SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I2P110

Volume 11 Issue 2, 92-102, February 2024

Original Article

A Hybrid Visionary for Unveiling Human Motion in the Face of Occlusion with Mask R-CNN, RNN, and MHT

Jeba Nega Cheltha[1], Chirag Sharma[2]

[1]*School of Computer Science & Engineering, Lovely Professional University, Punjab, India.*
[1]*Department of Computer Science and Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan, Rajasthan, India.*
[2]*School of Computer Science & Engineering, Lovely Professional University, Punjab, India.*

[1]*Corresponding Author : nega.cheltha@skit.ac.in*

Received: 02 November 2023 Revised: 06 December 2023 Accepted: 03 January 2024 Published: 06 February 2024

Abstract *- Addressing the intricate challenges of Human Motion Detection (HMD), this research presents a pioneering hybrid methodology integrating advanced computer vision and deep learning techniques. Focused primarily on mitigating the impact of occlusion in visual data, the proposed approach employs a Mask Region-based Convolutional Neural Network (Mask R-CNN) for precise motion segmentation. The dual challenges of self-occlusion and partial-occlusion are specifically targeted. The three-fold strategy encompasses motion segmentation, object classification, and tracking algorithms to discern and identify human motion accurately. Motion segmentation involves isolating the moving object within video frames, followed by object classification utilizing a Recurrent Neural Network (RNN) to determine the human presence and to tune the parameter of RNN; this work introduced a novel hybrid Whale Optimization Algorithm and Red Deer Algorithm (WOA-RDA), which gives better convergence speed with high accuracy. To tackle the persistence of occlusion, particularly self-occlusion, Multiple Hypothesis Tracking (MHT) is introduced for robustly tracking human gestures. An innovative aspect of the proposed approach lies in the integration of an RNN trained with 2D representations of 3D skeletal motion, enhancing the model's understanding of complex human movements. The proposed methodology is rigorously evaluated on diverse datasets, incorporating scenarios with and without occlusion. Experimental results underscore the effectiveness of the hybrid approach, showcasing its ability to accurately identify human motion under varying conditions, thereby advancing the field of human motion detection.*

Keywords *- Human Motion Detection, Mask R-CNN, Multiple Hypothesis Tracking, Occlusion, RNN.*

1. Introduction

Visual analysis of human motion is one of the most active areas of computer vision research, aiming to recognize, track, and detect people from image sequences, including humans, as well as to interpret human behaviors [1, 2].

Many studies have concentrated on the detection and tracking of basic human models by using data on skin tone or static backgrounds [3]. The automatic detection of human behaviors from pictures or video sequences can be summed up as a human motion recognition task [4].

For many years, motion analysis and tracking from monocular or stereo video have been suggested for use in surveillance and visual security. Many methods for detecting human motions have been developed using various kinds of sensors [5, 6]. The two primary types of these methods are wearable and non-wearable. Because wearable technologies connect sensors directly to body parts, they are able to record human behaviors in great detail. However, wearable sensors are easily worn out and frequently forgotten. Wearable techniques, therefore, lack dependability and convenience, and the unintended trade-off is that, despite their potential for outstanding performance in lab settings, they show a decline in performance in real-world settings due to occlusion, perspective, and illumination variations [7-9].

Conversely, non-wearable techniques get over these problems and end up being the more popular option. Usually, stationary sensors are positioned in fixed places. Human motion characteristics can be taken out of picture and video frames for motion recognition and classification by employing optical and depth cameras [10].

However, privacy concerns are raised by computer vision algorithms, which are severely limited by the light. Furthermore, compared to earlier versions, modern cameras perform noticeably better in low light because of recent technological advancements, particularly in the area of camera sensors [11, 12].

In addition to color video, specific cameras, often referred to as "RGB+depth" cameras, also record depth information. This allows for the robust extraction of human figures, for example, as a set of 3D points, even in low-light situations. Thus, occlusion presents the most significant limits among the three previously discussed issues [13, 14]. Henceforth, the proposed work incorporated the pioneering hybrid methodology integrating advanced computer vision and deep learning methods to understand the notion of occlusion.

The more accurate segmentation from the input data using the recently established Deep Learning techniques [15]. Consequently, Convolutional Neural Networks (CNNs) [16], which are frequently employed in standard Machine Learning techniques to analyze both simple and complicated actions, are inferior to deep learning-based HAR. Techniques based on CNNs have incredibly advanced state-of-the-art in motion estimation and semantic segmentation.

Given the variability in human activity duration and the ambiguity of activity boundaries, the continuous segmentation of input sensor sequence data is a complex operation. However, their suitability for the joint semantic motion segmentation challenge has not been investigated.

This task is intrinsically difficult because of a number of issues, such as the camera's ego motion, illumination variations between consecutive frames, motion blur, and shifting pixel displacements brought on by motion at different speeds [17, 18]. As a consequence, the proposed work utilized Mask R-CNN for precise motion segmentation. For classification, conventional CNN [19] and Artificial Neural Network (ANN) [20] techniques that incorporate multiple hidden layers and rely on learning representations from unprocessed input are collectively referred to as Deep Learning; the network learns numerous layers of non-linear information processing.

However, the recognition accuracy may be decreased by either of the two approaches if they result in inaccurate labeling. Therefore, the hybrid WOA-RDA optimized RNN is implemented in this work to determine human presence for a better classification process. Moreover Multiple Hypothesis Tracking (MHT) is introduced for robustly tracking human gestures, which is one of the first effective visual tracking systems. The significant contributions of the proposed work are discussed as follows,

- To mitigate the impact of occlusion in visual data, a Mask R-CNN is utilized for precise motion segmentation.
- To identify human motion accurately, the Recurrent Neural Network with the hybrid whale–red deer algorithm is employed for classification.
- To tackle the persistence of occlusion, particularly self-occlusion, Multiple Hypothesis Tracking (MHT) is introduced for robustly tracking human gestures.

