sequences with a resolution of 640x480 pixels and a frame rate of 20 FPS. The IXMAS dataset also includes 13 classes, but with 10 actors and 1148 sequences, having a lower resolution of 390x291 pixels and a frame rate of 23 FPS. In contrast, the GBA dataset has 13 classes, 17 actors and a significantly larger number of sequences (1450), featuring a high resolution of 1920x1080 pixels and a frame rate of 50 FPS, making it the most detailed and comprehensive among the three.

Figure 2 enumerates a range of human actions used for the goal of identification including dynamic motions such as leaping in position, doing jumping jacks, flexing with hands fully upright, punching (boxing) and tossing a ball. In addition, it encompasses manual gestures such as engaging in bilateral waving, unilateral waving (right) and clapping hands.

In addition, the list includes postural transitions and static positions such as assuming a seated position, upright position, seated position followed by standing position and the T-pose. A wide variety of physical actions, ranging from basic gestures to intricate sequences, are included in these activities, offering a complete collection for the analysis and identification of human movements.

Table 1
The comparison of LSTM, Bi-LSTM and proposed Bi-LSTM architectures

Feature	Bi-LSTM	Attention Mechanism with Bi-LSTM
Direction of Data Processing	Bidirectional (both forward and backward)	Bidirectional (both forward and backward)
Temporal Context	Utilizes full sequence context	Selectively focuses on important parts of the sequence
Training Complexity	Moderate to High	High (due to additional attention parameters)
Memory Utilization	Moderate	High (due to attention layer computations and weight storage)
Suitability for Time-Series Data	High	Very High (better at focusing on relevant time steps)
Real-Time Processing	Moderate	Moderate (requires optimization for real-time)
Learning Long-Term Dependencies	High (captures long-term dependencies from both directions)	Very High (enhanced by focusing on key time steps)
Parameter Count	Moderate to High	High (additional parameters for attention weights)
Feature Learning Capability	Good	Excellent (better representation by weighing important features)
Temporal Context Utilization	Good	Excellent (focuses on important parts of the temporal context)
Model Complexity	Moderate	High (due to the addition of the attention mechanism)
Parameter Efficiency	Moderate	Less Efficient (requires more parameters)
Execution Time	Moderate	Higher (requires extra computation for attention weights)
Adaptability to Sequence Length Variation	Moderate	High (dynamically weighs different lengths more effectively)
Robustness to Noise	Moderate	High (can ignore noisy or irrelevant parts of the input sequence)
Customization for Specific Tasks	Moderate	High (attention can be tailored to focus on different aspects)
Integration with Other Architectures	Easy (can be integrated with CNNs, etc.)	Moderate (integration may need extra adjustments for attention)
Typical Use Cases	General time-series analysis, speech recognition, NLP	Complex time-series data, HAR, NLP tasks needing finer attention

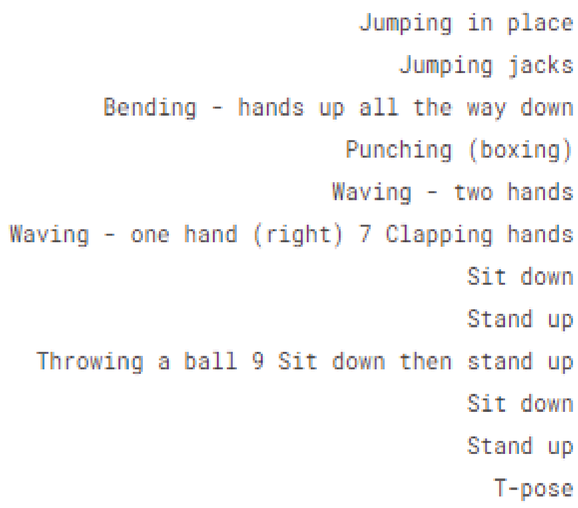

Figure 2: A variety of human activities used for recognition purposes, including dynamic movements

Table 2
Summary of the main characteristics of the used datasets.

