Heuristic Based Optimized Classifier For HAR

A **Project report** submitted in partial fulfilment of the requirements for the degree of

Bachelor of Technology

In

Electronics and Telecommunication Engineering

by

Swapnil Patil (12120131)

Devyani Ushir (12120175)

Under the guidance of

Prof. Milind Kamble

Academic Year: 2023-24 Sem I

Bansilal Ramnath Agarwal Charitable Trust'

Vishwakarma Institute of Technology, Pune -37

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

Bansilal Ramnath Agarwal Charitable Trust's

VISHWAKARMA INSTITUTE OF TECHNOLOGY, PUNE - 37

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

CERTIFICATE

This is to certify that the **Project Report** entitled **Heuristic Based Optimized Classifier for HAR** has been submitted in the academic year **2023-24**

By

Swapnil Patil (12120131)

Devyani Ushir (12120175)

under the supervision of **Prof. Milind Kamble** in partial fulfilment of the requirements for the degree of Bachelor of Technology in **Electronics and Telecommunication Engineering** as prescribed by Savitribai Phule Pune University.

Guide/Supervisor

Name: Prof. Milind Kamble

Signature:

Head of the Department

Name: Dr. Medha Wyawahare

Signature:

External Examiner

Name:

Signature:

Acknowledgment

We take this opportunity to acknowledge our gratitude to our project guide Prof. Milind Kamble Department of Electronics and Telecommunication Engineering of VIT, Pune for her valuable guidance. We are also thankful to her for her constant motivation and encouragement throughout the project and also for her innovative solutions to the problems that we faced during the course of this project. We sincerely thank her for providing each and every advice needed for the successful completion of this project.

We are also indebted to the faculty of the Department of Electronics and Telecommunication Engineering at VIT, Pune for their support. We express thanks to our parents and all our friends for their constant help, support, and encouragement.

SWAPNIL PATIL

DEVYANI USHIR

ABSTARCT

Human Activity Recognition plays a pivotal role in various domains, from healthcare to wearable technology. This Paper presents a novel approach that combines heuristic-based feature selection with Particle Swarm Optimization to optimize the performance of HAR classifiers. We begin by employing heuristics to intelligently select relevant features from raw sensor data, reducing dimensionality and enhancing classification efficiency. Subsequently, PSO is applied to fine-tune classifier hyperparameters, achieving superior model accuracy and robustness. Our approach initiates by harnessing heuristics to autonomously discern and select pertinent features from raw sensor data. This intelligent feature selection not only trims the data dimensionality but also elevates the overall efficiency of the classification process. Subsequently, we apply PSO, a nature-inspired optimization algorithm, to meticulously fine-tune classifier hyperparameters. Human Activity Recognition (HAR) is indeed a crucial field, impacting various sectors, from healthcare to technological advancements in wearables. The fusion of heuristic-based feature selection and Particle Swarm Optimization (PSO) stands out as an innovative approach in optimizing the performance of HAR classifiers. This method's strength lies in its initial use of heuristics, which intelligently identify and select pertinent features from raw sensor data. By doing so, it efficiently reduces the dimensionality of the dataset, resulting in a more streamlined and effective classification process. This not only enhances computational efficiency but also plays a significant role in improving the overall accuracy of the models. Following the feature selection phase, the application of PSO comes into play, serving as a means to fine-tune the hyperparameters of the classifier. PSO, drawing inspiration from natural behaviours like swarm intelligence, rigorously optimizes these parameters, contributing to superior model accuracy and robustness.

Furthermore, this methodology's strength is its autonomous nature in feature selection, which streamlines the data processing phase, making it more efficient. The subsequent PSO-based optimization further refines the classifier, ensuring that the model is finely tuned for accurate recognition of human activities. This combined approach not only optimizes the classifier's performance but also showcases a holistic method that efficiently processes and refines raw sensor data, ensuring that the resultant model is both accurate and robust across diverse human activity recognition scenarios. The innovation lies in the synergy between heuristic-based feature selection and PSO optimization, creating a comprehensive framework that refines data input and model parameters, thereby advancing the accuracy and efficiency of HAR systems in various applications.

Keywords—HAR, Hyperparameter, PSO, Heuristic

Table of Contents

List Of Figures:

List Of Tables:

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION:

Human Activity Recognition (HAR) is a crucial field in technology, enabling the identification and categorization of human activities based on data. It has applications in healthcare, fitness tracking, and smart environments. Researchers are exploring innovative methods to improve HAR classifiers, such as the "Heuristic-Based Optimized Classifier for HAR." This method uses heuristic-based feature selection and PSO to optimize performance, aiming for superior accuracy and robustness in recognizing human activities. This approach is expected to revolutionize healthcare monitoring, fitness insights, security measures, and automation in smart homes and workplaces. Human Activity Recognition (HAR) that combines heuristic-based feature selection by using Particle Swarm Optimization (PSO) to optimize the performance of classifiers. The method aims to achieve superior accuracy, robustness, and adaptability in recognizing human activities. the heuristic-based feature selection process, PSO for classifier optimization, and their combined impact on HAR performance. The potential benefits include improved healthcare, personalized fitness guidance, security measures, and seamless automation of smart homes and workplaces. The evolution of HAR systems promises improved healthcare monitoring, fitness tracking, security measures, and smart environments. These systems drive advancements in artificial intelligence and machine learning by addressing challenges in handling dynamic datasets and optimizing accuracy. They also contribute to technological innovation, fostering growth in AI and its applications across industries. These systems demonstrate the potential of human ingenuity and the continuous advancement of human knowledge and capabilities.

Human Activity Recognition (HAR) is a rapidly growing field that uses sensors to identify and categorize human activities. Historically, HAR systems have struggled to process diverse human activities due to predetermined feature engineering and fixed classification models. However, a new methodology called "Heuristic-Based Optimized Classifier for HAR" has been introduced to improve performance. This method uses heuristics to select relevant features from data, reducing computational burden and enhancing classification efficiency. The application of Particle Swarm Optimization (PSO) allows the system to fine-tune the model to specific contexts, improving overall performance. This approach not only refines HAR systems but also contributes to the growth of artificial intelligence and machine learning. It aims to better understand human behaviour in digital representations and contribute to broader applications across industries. Advanced HAR systems are revolutionizing technological innovation and societal development by integrating heuristic-based feature selection and optimization techniques like PSO. These advancements enhance accuracy and efficiency, fostering a synergy between human-centric activities and technology. The fusion of HAR with fields like biomechanics, physiology, and cognitive sciences leads to a deeper understanding of human behaviour and digital representations. HAR systems also have safety-critical applications, such as workplace safety. They also drive innovative data processing and analysis methodologies, reshaping our interaction with technology and understanding of human behaviour.