2. Proposed Methodology

Human Activity Recognition (HAR) tasks suffer from occlusion because it causes important motion data to be lost, which impairs the performance of awareness algorithms. Because it can impact the precision of computer vision techniques, including object identification, tracking, and recognition, occlusion handling is crucial in videos.

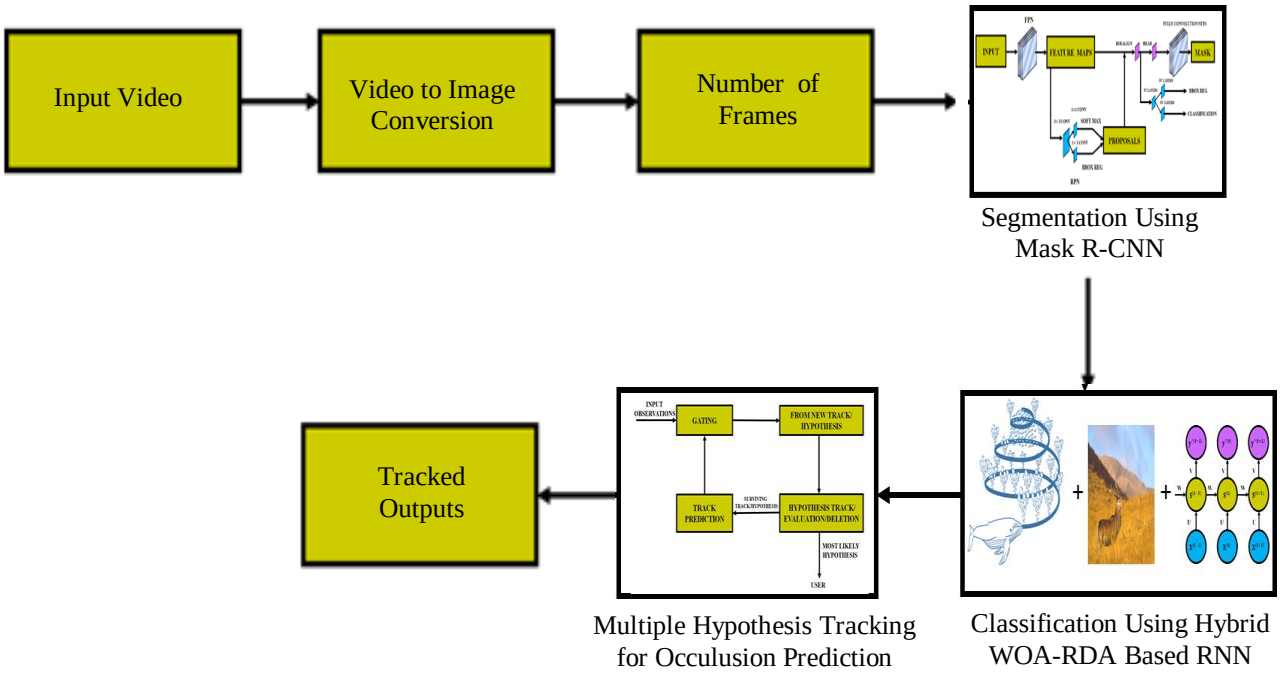


Fig. 1 Flow diagram for the proposed framework

Henceforth, the proposed research work implements the pioneering hybrid methodology integrating advanced computer vision and deep learning techniques for mitigating the impact of occlusion in visual data. The Flow diagram for the developed work is illustrated in Figure 1, which combines various topologies for mitigating occlusion.

To make the process of addressing occlusion easier, this technique initially converts the input video to video frames. With the application of Mask-R CNN, an input image is segmented into 265 pixels for a straightforward process of identifying human motion, and it involves isolating the moving object within video frames. The segmented image is fed to the classification process; the RNN technique is utilized to classify human motion with maximum accuracy and enhance the model's understanding of complex human movements. Furthermore, Multiple Hypothesis Tracking is implemented to tackle the persistence of occlusion, which can robustly track human gestures. Finally the tracked output is attained by utilizing the proposed techniques respectively.

2.1. Modelling of Mask-RNN for Segmentation

One of the main objectives of Mask R-CNN is to develop a framework for computer vision applications like image or instance segmentation. The field of computer vision examines methods to enhance computers' comprehension of digital visuals, such as pictures or movies.

The idea of Mask R-CNN is simple: For every candidate item, Faster R-CNN produces two outputs: a class label and a bounding box offset. Mask R-CNN adds a third branch to this output, which produces the object mask, a binary mask that shows the pixels in the bounding box where the object is located.

Since the extra mask output differs from the class and box outputs, the much finer spatial arrangement of an item needs to be retrieved. Pixel-to-pixel alignment, the primary characteristic of Faster R-CNN missing data, is then integrated. Figure 2 shows the schematic diagram for Mask R-CNN used in the segmentation process.