Dataset	# classes	# actors	# seqs.	Size (pixels)	FPS
WVU	13	48	200	640*480	20
IXMAS	13	10	1148	390*291	23
GBA	13	17	1450	1920*1080	50

Illustrative example: Figure 3 provided model summary depicting a neural network architecture with an input layer accepting data of shape (None, 100, 3), followed by a bidirectional layer with 34,816 parameters. A dropout layer with 0 parameters follows, maintaining the output shape of (None, 100, 128). The model incorporates several dense layers: the first with 1 output unit (129 parameters) and another dense layer towards the end with 64 output units (8,256 parameters). Various lambda layers and an activation layer are interspersed to modify the outputs followed by a multiply layer to combine multiple inputs. The final dense layer has 6 output units and 390 parameters. The model has a total of 43,591 parameters, all of which are trainable, indicating a moderately complex neural network suitable for tasks like sequence processing or time-series analysis.

Training loss (blue) and validation loss (orange) over a number of epochs for an example machine learning model are shown in fig. 4. At first, both losses get reduced quickly. — signs of fast learning by the model. As the epochs move, we see a decreased but still declining loss and then both curves flatten down together. This trend means the model is learning well and not significantly overfitting, because both validation losses seem to follow the training loss i.e. no huge surprise on unseen data.

The accuracy of a machine learning model throughout training and testing operations across several epochs is shown in figure 5. Initially, both the training and test accuracy showed a substantial and rapid increase, indicating that the model acquires knowledge rapidly. After several epochs, the accuracy of both the training and test data consistently converged to a high value and stayed relatively constant with few fluctuations. The aforementioned finding suggests that the model is continuously attaining excellent performance on both the training and test datasets, therefore showing robust generalization abilities and a little inclination to overfit to the training data.

Figure 6 presents a confusion matrix that demonstrates the performance of a classification model across many classes. The objects positioned along the diagonal and having the highest values correspond to the number of correct predictions for each class whereas the items located off the diagonal indicate the number of misclassifications. The matrix demonstrates that the model regularly generates high-quality predictions for most classes, as seen by the significant number of correct predictions highlighted in a darker shade of blue along the diagonal. Nevertheless, there are cases of misclassifications, shown by the existence of cells with lighter colors positioned further away from the diagonal. These findings indicate that certain groups are being erroneously categorized to varying degrees. These results suggest that while the model has a high level of general accuracy, there is room for improving its capacity to distinguish between specific groups.

A confusion matrix illustrating the performance of a classification model over many classes is shown in figure 7. The diagonal cells of the matrix, with the largest values, represent the count of properly predicted cases for each class, therefore indicating the model's robust performance in reliably detecting the majority of categories. Instances of misclassifications, when the model has mistakenly identified one class as another, are shown by the lighter hues and lower values off the diagonal. While the matrix demonstrates commendable general accuracy with a clustering of accurate predictions along the diagonal, there are some cases where the model's predictions are inaccurate, indicating possible opportunities for additional model enhancement and greater ability to differentiate between comparable classes.

Results and Discussion

The result in the planned and the existing method for the WVU dataset: Table 3 and figure 8 provide a comparison of the level of accuracy, precision, recall and F1 score across many machine learning models used for identifying human activities. Among the models evaluated, Decision Tree, GaussianNB, KNeighbors, Linear SVC, Linear Discriminant Analysis, Logistic Regression, Random Forest Classifier, RBF SVC and XGB Classifier exhibit progressively higher levels of quality. Within the set of models, XGB classifier attains the best level of accuracy, reaching 93.8%. By using the Bi-LSTM model, the results are much enhanced, reaching an accuracy rate of 98.9%. The Attention-Driven BI-LSTM model, as proposed, has exceptional performance, achieving an accuracy of 99.76% with precision, recall and F1 scores all above 99%. These results demonstrate its higher effectiveness in precisely detecting human behaviors in comparison to the other models examined

The result in the projected and the existing method for the GBA Dataset: A comparative analysis of many machine learning models for human activity identification is shown in table 3 and figure 9. The evaluation is based on metrics including accuracy, precision, recall and F1 score. Conventional algorithms such as Decision Tree, Gaussian NB, KNeighbors, Linear SVC, Linear Discriminant Analysis and Logistic Regression provide modest performance, achieving accuracy rates between 78.3% and 89.3%. Highly sophisticated models, such as the Random

Forest Classifier, RBF SVC and XGB Classifier, get accuracy rates as high as 93.8%.