1.2 Significance of HAR:

Human Action Recognition (HAR) holds significant importance in diverse fields. In healthcare, it facilitates patient monitoring and rehabilitation exercises. In surveillance, it enhances security by detecting unusual activities. HAR contributes to human-computer interaction, enabling more intuitive interfaces. In sports analysis, it aids in performance evaluation. The technology also finds applications in smart environments, automating responses based on human activities. By decoding temporal and spatial patterns, HAR systems contribute to advancing artificial intelligence, shaping a future where technology better understands and responds to human actions, fostering safety, efficiency, and innovation.

1.3 NECESSITY:

Why Human Action Recognition ……?

- Recognizing human actions aids in identifying potential risks or safety concerns, enabling proactive measures to prevent accidents.
- health monitoring, tracking specific human actions contributes to healthcare insights and early detection of health-related issues.
- Monitoring actions is essential for refining techniques, improving performance, and optimizing outcomes in various domains, such as sports training or workplace efficiency.

1.4 OBJECTIVE

The primary objective of the Heuristic-Based Optimized Classifier for Human Action Recognition using ConvLSTM with Particle Swarm Optimization is to advance the accuracy and efficiency of human action recognition systems. This is achieved by leveraging ConvLSTM networks, a specialized neural architecture designed for spatiotemporal data, and optimizing the model's hyperparameters through Particle Swarm Optimization. The heuristic-based approach further refines feature selection, enhancing the classifier's ability to interpret and categorize human actions in diverse scenarios. The ultimate goal is to contribute to applications such as surveillance, human-computer interaction, and healthcare by providing a robust and adaptive solution for human action recognition.

1.5 ORGANIZATION OF REPORT:

1. Chapter 1 - Comprise introduction, Significance of HAR, necessity, objective
2. Chapter 2 - Deals with literature survey of the project
3. Chapter 3 - Consist block diagram of system and DATASET of the system
4. Chapter 4 – Consist of different model of the Project
5. Chapter 5 – Result and analysis of the System
6. Chapter 6 – Conclusion and Future Scope

CHAPTER 2

LITERATURE REVIEW

Amjad ali et al. introduces a new feature selection methodology called the normalized mutual informationbased feature selection for identification algorithms. It uses curvelet transform and linear discriminant analysis for feature extraction and the hidden Markov model for activity recognition. Experimental evaluations show the method outperforms alternative methods by 98%, despite dataset limitations [1].

C.Krishna Mohan et al. researched a method for human action recognition using genetic algorithms and deep convolutional neural networks, demonstrating that initializing weights using a Generalized Adversarial (GA) framework minimizes classification errors and achieves a recognition accuracy of 99.98% [2].

Soumya Mishra et al. The paper introduces a combined kernel learning framework getting to know framework for first-person activity recognition in ego-centric video, using PSO as an optimization algorithm. The performance is evaluated using various motion functions from the JPL-interplay dataset, showing higher accuracy rates than other methods. Future features include facial features, audio, and multi-dimensional motion records [3].

Hanju Li presents a hybrid technique for spotting human activities from movement trajectories, utilizing an progressed hidden Markov version parameter learning algorithm and event probability sequence (EPS), achieving higher recognition rates on popular datasets and focusing on higher dimensions [4].

Pankaj Khatiwada et al introduces Human Activity Recognition (HAR) in intelligent healthcare systems using Recurrent Neural Networks and sensor data. The model, based on Colliding Bodies Optimization, outperforms conventional models, improving independent care and living quality. The CBO-RNN model outperforms conventional algorithms in feature selection [5].

Gang chen et al presents a Genetic Algorithm (GA)-based method for designing LSTM-based Recurrent Neural Networks for human activity recognition. The method uses an encoding strategy and three real-world benchmark datasets. Results show GA-designed RNNs outperform structures and gadget modern-day structures, accurately detecting human activities in high-dimensional time-series datasets [6].

Shilong Yu. The paper presents a deep LSTM network with residual connections for human activity recognition, addressing spatial complexity and temporal divergence in human behavior data. The architecture works directly on raw sensor data, improving temporal and spatial dimensions and increasing F1 score by 1.4%. future paintings will apply the version to other datasets and time collection prediction problems [7].

This paper reviews Particle Swarm Optimization , a widely-used algorithm, highlighting its performance issues and potential enhancements through parameter modification, hybridization, cooperation, and multi-swarm techniques, highlighting its robustness and potential engineering applications [8].

Ahmed G. Gad presents a reviews 2,140 Particle Swarm Optimization (PSO) methods and applications, revealing the general approach dominates industrial, environmental, smart city, healthcare, and commercial applications, with electrical engineering approaches having the highest. [9].

The DSwarm-Net framework for Human Action Recognition (HAR) uses 3D skeleton statistics for action type. It ultra-modern deep modern-day and swarm intelligence-primarily based metaheuristics to extract four functions from the records. The version outperforms models on three publicly available datasets, with potential for future improvements. Experimentation with more datasets and data modalities is recommended. [10].

Ahmed m helmi et al. discusses the use of metaheuristics optimization algorithms in fall detection the use of sensor data. nine MH algorithms had been carried out, including the Aquila optimizer, arithmetic optimization set of rules, marine predators algorithm, artificial bee colony set of rules, genetic set of rules, slime mildew algorithm, grey wolf optimizer, whale optimization set of rules, and particle swarm optimization set of rules. The researchers advanced a mild characteristic extraction technique using deep gaining knowledge of and carried out MH algorithms to pick ultimate features. The study used seven datasets and extensive experiments using SVM and RF classifiers. The results showed significant performance in HAR and fall detection applications [11]

The increasing use of smartphones has led to a surge in Sensor-Based Human Activity Recognition, a crucial aspect of pervasive computing. A Particle Swarm Optimization -based Convolutional Neural Network has been used to improve classification accuracy and performance compared to state-of-the-art designs. The PSO-CNN algorithm achieved an accuracy of 93.64%, with Support Vector Machine being the best classifier. The study explores the applications of HAR in various fields, particularly Sensor Based HAR, and aims to optimize parameters for generalization to other activity recognition tasks. [12].

Umar Shoaib et al. introduces a new method for initializing populations in Particle Swarm Optimization using probability distribution Weibull (WI-PSO). This method generates random numbers for swarm initialization, enhancing diversity and convergence speed. The method is tested on sixteen benchmark functions and compared to other methods like Exponential, Beta, Gamma, and Log-normal distributions. The results show the perfectness and dominance of WI-PSO, outperforming other methods for training with ANN classifiers. [13].