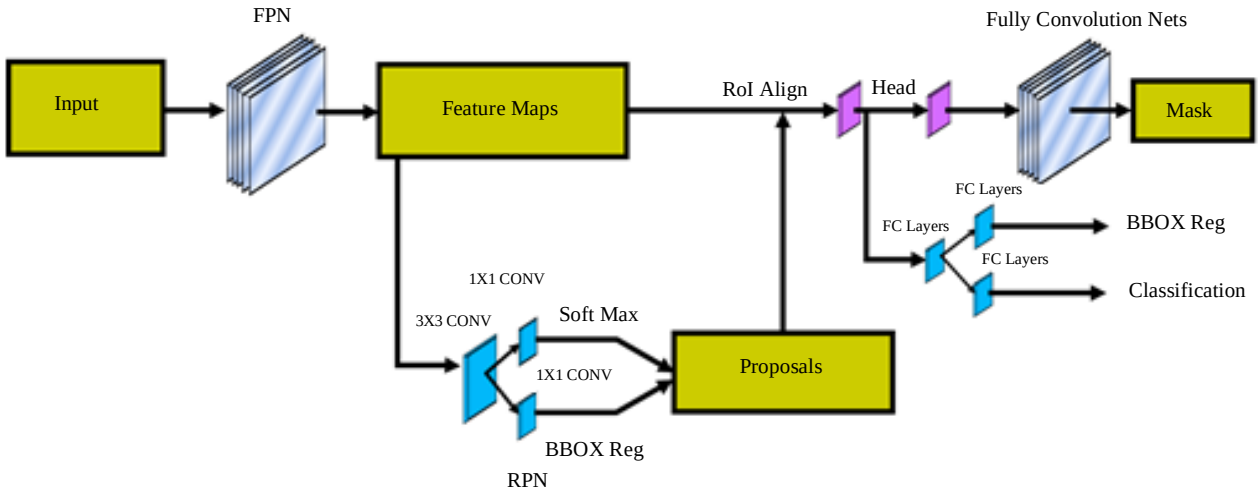


Fig. 2 Schematic diagram for mask R-CNN for segmentation

As demonstrated in Figure 2, the Mask-R-CNN is a versatile and straightforward to understand object instance segmentation technique. The region of interest extraction process makes use of a Region Proposal Network (RPN). It produces a binary mask while estimating box offset and class for every area of interest. Every Region of Interest (RoI) can have its segmentation mask predicted using an Fully Convolutional Network (FCN), as the mask shows the convolutional correlation between pixels directly.

Initially, Mask R-CNN creates a set of candidate regions that might contain objects using an RPN. An input image is fed into the RPN, a fully convolutional network, which yields a set of rectangular bounding boxes along with the associated objectness scores. Each anchor box is a preconfigured box with a specific size and aspect ratio.

The RPN is trained to categorize each box as either foreground or background and to regress the box's coordinates to match the object better. One way to formulate the RPN is as follows:

$$L(\{P_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(P_i, P_i*) + \lambda\frac{1}{N_{reg}}\sum_i P_i * L_{reg}(t_i, t_i*) \quad (1)$$

Here, t_i specifies the parameterized coordinates of the predicted bounding box, P_i* denotes ground truth table, P_i indicates predicted probability of anchor, t_i* represents the vector for the ground truth box, N_{cls} and N_{reg} specifies the normalization terms, λ denotes the balancing parameter, and L_{cls}, L_{reg} indicates the log loss function for classification and regression.

In the second stage, Mask R-CNN applies the RoI align layer to extract features from each suggested region. The faster R-CNN's RoI pooling layer, which can lead to misalignment between the input areas and the recovered features, is improved by the RoI align layer. The RoI align layer computes the precise values of the features at any spatial point by using bilinear interpolation rather than quantizing the areas. The layer of RoI alignment can be written as:

$$y = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{c} {x_{ij}}^{k}\, G_{ij} \quad (2)$$

Where, ${x_{ij}}^{k}$ represents the value of k[th] channel of the input feature map at (i, j), y indicates the output feature map of size $c \times n \times n$, n specifies the number of bins, and G_{ij} denotes the fraction of the bin that is covered by the region.

Two parallel branches are then fed with the characteristics that the RoI align layer extracted: one for mask prediction and the other for bounding box regression and classification. With the exception of predicting K class-specific bounding boxes rather than just one generic box, the first branch is comparable to the one utilized in faster R-CNN. A binary mask for each of the K classes is produced by the fully convolutional network in the second branch. One way to express the mask branch is as follows:

$$M = f(\phi, W_m) \quad (3)$$

Calculating the average binary cross-entropy loss over K classes and m2 pixels yields the mask loss.

$$L_{mask} = -\frac{1}{km^2}\sum_{k=1}^{K}\sum_{u=1}^{m}\sum_{v=1}^{m} y_{kuv} log\hat{y}_{kuv} + (1 - y_{kuv})\, log(1 - \hat{y}_{kuv}) \quad (4)$$

Here, $\hat{y}_{kuv}$ represents predicted mask value and y_{kuv} denotes ground truth mask value for class k and pixel (u, v).

The input image is efficiently segmented and provides better identification of human motion by employing the proposed Mask R-CNN. Following this, hybrid WOA-RDA optimized RNN is used for the classification process that is discussed as follows.

2.2. Modelling of Hybrid WOA-RDA Optimized RNN

2.2.1. Recurrent Neural Network

A DL algorithm having a recurrent feedback component is called RNN. An input layer, a hidden layer, and an output layer make up a standard RNN structure. RNN is able to retain previously processed data that has been altered from the input because every neuron in the hidden layer has a state feedback mechanism. Consequently, RNN is better equipped to handle sequential data. The ability of the RNN to process sequences of varying lengths due to its recurrent structure is another crucial characteristic, and the structure of the RNN is represented in Figure 3. Therefore, better motion classification and recognition are attained by utilizing RNN to extract sequential human motion features based on radar.

By examining the impact of the motion state at each time step on the state at the next step, RNNs are able to capture the patterns of motion evolution. Motion state is a high-level characteristic of the motion pattern that is extracted from the unprocessed observation of the human postures. The state at time step n, denoted as h_n, is dependent upon the previous state h_{n-1} as well as the related observation x_n.

$$h_n = \varphi(W_{h-h}h_{h-1} + W_{x-h}x_n + b) \quad (5)$$

Where, W_{x-h} specifies the weight connecting to the state, b indicates the bias, W_{h-h} specifies the weights that indicate the impact of the previous state on the current one, and φ denotes the non-linear function. Because h_{h-1} is likewise dependent on its previous state, all observations up to time n are used in the production of h_h.