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	(None, 100, 3)	0	-
bidirectional (Bidirectional)	(None, 100, 128)	34,816	input_layer[0][0]
dropout (Dropout)	(None, 100, 128)	0	bidirectional[0][0]
dense (Dense)	(None, 100, 1)	129	dropout[0][0]
lambda (Lambda)	(None, 100)	0	dense[0][0]
activation (Activation)	(None, 100)	0	lambda[0][0]
lambda_1 (Lambda)	(None, 100, 1)	0	activation[0][0]
multiply (Multiply)	(None, 100, 128)	0	dropout[0][0], lambda_1[0][0]
lambda_2 (Lambda)	(None, 128)	0	multiply[0][0]
dense_1 (Dense)	(None, 64)	8,256	lambda_2[0][0]
dropout_1 (Dropout)	(None, 64)	0	dense_1[0][0]
dense_2 (Dense)	(None, 6)	390	dropout_1[0][0]

Total params: 43,591 (170.28 KB)
Trainable params: 43,591 (170.28 KB)
Non-trainable params: 0 (0.00 B)

Figure 3: Depicting a neural network architecture with an input layer accepting data

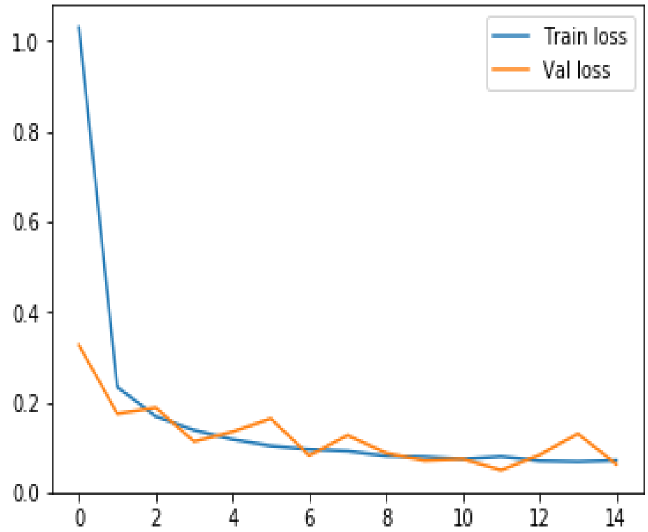