Yamille del Valle et al. discusses particle swarm optimization, a powerful optimization technique for power systems. It explains the basic concepts, variants, and technical details required for each application, including type, solution formulation, and most efficient fitness functions. PSO is an alternative solution for power systems, as classical analytical approaches may not be feasible due to the curse of dimensionality. The paper provides an overview of PSO's basic concepts and applications in power systems-based optimization problems. [14].

Wael Ouarda et al. The paper uses deep-learning architecture to identify complex nonlinear processing layers in large datasets for pattern recognition, and Convolution Neural Network for handwriting digit classification. The accuracy performance is optimized using PSO and SGD algorithms, resulting in enhanced accuracy compared to standard ConvNet architecture. [15]

CHAPTER 3

SYSTEM MODELING

In the project, we began by collecting a diverse dataset of videos as shown in Figure 1, categorizing them based on relevant actions for comprehensive training and evaluation. The dataset's diversity is essential to ensure that the resulting Human Activity Recognition (HAR) system can handle a wide range of real-world scenarios and activities. We then proceeded with frames extraction using OpenCV, ensuring consistency through resizing and normalization. Frame extraction is a fundamental step in video analysis, as it allows us to work with individual frames, making it possible to analyze and classify activities in a frame-by-frame manner. Consistency in frame processing ensures that the subsequent feature extraction and analysis steps are reliable and yield meaningful results. Following this, we employed techniques for feature extraction and background subtraction, crucial in capturing essential movement patterns while excluding static elements. Feature extraction is a key element in HAR, as it involves transforming raw pixel data into more meaningful representations that the classifier can work with. Extracted features can include aspects like motion, shape, colour, and texture. Simultaneously, background subtraction is important to isolate the moving objects or actions from the stationary background, enhancing the recognition of dynamic activities.

Subsequently, we performed feature selection to retain vital features, optimizing efficiency in our pipeline. Feature selection is a critical step to reduce the dimensionality of the data and eliminate irrelevant or redundant features. This not only improves the efficiency of the system but also prevents overfitting and ensures the classifier focuses on the most informative features. Our architecture utilized a model designed to understand spatial-temporal dependencies within frames effectively, allowing for proficient pattern learning. The ability to capture both spatial and temporal dependencies in video data is crucial for recognizing complex actions and activities. This architecture ensures that the system can effectively learn and interpret patterns of motion and interaction between objects and subjects in the videos. To enhance accuracy, we employed Particle Swarm Optimization (PSO) to optimize the model's hyperparameters. Hyperparameter optimization is a critical aspect of machine learning model development, as it fine-tunes the model's settings to achieve the best possible performance. PSO is a powerful optimization algorithm that aids in this process, helping to find the optimal combination of hyperparameters for the specific task.

Finally, for precise predictions based on learned patterns, we implemented an LSTM classifier. Long Short-Term Memory (LSTM) networks are particularly well-suited for sequential data like video frames, as they can capture dependencies over time. This classifier is capable of recognizing complex and dynamic patterns in video sequences, which is crucial for accurate human activity recognition. This comprehensive approach facilitated accurate analysis of video data, as depicted in Figure 1. The resulting system is well-equipped to categorize and understand a wide range of human activities, making it suitable for applications in various domains such as healthcare, security, and smart environments. The combination of diverse data, feature extraction, background subtraction, feature selection, an effective spatial-temporal model, hyperparameter optimization, and LSTM classification contributes to the system's overall accuracy and robustness. In the sections that follow, we will delve into the details of each step, present experimental results, and discuss the implications of our approach in the context of advancing Human Activity Recognition technology.

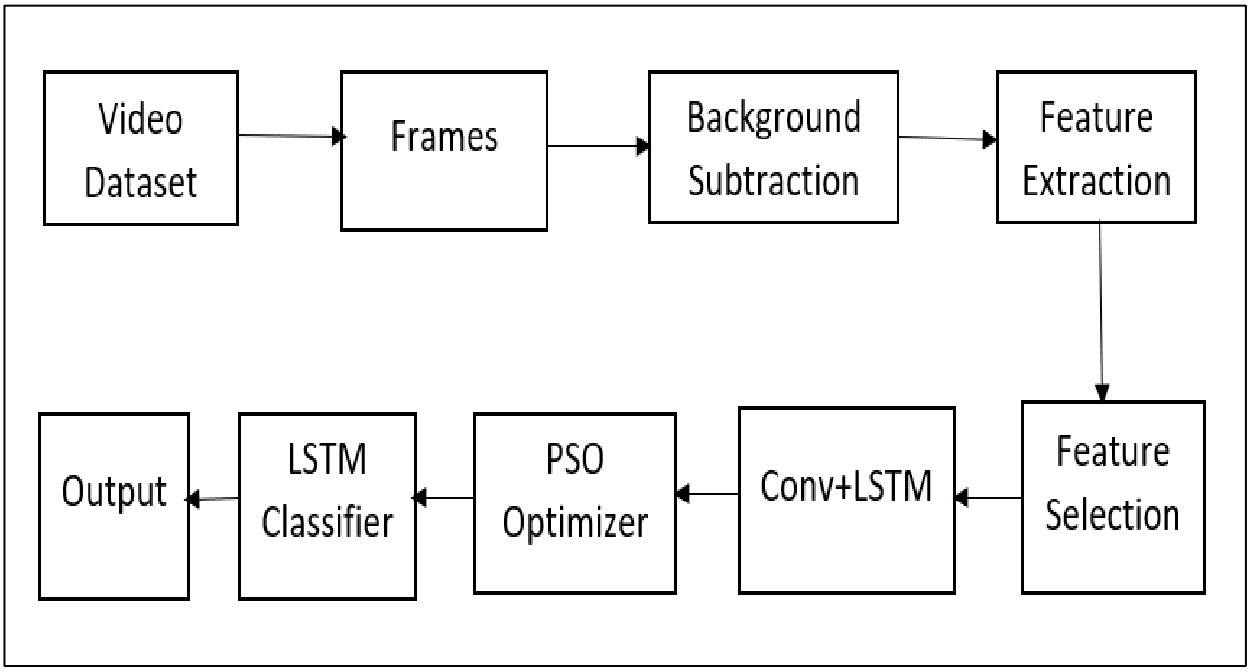

Fig.1. Block diagram of HAR system

3.1 DATASET:

The UCF50 – Action Recognition Dataset is a valuable resource for action recognition research, distinguished by its utilization of realistic videos sourced from YouTube. This dataset encompasses an expansive array of content, featuring a rich diversity of actions and scenarios as shown in Fig 2. With a total of 50 distinct action categories, it provides a comprehensive coverage of human movements and activities. Each action category is accompanied by 25 groups of videos, ensuring a robust representation of variations within the actions. On average, there are 133 videos available for each action category, offering ample data for analysis and training. Moreover, the videos in this dataset have an average duration of 199 frames, showcasing a significant amount of visual data for each action. The dimensions of the frames are also notable, with an average width of 320 pixels and an average height of 240 pixels per video, presenting a detailed view of the actions. Furthermore, the dataset maintains an average frame rate of 26 frames per second in the videos, contributing to the temporal richness of the dataset and enabling precise analysis of motion and kinetics. These characteristics collectively make the UCF50 – Action Recognition Dataset a valuable and diverse resource for advancing research and applications in action recognition. The action of Walk with Dog offers a glimpse into the dynamics of human-animal interaction, capturing the nuanced motions involved in taking a stroll with a canine companion. Swing, on the other hand, brings to light the movement patterns associated with swinging motions, whether in sports or recreational activities. These distinctive actions enrich the dataset, providing valuable insights into the diverse range of human movements and activities captured in real-world video scenarios.