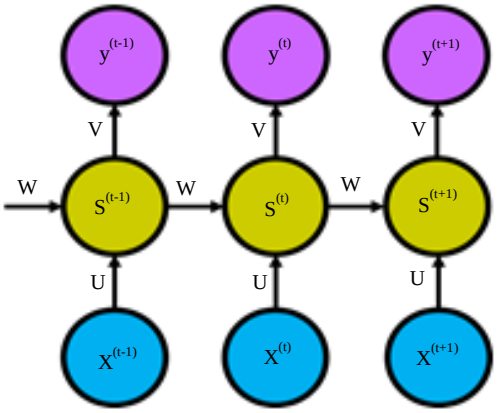


Fig. 3 Structure of RNN

$$S^{(t)} = f\left(Ux^{(t)} + Ws^{(t-1)} + {b^{(t)}}_{s}\right) \quad (6)$$

X(t) specifies input into the network at time t. The state, or s(t), of the hidden layer, is the outcome of non-linear mapping that begins with the weighted sum of the current network input and the historical state of the preceding instant. This mapping is written as,

Ultimately, the comparable non-linear mapping with respect to the weighted total of states yields a network estimate $\hat{y}\ (t)$, which is stated as,

$$\hat{y}^{(t)} = f\,(Vs^{(t)} + {b^{(t)}}_{y} \quad (7)$$

Furthermore, to tune the parameter of the developed RNN technique, the optimization algorithm of the hybrid WOA-RDA technique is introduced in this work, which is described below.

2.2.2. Hybrid WOA-RDA Algorithm

The Advantages of two nature-inspired algorithms are combined in a revolutionary meta-heuristic technique called the hybrid Red Deer Algorithm (RDA) and Whale Optimization Algorithm (WOA). The RDA explores the search space and identifies the optimal solutions by imitating red deer's natural behaviors, such as fighting, shouting, and mating. To determine the most promising areas and improve the solutions, the WOA mimics the humpback whale's hunting tactics, which include encircling, bubble-netting, and spiral attacks.

In order to balance population diversity and convergence, the hybrid algorithm uses the RDA for global exploration and the WOA for local exploitation. It also makes use of a dynamic switching mechanism, and the flowchart for the developed hybrid WOA-RDA algorithm is represented in Figure 4.

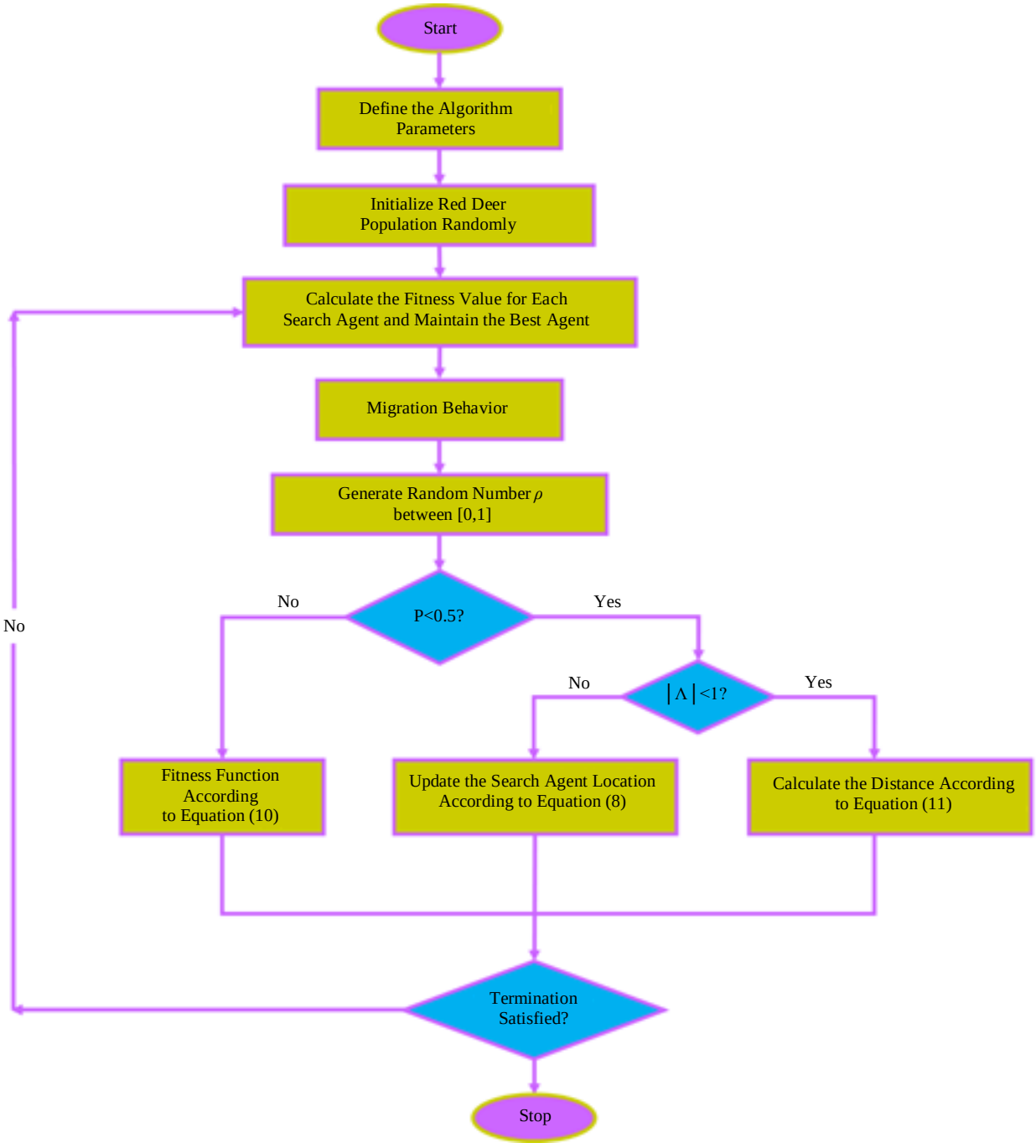


Fig. 4 Flowchart for the proposed hybrid WOA-RDA algorithm

In other words, whales randomly search based on one another's positions. The mathematical model can be written like this:

$$\vec{D} = |\vec{C}.\vec{X}_{rand} - \vec{X}| \quad (8)$$

Here, $\vec{X}_{rand}$ represents the position vector of individual whales arbitrarily selected from the current population.