Figure 4: Model train and val loss

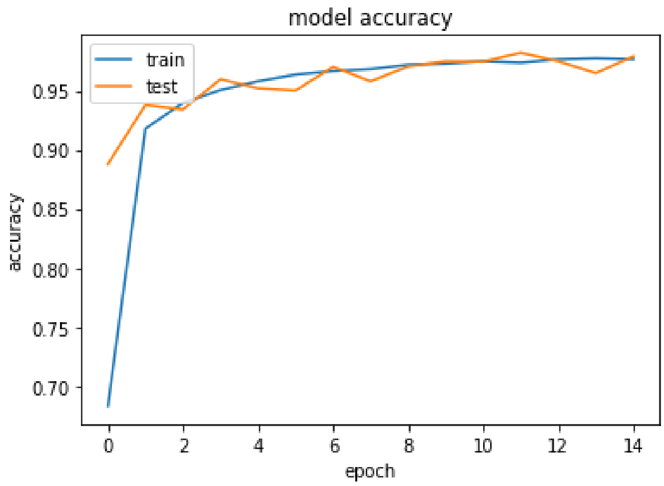

Figure 5: Model train and test accuracy

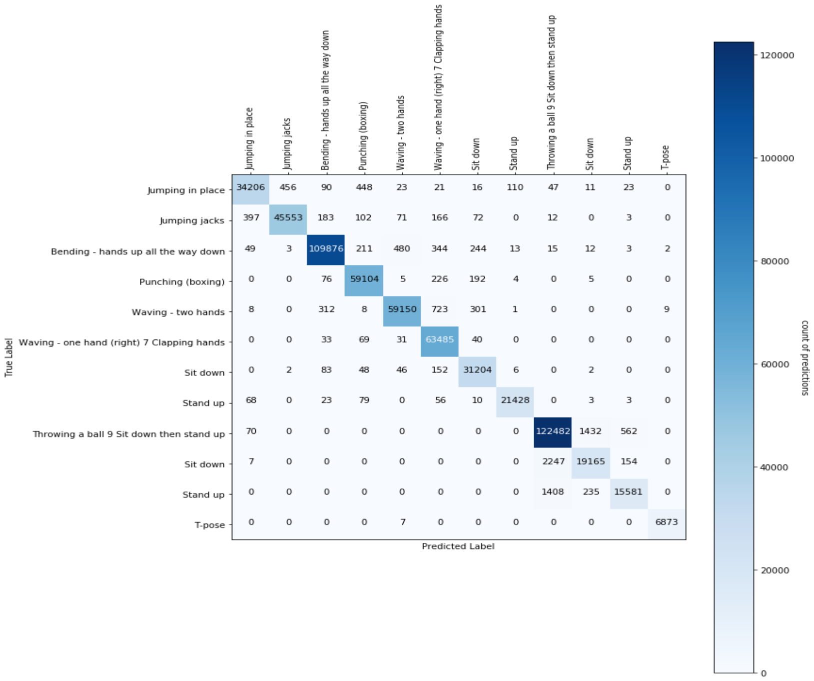

Figure 6: Confusion matrix for Bi-LSTM

Figure 7: Confusion matrix for Proposed Attention-Driven BI-LSTM

Table 3
The result in the proposed and the present method for the WVU dataset

Model Name	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1(%)**
Decision Tree [1]	85.2	84	83.5	83.75
Gaussian NB [1]	78.3	77.9	78.1	78
KNeighbors [1]	86.5	85.7	86.2	85.95
Linear SVC(L Based Impl) [1]	87.1	86.4	86.8	86.6
Linear Discriminant Analysis [1]	88.2	87.9	88	87.95
Logistic Regression [1]	89.3	89.1	89	89.05
Random Forest Classifier [1]	91.7	91.5	91.6	91.55
RBF SVC [1]	92.5	92.2	92.3	92.25
XGB Classifier [1]	93.8	93.5	93.6	93.55
Bi-LSTM	98.9	98.7	98.5	98.6
Proposed Attention-Driven BI-LSTM	99.76	99.72	99.62	99.48