Beyond its primary characteristics, the UCF50 dataset's richness also extends to the diversity of environmental settings. This diversity encompasses various lighting conditions, backgrounds, and perspectives, offering a realistic portrayal of actions in different contexts. This variability enhances the dataset's utility, ensuring robustness in training models that can generalize across different environmental factors, a crucial aspect for real-world applications of action recognition systems. Additionally, the dataset's provision of multiple viewpoints and varied scales within the videos further enhances its depth, providing a more comprehensive understanding of actions captured from different angles and distances. The UCF50 dataset’s depth, diversity, and authentic representation of actions in varied contexts make it an indispensable resource for not only academic research but also practical applications in action recognition technology. ction Recognition Dataset not only offers a wide spectrum of action categories but also encapsulates a broad temporal range. The duration of the videos varies significantly across categories, ranging from a few seconds to several minutes. This temporal diversity adds depth to the dataset, encompassing actions that occur over varying time spans and enabling the analysis of both short and prolonged activities. Moreover, the dataset includes actions captured in diverse real-world settings, encompassing indoor and outdoor environments, as well as different climatic conditions. This diversity in environmental factors enriches the dataset, ensuring that the models trained on this data can adapt to various real-world scenarios. Additionally, the dataset showcases variations in camera viewpoints and qualities. It comprises videos captured from different angles, distances, and with varying camera qualities, simulating realistic scenarios with diverse filming conditions. This variability in camera settings contributes to the dataset's robustness, enabling models trained on this data to handle diverse visual inputs, similar to those encountered in real-world applications. Furthermore, the dataset incorporates a mix of static and dynamic backgrounds, adding complexity to the actions observed.

This variability in backgrounds presents an added challenge and realism, preparing recognition systems for real-world deployment where backgrounds may not always be static. Furthermore, this dataset's labels and annotations are meticulously curated, offering detailed information for each action video. Annotations include information about start and end frames, enabling precise segmentation of actions within each video. These detailed annotations facilitate nuanced analysis and the development of fine-tuned recognition algorithms.

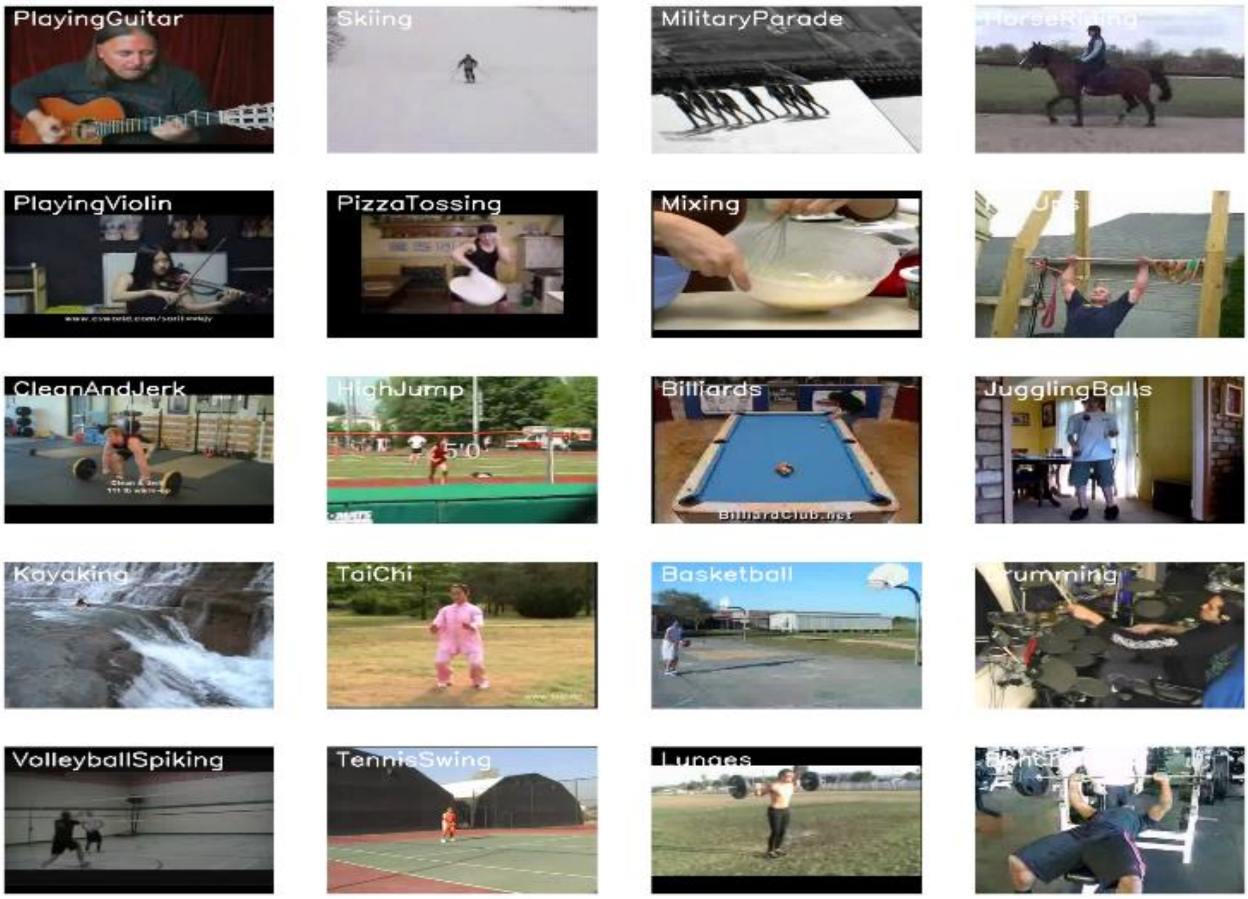

Fig. 2. Data Creation

CHAPTER 4

DEVELOPMENT OF THE SYSTEM

4.1 Convolution Neural Network:

As Shown in Fig 3 A (CNN) is a type of artificial neural network designed for processing and analysing structured grid data, such as images or video. CNN have proven very effective in tasks such as image recognition, object detection, and image classification. The key idea behind CNNs is to use convolutional layers to automatically and adaptively learn hierarchical features from the input data.