Using Equation (1), which creates the red deer population initially, it is feasible to model the RDA mathematically.

$$RD = X_1, X_2, X_3, \ldots., X_{Nvar} \quad (9)$$

Following that, using Equation (2), the fitness of each individual in the population is determined.

$$Value = f(RD) = f(X_1, X_2, X_3, \ldots., X_{Nvar}) \quad (10)$$

Using the Equation (11), determine the separation between all HN and the male RD,

$$T_j = (\sum_{r \in R}(stag_r - HN^j{}_r)^2)^{\frac{1}{2}} \quad (11)$$

By implementing the proposed hybrid WOA-RDA optimized RNN technique, better classification accuracy, precision, recall, and F1 score are attained with rapid convergence speed correspondingly.

2.3. Modeling of Multiple Hypothesis Tracking

The MHT algorithm is selected to carry out human tracking. For the same reason, the MHT algorithm was chosen to carry out the tracking duties. The ability of MHT to retain a large number of potential hypotheses over several scans indicates that it may be a viable method for carrying out reliable human tracking. Here, a different data association hypothesis is developed in the event of a conflict, as seen in Figure 5. In modern systems, it is the most preferred data association mechanism.

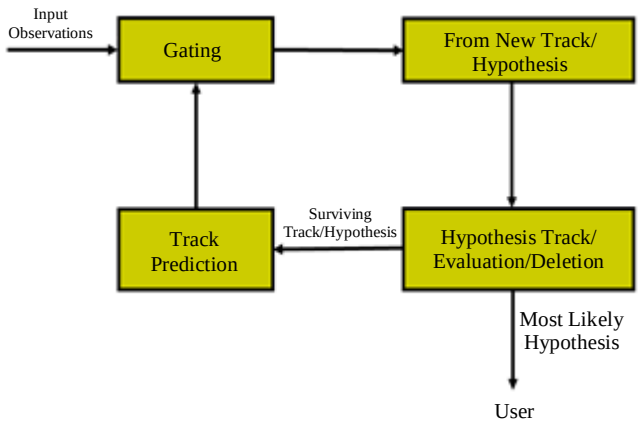


Fig. 5 Flow diagram for MHT

It will present the idea behind the Kalman filter-based tracker for tracking human targets. After that, the specifics of applying MHT to solve the data association problem will be covered, including the formula for determining the likelihood of each hypothesis, the MHT pruning phase for eliminating hypotheses, and parameter tuning.

2.3.1. Kalman Filter-Based Target Track

To track the target human, a tracker based on the Kalman filter is employed. This will focus briefly on the updating target track tracking cycle. A series of measurements taken with the assumption that they are from the same moving target is called a target human track. A vector comprising the tracked location and velocity, for instance, is used to describe the state of the track. This vector is initialized with the first possible human measurement that is received by the human extraction approach.

The target motion state and state covariance are computed using the conventional prediction equations when using a constant velocity model for motion prediction. The measurement prediction equation is updated whenever a new measurement is acquired using the difference between the actual and expected measurements. The following characteristics can be applied to the target motion state and state covariance prediction,

$$\overline{u_t} = Au_{t-1} \quad (12)$$

$$\overline{\Sigma_t} = A\sum_{t-1}A^T + R \quad (13)$$

Here, $\overline{\Sigma_t}$ denotes the prediction about state covariance, $\overline{u_t}$ indicates the target state prediction, where is the motion noise's covariance. The standard formulation that follows can be used to update the Kalman gain, target prediction, and state covariance when a new measurement is obtained.

$$K_t = \overline{\Sigma_t}H^T(H\overline{\Sigma_t}H^T + Q)^{-1} \quad (14)$$

$$u_t = \overline{u_t} + K_t(Z_t{}^i - H\overline{u_t}) \quad (15)$$

$$\Sigma_t = (I - K_tH)\,\overline{\Sigma_t} \quad (16)$$

When an individual is entirely obscured by a larger object, such as a pillar, this system switches from tracking them to observation mode. In other words,

$$F(S1, \ldots, S^{-k}{}_n) = F(S1, \ldots, S_{n-1}).P(S^{-k}{}_n|S_{n-1}) \quad (17)$$

The person's pre-occluded and post-reappearance velocities are used to estimate their movements. The person's location throughout the observation period, S_{ki} can be retrieved based on the estimated motion. The proposed MHT tracking technique effectively tackles the persistence of occlusion, particularly self-occlusion, and robustly tracks the human gesture efficiently.

3. Results and Discussion

In this work, a pioneering hybrid methodology integrating advanced computer vision and deep learning techniques is proposed work for mitigating the impact of occlusion in visual data. By utilizing Mask R-CNN, an input image is efficiently segmented and gives better identification of human motion.

Furthermore, better classification accuracy for precise human motion is achieved by adopting the RNN technique, and the implemented Multiple Hypothesis Tracking approach efficiently tackles the persistence of self-occlusion. The overall results obtained by utilizing the proposed method are discussed as follows.