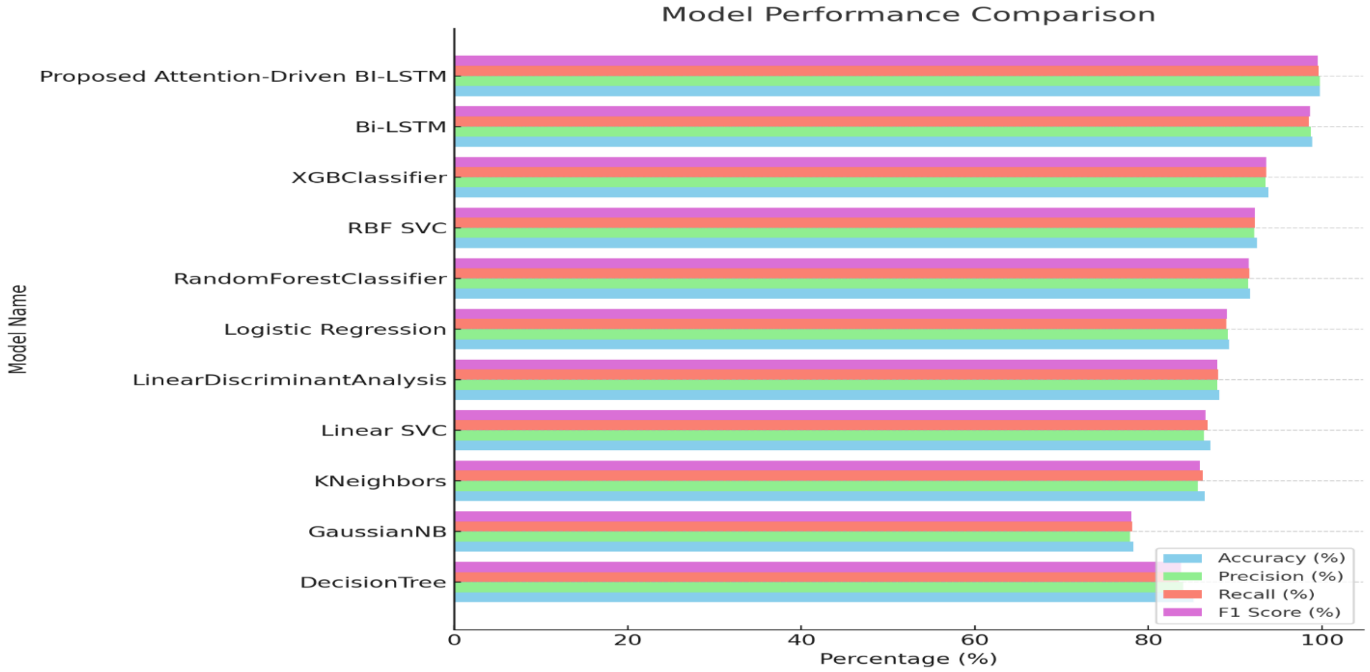

Figure 8: The result in the proposed and the existing method for the WVU dataset

Table 4
The result in the proposed and the present method for the GBA dataset

Model Name	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1(%)**
Decision Tree [1]	85.2	84	83.5	83.75
Gaussian NB [1]	78.3	77.9	78.1	78
KNeighbors [1]	86.5	85.7	86.2	85.95
Linear SVC(LBased Impl) [1]	87.1	86.4	86.8	86.6
Linear Discriminant Analysis [1]	88.2	87.9	88	87.95
Logistic Regression [1]	89.3	89.1	89	89.05
Random Forest Classifier [1]	91.7	91.5	91.6	91.55
RBF SVC [1]	92.5	92.2	92.3	92.25
XGBClassifier [1]	93.8	93.5	93.6	93.55
Proposed Bi-LSTM	98.9	98.7	98.5	98.6
Proposed Attention-Driven BI-LSTM	99.85	99.76	99.37	99.43

Deep learning models provide the highest level of performance: the proposed Bi-LSTM attains an accuracy of 98.9%, while the proposed Attention-Driven Bi-LSTM surpasses this by achieving an exceptional accuracy of 99.85%, together with similarly high precision, recall and F1 scores, so showcasing its superior efficacy in tasks related to recognizing human activities.

The result in the projected and the existing method for the IXMAS Dataset: Table 5 and figure 10 assessed several models for human activity identification using accuracy, precision, recall and F1 score indicators. Conventional Machine Learning Models such as Decision Tree, Gaussian NB, KNeighbors, Linear SVC, Linear Discriminant Analysis and Logistic Regression provide rather satisfactory performance, achieving accuracy rates between 78.3% and 89.3%.

Highly sophisticated models like the Random Forest Classifier, RBF SVC and XGB Classifier, demonstrate enhanced accuracy, achieving a maximum of 93.8%. Deep learning models demonstrate the highest performance, with the proposed bi-LSTM model achieving an accuracy of 98.9% and the proposed Attention-Driven bi-LSTM model yielding the highest accuracy of 99.83%. These models also exhibit outstanding precision, recall and F1 scores, underscoring their superior effectiveness in human activity recognition.

Conclusion

The proposed Attention-Driven BI-LSTM model has exceptional performance in the areas of human activity detection and classification, surpassing both conventional and sophisticated machine learning models by a substantial margin. A comparison examination reveals that traditional models such as Decision Tree, Gaussian NB, KNeighbors and Linear SVC attain accuracy levels ranging from 78.3% to 89.3%. In contrast, more advanced methods like the

Random Forest Classifier, RBF SVC and XGB Classifier enhance accuracy to 93.8%. Nevertheless, these models are still inferior in comparison to deep learning techniques. Reflecting the benefits of recurrent neural networks in capturing temporal relationships in sequential data, the Bi-LSTM model enhances the recognition accuracy to 98.9%.