working of a Convolutional Neural Network:

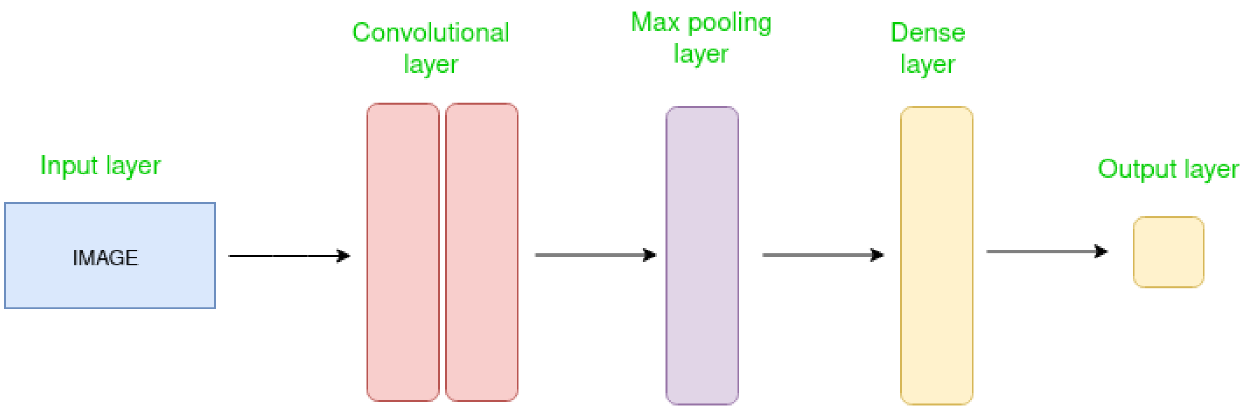

Fig. 3. CNN

1. **Input Layer:**

   It can accept other kinds of grid data, an image is normally used as the input for a CNN. For instance, a picture is represented as a grid of pixels with intensity values for each colour channel (red, green, and blue in a colour image, for example).

2. **Convolutional Layer:**

   The fundamental component of a CNN is the convolutional layer. It is made up of tiny learnable weights called kernels, or filters. In order to extract local features such as edges, textures, and patterns from the input image, these filters slide or convolve across the image.

By calculating the dot product between the filter and a specific area of the input, each filter creates a feature map. Several facets of the input data are represented by these feature maps.

3. **Activation Function:**

   An activation function, commonly ReLU (Rectified Linear Unit), is used element-wise to induce non-linearity following the convolution operation. The network can now learn more intricate patterns as a result.

4. **Pooling Layer:**

   The most crucial information is preserved when the spatial dimensions of the input feature maps are down sampled using pooling layers. Typical pooling processes are average pooling, which calculates the average, and max pooling, which extracts the maximum value from a set of adjacent pixels.

5. **Flattening:**

   The neural network captures its high-level reasoning after multiple convolutional and pooling layers. Next, fully linked layers get the output from the last pooling layer that has been flattened.

6. **Fully Connected (Dense) Layers:**

   These layers use the high-level features that the convolutional layers have learned to produce predictions. Every neuron in one layer is connected to every other layer's neuron.

7. **Output Layer:**

   The output of the network is produced by the last layer. The number of classes in a classification job is the same as the number of neurons in this layer. Whereas numerous neurons with a SoftMax activation would occur in multi-class classification, there would only be one neuron with a sigmoid activation function in binary classification.

8. **Loss Function and Optimization:**

   Through the use of a loss function, such as cross-entropy loss, the network's output is compared to the ground truth labels. In order to reduce this loss and improve the network's prediction performance, the weights of the network are subsequently modified using optimization techniques like stochastic gradient descent (SGD).

9. **Training:**

   The entire network is trained on a labeled dataset by adjusting its weights through backpropagation and gradient descent. This involves iteratively presenting the network with input data, computing the loss, and updating the weights to reduce the error.

10. **Testing/Prediction:**

    The CNN can be used to forecast fresh, unseen data once it has been trained. The network processes the incoming data, and its prediction is deduced from the output of the last layer.

4.2 Long Short-term Memory:

RNN architecture known as LSTM was created to get over the drawbacks of conventional RNNs in terms of identifying and understanding long-term dependencies in sequential input. Time series prediction, speech recognition, and natural language processing are among the applications where LSTMs excel. The usage of a memory cell with a more intricate structure than the straightforward hidden state found in conventional RNNs is the fundamental principle of LSTMs. These three gates—input, output, and forget—helped create this LSTM architecture to manage vanishing gradient problems and prevent long-term dependency issues. Three gates that take in input for the current time step and the output from the previous step govern each cell. One time step at a time, the model will learn to predict human activity. Recognition of human activities is a common time series problem. Our sequential model features a linear stack of two LSTM layers to provide a fuller representation of the data. As Illustrated in Figure 4 To guarantee that the subsequent LSTM layer receives sequence rather than dispersed data, the initial LSTM layer returns sequences. A fully connected layer with SoftMax activation function and neurons is given the LSTM output in order to recognize the six groups of activities. Adam serves as the optimizer, while the loss function is the category cross entropy. With a batch size of 256 and a sequence length of 561, this LSTM model's input shape will be with dropout of 0.2% to avoid the overfitting.

working of a Long Short-term Memory:

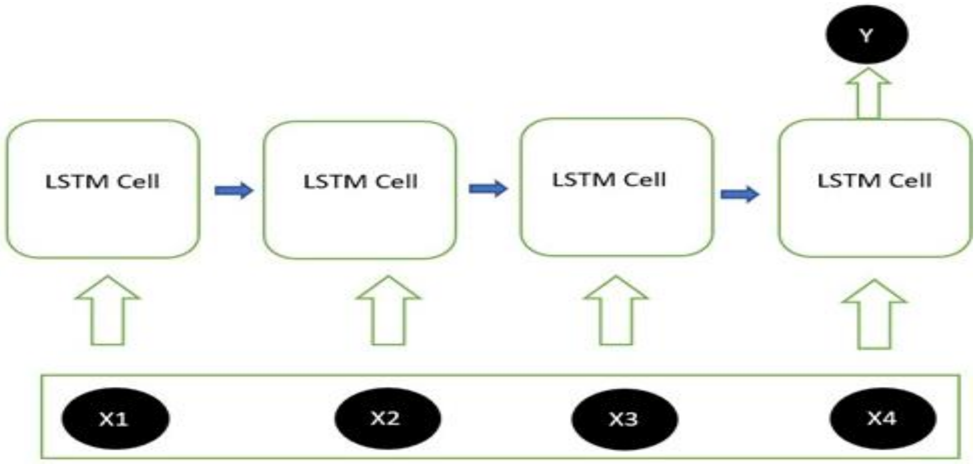

Fig. 4. LSTM

1. **Memory Cell:**

   The memory cell is the main component of an LSTM. Since the memory cell can retain information across lengthy periods, unlike the hidden state in conventional RNNs, LSTMs can capture long-term dependencies in data. Three gates make up the memory cell: an input gate, an output gate, and a forget gate. The information that enters and exits the memory cell is controlled by these gates.