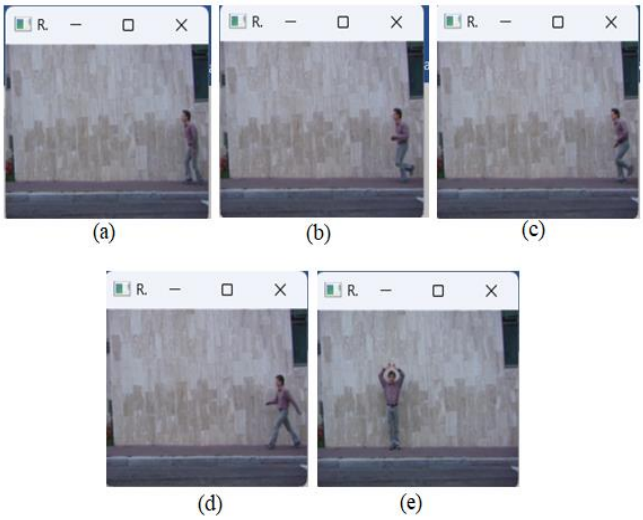


Fig. 6 Frames from Weizmann dataset

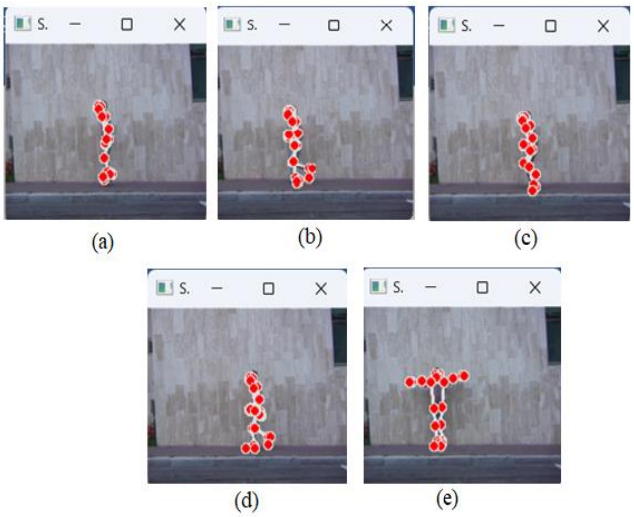


Fig. 7 Critical points from Weizmann dataset using media pipe library

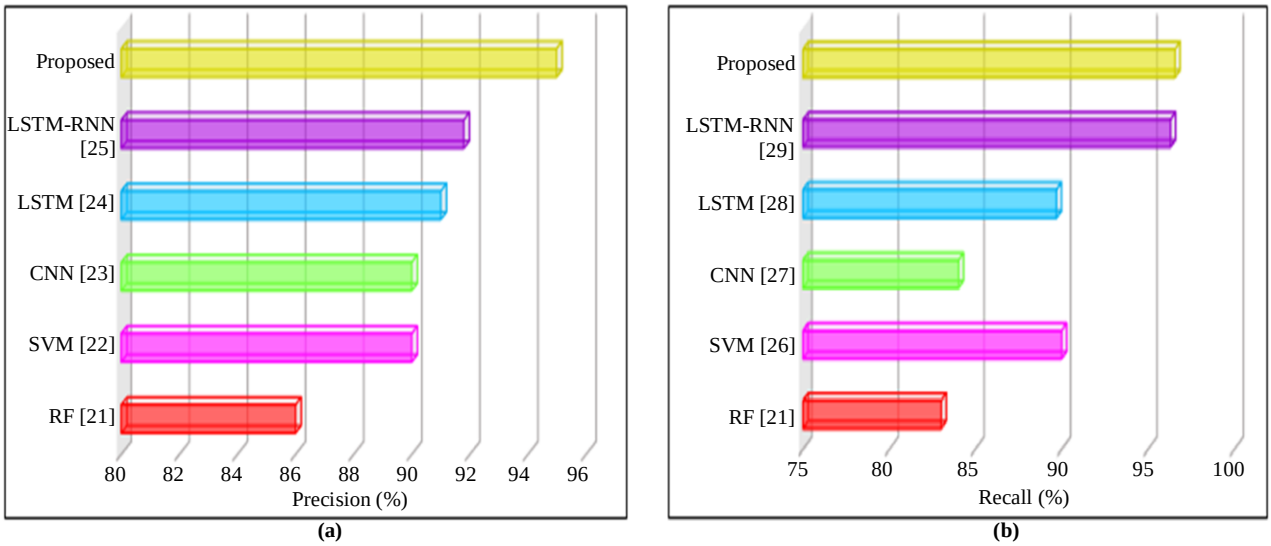
Figure 6 illustrates the frames from the Weizmann dataset, which specifies the different human motions. From the above-stated dataset, human motion is monitored. By using the Mask R-CNN, the image is effectively segmented, as presented in the figure.

Fig. 8 Motion of human from Weizmann dataset (a) Jumping, (b) Running, (c) Skipping, (d) Walking, and (d) Waving.

Key points from the Weizmann dataset using media pipe library are illustrated in Figures 7(a) to 7(d), which analyzed that by utilizing the proposed novel RNN with a hybrid WOA-RDA algorithm, the accuracy is efficiently attained. As illustrated in Figure 8, the motion of Humans from the Weizmann dataset (a) Jumping, (b) Running, (c) Skipping, (d) Walking, and (d) Waving, which is efficiently tracked by the developed multiply hypothesis tracker and the clear image is shown as specifies in the above Figure 8. The proposed RNN-Hybrid WOA-RDA technique is compared with the other conventional topologies like RF, SVM, CNN, LSTM, and LSTM-RNN to determine the best precision and recall, as illustrated in Figure 9. From the graph, it is evident that the proposed system achieves high precision and recall by the values of 94.98% and 96.56%, respectively.