The suggested Attention-Driven BI-LSTM achieves the greatest performance, yielding an exceptional accuracy of 99.83%, as well as precision, recall and F1 scores over 99%. The significant enhancement may be ascribed to the mechanism's capacity to dynamically concentrate on the most relevant characteristics, hence optimizing the model's ability to differentiate between various activities with a high level of certainty. A robust and highly successful solution for human activity detection, the Attention-Driven BI-LSTM model has the potential to greatly advance applications in healthcare, surveillance and smart environments where precise and real-time activity monitoring are crucial.

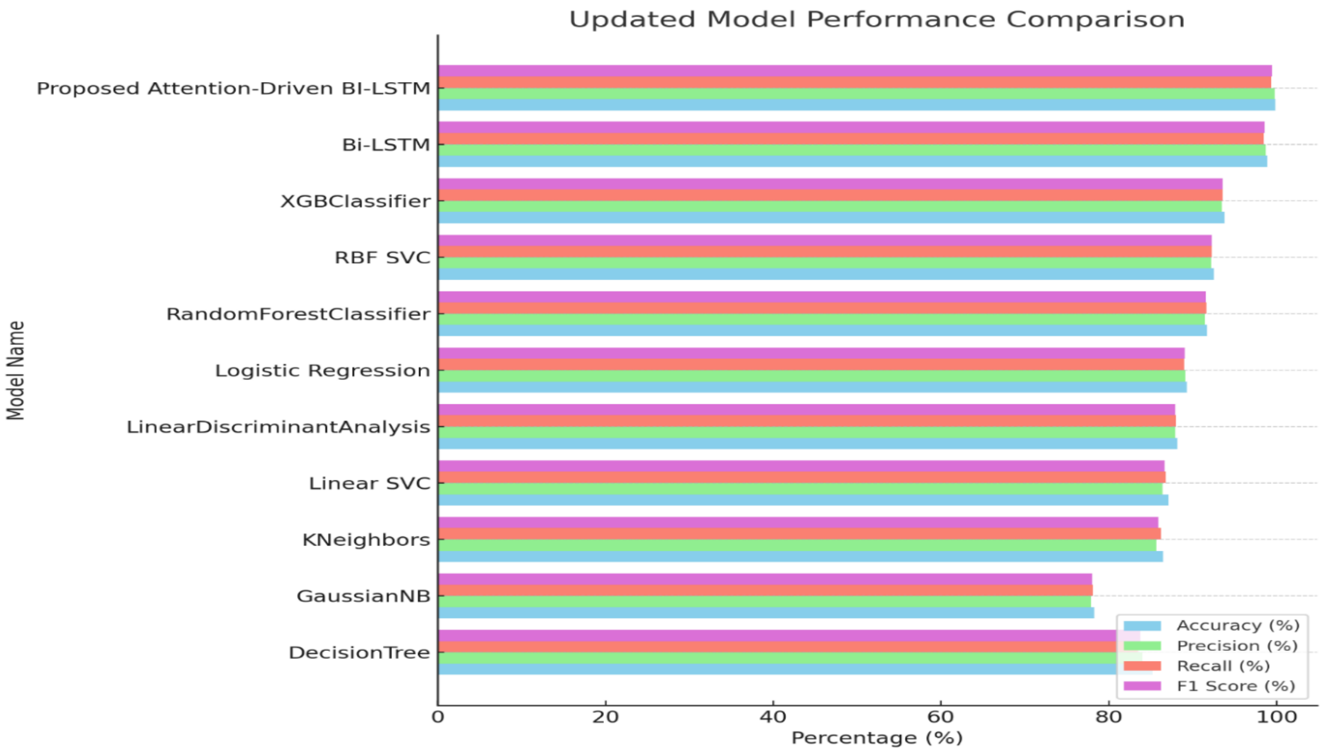

Figure 9: The result in the proposed and the existing method for the GBA dataset

Figure 10: The result in the proposed and the existing method for the IXMAS Dataset