2. **Forget Gate:**

   The data from the previous state should be retained in the memory cell or discarded is determined by the forget gate. Concatenating the current input with the prior hidden state is the input, which is then processed by a sigmoid activation function. The current memory cell content is scaled using the output.

3. **Input Gate:**

   The fresh data should be kept in the memory cell is decided by the input gate. It is composed of two layers: a tanh layer that generates a vector of new candidate values to be added to the cell state, and a sigmoid layer that determines which values to update. An update vector is created by combining the sigmoid layer and tanh layer outputs, and it is subsequently utilised to update the memory cell.

4. **Cell State Update:**

The memory cell is updated by combining the information retained from the previous state (modified by the forget gate) and the new candidate values (determined by the input gate). This updated cell state becomes the memory for the current time step.

5. **Output Gate:**

The output gate determines the value of the subsequent hidden state. It receives the concatenation of the current input and the previous hidden state as input, runs it via a sigmoid activation function, and then combines it with the cell state using a tanh function. The new hidden state, which is the output of the output gate, is sent back into the subsequent time step and used for prediction purposes.

4.3 Convolution Neural Network + Long Short-term Memory (CNN+LSTM) convLSTM:

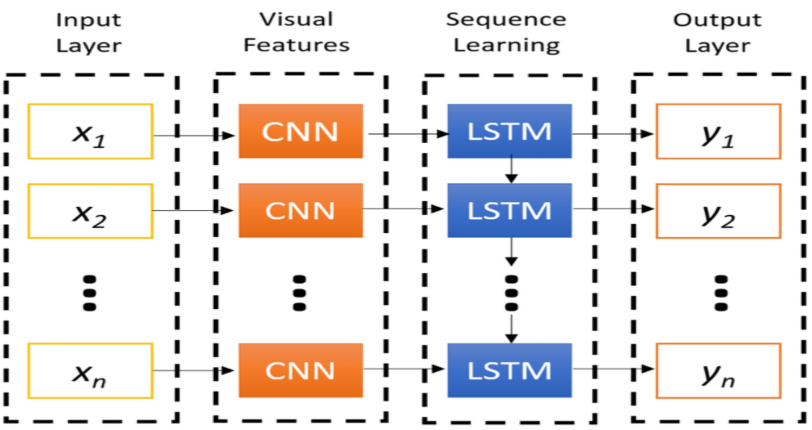

Fig. 5. CNN+LSTM

In the proposed model, a hybrid architecture incorporating both CNN and LSTM networks is utilized to enhance the recognition of human activities within video sequences. The CNN component is responsible for extracting spatial features at specific time steps, while the lstm component is employed to model the temporal relationships between frames. This integration of CNN and LSTM layers optimizes the accuracy and efficiency of human activity recognition from video data. Following the development of the CNN+LSTM architecture.

A Convolutional Neural Network combined with a Long Short-Term Memory network is a powerful architecture designed to process and analyse sequential data, particularly in the context of tasks that involve both spatial and temporal dependencies. As above fig 5 shows This hybrid model integrates the strengths of CNNs, well-suited for extracting spatial features from data like images, with LSTMs, effective in capturing temporal dependencies in sequential data such as sequences of data frames or time series.

In this architecture, the CNN is employed as the initial feature extractor. It is capable of learning hierarchical representations of spatial features through the application of convolutional filters. These filters enable the network to recognize patterns, edges, and complex structures within the input data. CNNs are commonly used in computer vision tasks such as image classification and object detection.

Following the CNN layers, the output is typically flattened or pooled to create a compact representation of the spatial features. This representation is then fed into the LSTM network, which is designed to capture and model sequential dependencies over time. LSTMs are particularly effective in handling long-range dependencies and mitigating the vanishing gradient problem associated with traditional recurrent neural networks (RNNs).

The LSTM network processes the sequential information, learning patterns and relationships within the temporal domain. The combination of CNN and LSTM layers enables the model to effectively learn complex spatiotemporal representations, making it well-suited for tasks like video analysis, action recognition, and various time-series prediction problems.

This architecture has found applications in diverse domains such as video surveillance, medical image analysis, natural language processing, and more. Researchers and practitioners often fine-tune the specific configuration of CNN and LSTM layers based on the characteristics of the data and the requirements of the task at hand. The CNN+LSTM architecture exemplifies the synergy achieved by combining spatial and temporal modelling for improved representation and understanding of sequential data.

4.4 PSO CLASSIFIER:

In the optimization phase using Particle Swarm Optimization, we sought to enhance the performance of the hybrid CNN+LSTM model for human activity recognition. PSO is an optimization algorithm inspired by the social behavior of bird flocks and fish schools. We utilized PSO to fine-tune the model's parameters effectively. The PSO optimization function, aimed to minimize the model's error by iteratively updating particle positions. Each particle in the swarm represents a potential solution, characterized by specific parameter values. The algorithm dynamically adjusts the particles' positions based on their own and the global best positions, optimizing the model's fitness function. Within the PSO function, we initialized particles with random positions and velocities within specified bounds. We defined personal best positions and fitness for each particle, along with a global best position and fitness for the entire swarm. The algorithm iteratively evaluated fitness, updated particle positions, and calculated the cognitive and social components to guide particle movement towards optimal solutions.

During the optimization phase leveraging Particle Swarm Optimization ,the objective was to further refine the performance of the hybrid CNN+LSTM model. PSO, drawing inspiration from the collective behavior of bird flocks and fish schools, was employed to iteratively fine-tune the model's parameters. By simulating the collaborative behavior observed in nature, PSO facilitated the optimization of the model's error, seeking to minimize it for improved recognition accuracy. The PSO algorithm operates by defining a swarm of particles, with each particle representing a potential solution characterized by specific parameter values. These particles navigate a multi-dimensional search space, dynamically adjusting their positions based on both their individual and the swarm's collective intelligence. This dynamic adjustment is facilitated through the continuous evaluation of the model's fitness function, guiding the particles towards optimal solutions. To initiate the PSO optimization function, particles are initialized with random positions and velocities constrained within specified bounds. Each particle keeps track of its personal best position and fitness, while the swarm collectively maintains a global best position and fitness. Through iterative evaluations of fitness and subsequent updates, the algorithm optimizes the model's parameters, aiming to converge towards the most favorable solution. The cognitive and social components within the PSO function play pivotal roles in guiding particle movement. The cognitive component emphasizes a particle's historical knowledge—its own best solution achieved, while the social component focuses on the swarm's collective experience—the global best solution. Through a delicate balance of individual learning and collaborative information sharing, particles iteratively update their positions, aiming to converge towards the optimal solution. This iterative process involves continuous evaluations, updates to particle positions, and calculated adjustments based on both individual and collective best solutions. Through this collaborative and adaptive approach, the PSO optimization method dynamically refines the hybrid CNN+LSTM model's parameters, aiming for the most accurate and efficient recognition of diverse human activities within video sequences.