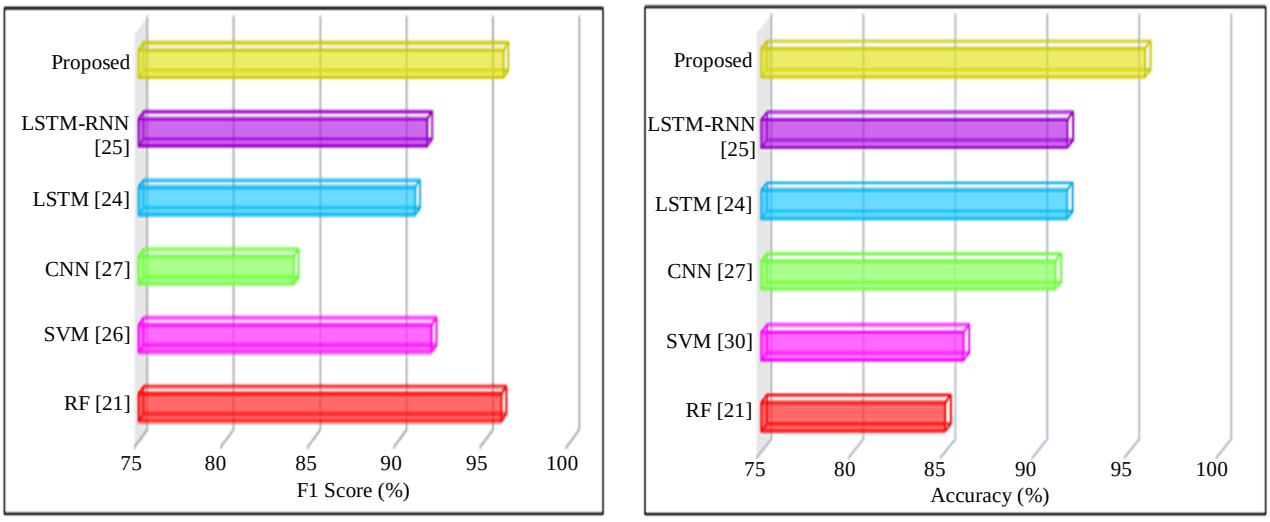
Figure 10 specifies the comparison of F1 score and accuracy with the proposed and existing topologies to foreshow the proficiency of the developed work, which shows that the proposed RNN-hybrid WOA-RDA technique attains better F1 score value and most excellent accuracy compared to the other conventional topologies by the value of 96.12% and 95.87%.

Table 1 compares the proposed novel RNN-hybrid WOA-RDA technique with existing topologies, demonstrating that the developed system performs better in terms of precision, recall, F1 score, and accuracy, achieving values of 94.98%, 96.56%, 96.12%, and 95.87%, respectively. The comparative analysis and overview of relevant literature for human motion detection across a range of classification algorithms, together with their respective accuracy, are shown in Table 2.



(a) **(b)**

Fig. 9 Comparison of (a) Precision, and (b) Recall.



(a) **(b)**

Fig. 10 Comparison of (a) F1 score, and (b) Accuracy.

Table 1. Comparison of performance metrics

No.	Model Name	Comparison of Performance Metrics			
		Precision	Recall	F1-Score	Accuracy
1.	RF	86	83	96	85
2.	SVM	90	90	90.1	86
3.	CNN	90.4	84	84	91
4.	LSTM	91	89.7	91	91.63
5.	LSTM-RNN	91.79	96.3	91.7	91.65
6.	RNN-Hybrid WOA-RDA	94.98	96.56	96.12	95.87

Table 2. Comparative analysis and related literature summary for human motion detection

Ref	Approach	Dataset	Accuracy	Research Goal
[31]	ELM	G3D and UTKinectAction3D Dataset	89%	ELM technique is employed to achieve recognition of human motion.
[32]	ML	Benchmark Dataset	92.47%	With the ML model created with the dataset, an application that offered a quasi-real-time classification of standing or sitting status was created.
[33]	CNN	The Florence 3D Public Dataset	94.08%	The purpose of this study is to use the CNN approach to identify nine activities based on changes in joint distance and movement characteristics.
[34]	U-net	self-Collected Sanitation Dataset	94.75%	A comprehensive U-Net-based HAR framework is employed to achieve motion sensor data-dense prediction.
[35]	LSTM-RNN	Benchmark Dataset	91%	In order to categorize the activities according to their motion characteristic, the LSTM-RNN algorithm was used.
[36]	LSTM	UCF101	80.12%	To identify and follow suspicious conduct in any surveillance setting, development of an intelligent human activity monitoring system based on LSTM.
Proposed	RNN-Hybrid WOA-RDA	Weizmann Dataset	94.98%	The proposed work utilized RNN-hybrid WOA-RDA to achieve better accuracy for human motion detection.

4. Conclusion

In this work, a pioneering hybrid methodology integrating advanced computer vision and deep learning techniques is implemented to address the intricate challenges of HMD, which primarily focuses on mitigating the impact of occlusion in visual data. The three-pronged approach uses tracking algorithms, object classification, and motion segmentation to detect and classify human motion precisely. Henceforth, the proposed approach employs a Mask R-CNN, which effectively identifies the human motion specifically. Furthermore, improved classification accuracy (94.98%), precision (96.56 %), recall (96.12%), and F1 score (95.87%) is attained by utilizing the novel hybrid WOA-RDA optimized RNN topology with rapid convergence speed.

The persistence of occlusion, particularly self-occlusion, is robustly tracked the human gestures by adopting the MHT technique. Using a variety of datasets, including both occlusion-free and occlusion-filled scenarios, the proposed approach is thoroughly assessed. The hybrid approach's accomplishment is demonstrated by the experimental results, which also show its capacity to precisely detect human motion in a variety of settings, contributing to the advancement of human motion detection research.