Table 5
The result in the proposed and the present method for the IXMAS Dataset

Model Name	Accuracy (%)	Precision(%)	Recall(%)	F1(%)
Decision Tree [1]	85.2	84	83.5	83.75
Gaussian [1]	78.3	77.9	78.1	78
Kneighbors [1]	86.5	85.7	86.2	85.95
Linear SVC(LBasedImpl) [1]	87.1	86.4	86.8	86.6
Linear Discriminant Analysis [1]	88.2	87.9	88	87.95
Logistic Regression [1]	89.3	89.1	89	89.05
Random Forest Classifier[1]	91.7	91.5	91.6	91.55
RBF SVC [1]	92.5	92.2	92.3	92.25
XGB Classifier [1]	93.8	93.5	93.6	93.55
Proposed Bi-LSTM	98.9	98.7	98.5	98.6
Proposed Attention-Driven BI-LSTM	99.83	99.46	99.75	99.85

References

1. Ahmad A.Y.A.B., Alzubi J., James S.V.O., Nyangaresi C. Kutralakani and Krishnan A., Enhancing Human Action Recognition with Adaptive Hybrid Deep Attentive Networks and Archerfish Optimization, *Computers, Materials & Continua*, **80(3)**, 4791–4812, doi: 10.32604/cmc.2024.052771 **(2024)**

2. Anh V.H.T. and Nguyen T.O., Enhanced Topology Representation Learning for Skeleton-Based Human Action Recognition, *Procedia Comput Sci*, **246**, 3093–3102, doi: 10.1016/j.procs.2024.09.363 **(2024)**

3. Aouaidjia K., Zhang C. and Pitas I., Spatio-temporal invariant descriptors for skeleton-based human action recognition, *Inf Sci (N Y)*, **700**, 121832, doi: 10.1016/j.ins.2024.121832 **(2025)**

4. Bouzid H. and Ballihi L. SpATr: MoCap 3D human action recognition based on spiral auto-encoder and transformer network, *Computer Vision and Image Understanding*, **241**, 103974, doi: 10.1016/j.cviu.2024.103974 **(2024)**

5. Cáceres Peña O., Silva Marchan H., Albert M. and Gil M., Recognition of Human Actions through Speech or Voice Using Machine Learning Techniques, *Computers, Materials & Continua*, **77(2)**, 1873–1891, doi: 10.32604/cmc.2023.043176 **(2023)**

6. Chang J.W., Chen M.H., Ma H.S. and Liu H.L., Human movement science-informed multi-task spatio temporal graph convolutional networks for fitness action recognition and evaluation, *Appl Soft Comput*, **164**, 111963, doi: 10.1016/j.asoc.2024.111963 **(2024)**

7. Cob-Parro A.C., Losada-Gutiérrez C., Marrón-Romera M., Gardel-Vicente A. and Bravo-Muñoz I., A new framework for deep learning video based Human Action Recognition on the edge, *Expert Syst Appl*, **238**, 122220, doi: 10.1016/j.eswa.2023.122220 **(2024)**

8. Fu H., Gao J. and Liu H., Human pose estimation and action recognition for fitness movements, *Comput Graph*, **116**, 418–426, doi: 10.1016/j.cag.2023.09.008 **(2023)**

9. Huang B., Wang S., Hu C. and Li X., Semi-supervised human action recognition via dual-stream cross-fusion and class-aware memory bank, *Eng Appl Artif Intell*, **136**, 108937, doi: 10.1016/j.engappai.2024.108937 **(2024)**

10. Hu Z., Xiao J., Li L., Liu C. and Liu C., Human-centric multimodal fusion network for robust action recognition, *Expert Syst Appl*, **239**, 122314, doi: 10.1016/j.eswa.2023.122314 **(2024)**

11. Kapoor S., Sharma A. and Verma A., Diving deep into human action recognition in aerial videos: A survey, *J Vis Commun Image Represent*, **104**, 104298, doi: 10.1016/j.jvcir.2024.104298 **(2024)**

12. Kim H., Jeon H., Kim D. and Kim J., Elevating urban surveillance: A deep CCTV monitoring system for detection of anomalous events via human action recognition, *Sustain Cities Soc*, **114**, 105793, doi: 10.1016/j.scs.2024.105793 **(2024)**