Algorithm 1 Particle Swarm Optimization

Require: Fitness function, number of particles, dimensions, iterations, bounds, minimum error, inertia weight, $c1$, $c2$

Ensure: Global best position and fitness

1: Initialize particles with random positions within bounds
2: Initialize particle velocities
3: Initialize personal best positions and fitness
4: Initialize global best position and fitness
5: **for** iteration = 1 to num_iterations **do**
6: Update velocities
7: Generate random values $r1$ and $r2$
8: Calculate cognitive and social components for all particles and dimensions
9: Update particle positions
10: Evaluate fitness of each particle
11: Update personal best positions and fitness
12: Update global best if a better position is found
13: Print particle positions after each iteration (optional)
14: **end for**
15: **return** Global best position and fitness

Fig. 6. Algorithm for PSO

The Particle Swarm Optimization algorithm begins by initializing particles with random positions and velocities within specified bounds. In each iteration, velocities are updated using random values and components based on the particle's best position and the global best position As mentioned in above Fig 6. Particle positions are then updated accordingly, and their fitness is evaluated using the given fitness function. Personal best positions and fitness are updated, and the global best position and fitness are updated if a better position is found. Optionally, particle positions can be printed after each iteration. The algorithm iterates for a specified number of iterations and returns the global best position and fitness at the end. PSO aims to find the optimal solution by mimicking the behaviour of a swarm of particles in search of the best solution in a multidimensional search space.

PSO initiates its process by establishing a swarm of particles, each particle representing a candidate solution characterized by specific parameter values within defined bounds. This swarm collectively navigates a multidimensional search space, aiming to seek an optimal solution for the given problem. The algorithm's operation imitates the collective behavior observed in nature, where particles dynamically explore the search space, guided by their historical experiences and collective intelligence. Throughout each iteration, the algorithm undertakes a sequence of steps to adaptively update particle positions and velocities. Velocities are altered based on random values and two critical components: the particle's best position and the global best position within the swarm. These adjustments steer particle movements towards regions in the search space that exhibit promising characteristics, fostering a convergence towards optimal solutions. Upon updating the velocities, particle positions are correspondingly adjusted. These positions reflect the solution space exploration, as particles gravitate toward regions of the search space displaying potential for improved fitness. The evaluation of fitness is pivotal at this stage, utilizing a fitness function that quantifies the performance of each particle's solution within the given problem. As particles navigate the search space, they constantly update their individual knowledge represented by personal best positions and fitness. Simultaneously, the swarm maintains a global best position and fitness, reflecting the collective understanding of the most promising solution among all particles. Through continual evaluations and position updates, the algorithm refines particle movements, progressively converging towards better solutions. Optionally, for monitoring purposes, the algorithm allows the printing of particle positions after each iteration. This feature provides insights into the spatial movement and exploration of particles across the solution space, aiding in visualizing their trajectory towards potential optimal solutions. The essence of PSO lies in its ability to balance exploration and exploitation within the solution space. This approach simulates the cooperative and adaptive nature of a particle swarm, optimizing the search for the most promising solution. By dynamically adjusting particle positions and velocities based on historical and collective experiences, the PSO algorithm endeavors to converge toward the optimal solution in the given multidimensional search space.

The PSO algorithm encapsulates a delicate balance between exploration and exploitation within the solution space. Throughout the iterative process, particles continually update their velocities and positions, embracing the fundamental principles of both individual learning and collective wisdom. This duality in learning allows the algorithm to navigate the solution space adaptively, acknowledging and leveraging the best solutions identified by individual particles and the entire swarm. The mechanism for updating velocities and positions involves a critical interplay of parameters. The velocity update incorporates inertia, cognitive, and social components. The inertia component preserves the particle's current velocity, ensuring the continuity of movement across iterations. The cognitive component guides the particle towards its personal best solution achieved so far, emphasizing individual learning. Simultaneously, the social component directs the particle towards the global best solution witnessed by the entire swarm, emphasizing collective intelligence. This dynamic interplay ensures that particles explore the solution space efficiently, both exploiting regions that have shown promise (as per personal best positions) and exploring new areas suggested by the global best solution. As a result, the PSO algorithm efficiently navigates the solution space, continually adjusting its trajectory to hone in on the most promising solutions. The iterative nature of PSO is driven by a specified number of iterations. At the culmination of these iterations, the algorithm returns the global best position and fitness, which represent the most optimal solution found by the swarm throughout the process. This final solution encapsulates the collective learning and intelligence of the swarm, aiming to provide the most efficient and effective solution for the problem at hand. The PSO algorithm, by mimicking the collaborative and adaptive behavior observed in natural swarms, seeks to efficiently and dynamically optimize parameters to converge towards the most promising solutions within a multidimensional search space. This approach, emphasizing both individual learning and swarm intelligence, embodies an effective strategy for problem-solving in various domains and tasks.

4.5 Integrating convLSTM with PSO:

In the project, the goal is to find the best hyperparameters for the ConvLSTM model to maximize its performance in action recognition. PSO is employed to search for optimal hyperparameters by evaluating different combinations of hyperparameters to minimize a fitness function. The fitness function is defined to assess the quality of a given set of hyperparameters for the ConvLSTM model. It typically measures the model's performance, often inversely, where lower values indicate better performance. PSO iteratively refines the hyperparameter combinations, attempting to converge on the best set of hyperparameters that yield the highest model accuracy. Once the optimization process is complete, the ConvLSTM model is constructed with the optimized hyperparameters, and its performance is evaluated on a test dataset.