References

[1] Ying Liu et al., "Human Motion Image Detection and Tracking Method Based on Gaussian Mixture Model and CAMSHIFT," *Microprocessors and Microsystems*, vol. 82, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] Ivan Mutis, Abhijeet Ambekar, and Virat Joshi, "Real-Time Space Occupancy Sensing and Human Motion Analysis Using Deep Learning for Indoor Air Quality Control," *Automation in Construction*, vol. 116, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Md. Milon Islam, Md. Repon Islam, and Md. Saiful Islam, "An Efficient Human Computer Interaction through Hand Gesture Using Deep Convolutional Neural Network," *SN Computer Science*, vol. 1, pp. 1-9, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Xiangui Bu, "Human Motion Gesture Recognition Algorithm in Video Based on Convolutional Neural Features of Training Images," *IEEE Access*, vol. 8, pp. 160025-160039, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Yubin Wu et al., "Hybrid Motion Model for Multiple Object Tracking in Mobile Devices," *IEEE Internet of Things Journal*, vol. 10, no. 6, pp. 4735-4748, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jilong Wang, Chunhong Lu, and Kun Zhang, "Textile-Based Strain Sensor for Human Motion Detection," *Energy & Environmental Materials*, vol. 3, no. 1, pp. 80-100, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Fatemeh Serpush et al., "Wearable Sensor-Based Human Activity Recognition in the Smart Healthcare System," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1-31, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Carolin Helbig et al., "Wearable Sensors for Human Environmental Exposure in Urban Settings," *Current Pollution Reports*, vol. 7, pp. 417-433, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Sakorn Mekruksavanich, and Anuchit Jitpattanakul, "Biometric User Identification Based on Human Activity Recognition Using Wearable Sensors: An Experiment Using Deep Learning Models," *Electronics*, vol. 10, no. 3, pp. 1-21, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Miao He, Guangming Song, and Zhong Wei, "Human Behavior Feature Representation and Recognition Based on Depth Video," *Journal of Web Engineering*, vol. 19, no. 5-6, pp. 883-902, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Thamer Alanazi, Khalid Babutain, and Ghulam Muhammad, "A Robust and Automated Vision-Based Human Fall Detection System Using 3D Multi-Stream CNNs with an Image Fusion Technique," *Applied Sciences*, vol. 13, no. 12, pp. 1-20, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Jing Zhao et al., "Reconstructing Clear Image for High-Speed Motion Scene with a Retina-Inspired Spike Camera," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 12-27, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Snehasis Mukherjee, Leburu Anvitha, and T. Mohana Lahari, "Human Activity Recognition in RGB-D Videos by Dynamic Images," *Multimedia Tools and Applications*, vol. 79, pp. 19787-19801, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Matteo Zago et al., "3D Tracking of Human Motion Using Visual Skeletonization and Stereoscopic Vision," *Frontiers in Bioengineering and Biotechnology*, vol. 8, pp. 1-11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Monica Gruosso, Nicola Capece, and Ugo Erra, "Human Segmentation in Surveillance Video with Deep Learning," *Multimedia Tools and Applications*, vol. 80, pp. 1175-1199, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Ghazaleh Khodabandelou et al., "A Fuzzy Convolutional Attention-Based GRU Network for Human Activity Recognition," *Engineering Applications of Artificial Intelligence*, vol. 118, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Naoya Yoshimura et al., "Acceleration-Based Activity Recognition of Repetitive Works with Lightweight Ordered-Work Segmentation Network," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 2, pp. 1-39, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] Chengqun Song et al., "Spatial-Temporal 3D Dependency Matching with Self-Supervised Deep Learning for Monocular Visual Sensing," *Neurocomputing*, vol. 481, pp. 11-21, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[19] Xiaoli Liu et al., "TrajectoryCNN: A New Spatio-Temporal Feature Learning Network for Human Motion Prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 6, pp. 2133-2146, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Marcin Woźniak et al., "Body Pose Prediction Based on Motion Sensor Data and Recurrent Neural Network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2101-2111, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[21] I.A. Bustoni et al., "Classification Methods Performance on Human Activity Recognition," *Journal of Physics: Conference Series*, vol. 1456, pp. 1-9, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] Siqi Cai et al., "Automatic Detection of Compensatory Movement Patterns by a Pressure Distribution Mattress Using Machine Learning Methods: A Pilot Study," *IEEE Access*, vol. 7, pp. 80300-80309, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[23] X. Wang, and S. Hosseinyalamdary, "Human Detection Based on a Sequence of Thermal Images Using Deep Learning," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 127-132, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[24] Manahil Waheed et al., "An LSTM-Based Approach for Understanding Human Interactions Using Hybrid Feature Descriptors over Depth Sensors," *IEEE Access*, vol. 9, pp. 167434-167446, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] Divya Gaur, and Sanjay Kumar Dubey, "Development of Activity Recognition Model Using LSTM-RNN Deep Learning Algorithm," *Journal of Information and Organizational Sciences*, vol. 46, no. 2, pp. 277-291, 2022. [Google Scholar] [Publisher Link]

[26] Alefiya Laturwala et al., "Human Pose Estimation Using Machine Learning," *Journal of Emerging Technologies and Innovative Research*, vol. 8, no. 12, pp. 403-406, 2021. [Publisher Link]

[27] Xunyun Chang, and Liangqing Peng, "Visual Sensing Human Motion Detection System for Interactive Music Teaching," *Journal of Sensors*, vol. 2021, pp. 1-10, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[28] Sha Ji, and Chengde Lin, "Human Motion Pattern Recognition Based on Nano-Sensor and Deep Learning," *Information Technology and Control*, vol. 52, no. 3, pp. 776-788, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[29] George A. Oguntala et al., "Passive RFID Module with LSTM Recurrent Neural Network Activity Classification Algorithm for Ambient-Assisted Living," *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10953-10962, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[30] Phat Nguyen Huu, and Tan Phung Ngoc, "Hand Gesture Recognition Algorithm Using SVM and HOG Model for Control of Robotic System," *Journal of Robotics*, vol. 2021, pp. 1-13, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[31] Anzhu Miao, and Feiping Liu, "Application of Human Motion Recognition Technology in Extreme Learning Machine," *International Journal of Advanced Robotic Systems*, vol. 18, no. 1, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[32] William Taylor et al., "An Intelligent Non-Invasive Real-Time Human Activity Recognition System for Next-Generation Healthcare," *Sensors*, vol. 20, no. 9, pp. 1-20, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[33] Endang Sri Rahayu et al., "Human Activity Classification Using Deep Learning Based on 3D Motion Feature," *Machine Learning with Applications*, vol. 12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[34] Yong Zhang et al., "Human Activity Recognition Based on Motion Sensor Using U-Net," *IEEE Access*, vol. 7, pp. 75213-75226, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[35] Fabio Carrara et al., "LSTM-Based Real-Time Action Detection and Prediction in Human Motion Streams," *Multimedia Tools and Applications*, vol. 78, pp. 27309-27331, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[36] Rajiv Vincent et al., "Human Activity Recognition Using LSTM/BiLSTM," *International Journal of Advanced Science and Technology*, vol. 29, no. 4, pp. 7468-7474, 2020. [Google Scholar] [Publisher Link]