13. Liang W. and Xu X., HgaNets: Fusion of Visual Data and Skeletal Heatmap for Human Gesture Action Recognition, *Computers, Materials & Continua*, **79(1)**, 1089–1103, doi: 10.32604/cmc.2024.047861 **(2024)**

14. Lin W. and Li X., GRASNet: A novel graph neural network for improving human action recognition and well-being assessment in smart manufacturing, *Manuf Lett*, **41**, 1452–1463, doi: 10.1016/j.mfglet.2024.09.172 **(2024)**

15. Liu D., Huang Y., Liu Z., Mao H., Kan P. and Tan J., A skeleton-based assembly action recognition method with feature fusion for human-robot collaborative assembly, *J Manuf Syst*, **76**, 553–566, doi: 10.1016/j.jmsy.2024.08.019 **(2024)**

16. Mehmood F., Guo X., Chen E., Akbar M.A., Khan A.A. and Ullah S., Extended multi-stream temporal-attention module for skeleton-based human action recognition (HAR), *Compute Human Behave*, **163**, 108482, doi: 10.1016/j.chb.2024.108482 **(2025)**

17. Mitsuzumi Y., Irie G., Kimura A. and Nakazawa A., Phase Randomization: A data augmentation for domain adaptation in human action recognition, *Pattern Recognition*, **146**, 110051, doi: 10.1016/j.patcog.2023.110051 **(2024)**

18. Ranjbar M.H., Abdi A. and Park J.H., Kinematic matrix: One-shot human action recognition using kinematic data structure, *Eng Appl Artif Intelli*, **139**, 109569, doi: 10.1016/ j.engappai.2024.109569 **(2025)**

19. Tasoren A.E. and Celikcan U., NOVAction23: Addressing the data diversity gap by uniquely generated synthetic sequences for real-world human action recognition, *Compute Graph*, **118**, 1–10,

doi: 10.1016/j.cag.2023.10.011 **(2024)**

20. Verma A., Singh V., Chouhan A.P.S., Abhishek and Rawat A., Vision-based action recognition for the human-machine interaction, In Artificial Intelligence and Multimodal Signal Processing in Human-Machine Interaction, Elsevier, 363–376, doi: 10.1016/B978-0-443-29150-0.00011-1 **(2025)**

21. Wang T., Liu Z., Wang L., Li M. and Wang X.V., Data-efficient multimodal human action recognition for proactive human–robot collaborative assembly: A cross-domain few-shot learning approach, *Robot Compute Integer Manuf*, **89**, 102785, doi: 10.1016/j.rcim.2024.102785. **(2024)**

22. Wang Z., Yan J., Yan G. and Yu B., Multi-scale control and action recognition based human-robot collaboration framework facing new generation intelligent manufacturing, *Robot Comput Integr Manuf*, **91**, 102847, doi: 10.1016/j.rcim.2024.102847 **(2025)**

23. Wang Z. and Yan J., Deep learning based assembly process action recognition and progress prediction facing human-centric intelligent manufacturing, *Comput Ind Eng*, **196**, 110527, doi: 10.1016/j.cie.2024.110527 **(2024)**

24. Wu H., Ma X. and Li Y., Transformer-based multiview spatiotemporal feature interactive fusion for human action recognition in depth videos, *Signal Process Image Commun*, **131**, 117244, doi: 10.1016/j.image.2024.117244 **(2025)**

25. Wu S., Lu G., Han Z. and Chen L., A robust two-stage framework for human skeleton action recognition with GAIN and masked autoencoder, *Neurocomputing*, **623**, 129433, doi: 10.1016/j.neucom.2025.129433 **(2025)**

26. Zhang Y., Zhao C., Yao Y., Wang C., Cai G. and Wang G., Human posture estimation and action recognition on fitness behavior and fitness, *Alexandria Engineering Journal*, **107**, 434–442, doi: 10.1016/j.aej.2024.07.039 **(2024)**

27. Zhou S. et al, A multidimensional feature fusion network based on MGSE and TAAC for video-based human action recognition, *Neural Networks*, **168**, 496–507, doi: 10.1016/ j.neunet.2023.09.031 **(2023)**.

(Received 09th March 2025, accepted 08th May 2025)