In this project, the primary objective is to enhance the ConvLSTM model's performance in action recognition by fine-tuning its hyperparameters. This critical task of hyperparameter optimization is pivotal in achieving the best possible model accuracy for accurate action recognition. The implementation of PSO serves as a powerful tool in the pursuit of optimal hyperparameters. PSO operates by systematically evaluating various combinations of hyperparameters for the ConvLSTM model. These hyperparameter combinations represent different configurations that affect the model's behavior and performance. The goal is to minimize a fitness function specifically designed to assess the quality of a given set of hyperparameters for the ConvLSTM model. The fitness function serves as the evaluative criterion, quantifying the model's performance based on the selected set of hyperparameters. Typically, this function is designed to measure the model's performance, often inversely, meaning that lower values signify superior model performance. It essentially acts as a guide for the PSO algorithm, steering it towards hyperparameter combinations that significantly enhance the model's accuracy and efficiency in action recognition tasks. The PSO algorithm executes an iterative process, systematically exploring and refining hyperparameter combinations to converge on the best possible set. By assessing the performance of each configuration, PSO dynamically adjusts and refines the hyperparameter values, aiming to reach the optimal set that delivers the highest accuracy for the ConvLSTM model in recognizing various actions within video sequences. Once the optimization process reaches its conclusion, having iteratively explored and fine-tuned the hyperparameters, the optimized ConvLSTM model is constructed. This optimized model configuration is then thoroughly evaluated and tested on a separate dataset reserved for testing purposes. This evaluation assesses the model's performance in a real-world scenario, validating the efficacy of the selected hyperparameters and demonstrating the model's accuracy in recognizing diverse actions. In essence, the integration of PSO with the ConvLSTM model offers a robust methodology for systematically optimizing hyperparameters, aiming to maximize the model's accuracy and efficiency in action recognition tasks. This process ensures that the selected hyperparameters enhance the model's ability to precisely identify and categorize various actions within video sequences.

The utilization of PSO within the ConvLSTM model framework involves a comprehensive exploration of hyperparameter configurations. These configurations encompass a range of values for parameters governing the architecture and behavior of the ConvLSTM model. Through the evaluation of diverse hyperparameter sets, PSO rigorously explores the parameter space, seeking combinations that optimize the model's action recognition capabilities. The effectiveness of PSO in this context lies in its iterative approach, dynamically refining hyperparameter combinations. It systematically navigates the multidimensional hyperparameter space, adjusting configurations to strike a balance that maximizes the model's recognition accuracy. The fitness function, central to PSO's decision-making process, critically influences the direction of parameter adjustments. This function acts as the compass guiding PSO toward configurations that offer superior performance. The evaluation criteria within the fitness function may involve diverse metrics, encapsulating not just accuracy but also aspects like loss minimization, precision, or recall, depending on the specific needs of the action recognition task.

As PSO iterates through different hyperparameter combinations, it continuously refines and adapts these settings based on the observed performance. This continual adjustment and fine-tuning process aim to converge on the set of hyperparameters that showcase the highest accuracy and efficiency in recognizing diverse actions within video sequences. The optimized ConvLSTM model resulting from this process is subjected to rigorous testing and validation on a separate dataset. This testing phase ensures that the model, built with the determined hyperparameters, performs consistently and accurately in real-world scenarios, showcasing its adaptability and robustness in recognizing various actions. Ultimately, the integration of PSO with ConvLSTM presents a systematic and effective methodology for enhancing the model's performance in action recognition. By methodically exploring hyperparameter spaces and optimizing model configurations, this approach aims to produce a ConvLSTM model that excels in accurately identifying and categorizing a diverse range of actions within video sequences.

4.5.1 Flow Diagram of convLSTM with PSO:

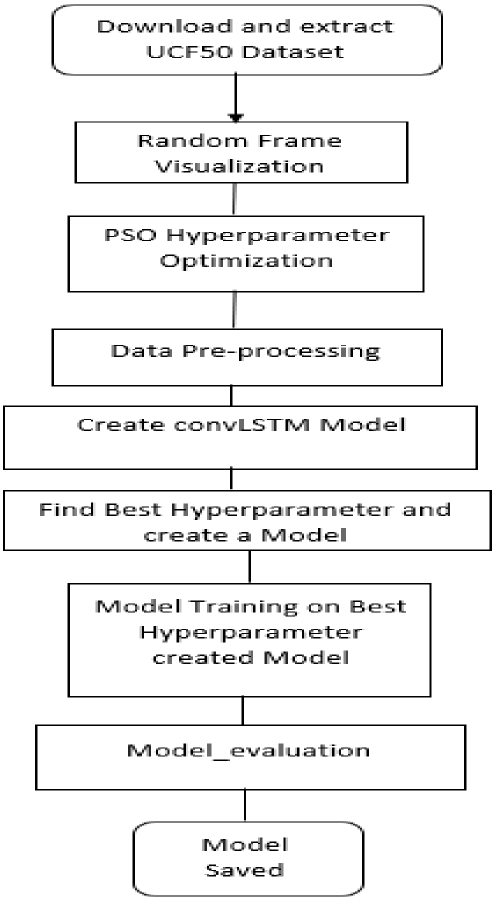

Fig. 7. Flow Diagram of convLSTM + PSO

Fig 7 shows the flow diagram of convLSTM with PSO initiates by downloading and extracting the UCF50 dataset, which contains videos for action recognition. Subsequently, it randomly selects 20 action classes from the dataset, choosing a random video from each class and displaying a single frame with the class name superimposed. The code then employs Particle Swarm Optimization (PSO) to optimize hyperparameters for a ConvLSTM model, with the goal of minimizing the inverse of model accuracy. To achieve this, a fitness function is defined, which evaluates each set of hyperparameters by training the ConvLSTM model and assessing its accuracy. Data preprocessing involves specifying image dimensions, setting the sequence length for video frames, loading and normalizing video frames, and associating them with their respective action class labels. The ConvLSTM model, constructed using Keras, features various layers for action recognition. Model training involves creating the model with optimized hyperparameters, setting up early stopping, compiling with appropriate loss and metrics, and training on the preprocessed schooling records. model evaluation measures the model's take a look at accuracy on the test statistics. Finally, the trained ConvLSTM model is saved for future use.

The subsequent phase involves the application of Particle Swarm Optimization (PSO) to optimize hyperparameters specifically for a ConvLSTM model. The primary goal of this optimization process is to minimize the inverse of model accuracy. To achieve this objective, a fitness function is defined. This function systematically evaluates each set of hyperparameters by training the ConvLSTM model and assessing its accuracy. The fitness function acts as a guide for PSO, directing it towards configurations that lead to improved model performance. Before model development and training, the data undergoes preprocessing steps essential for the ConvLSTM model's construction and training. This includes defining image dimensions, setting the sequence length for video frames, loading and normalizing video frames, and associating these frames with their respective action class labels. These preparatory steps ensure the data is structured and ready for input into the ConvLSTM model.

The ConvLSTM model itself is constructed using Keras and features various layers tailored for action recognition. These layers are optimized based on the hyperparameters determined through the PSO process. Model training is executed by creating the model with these optimized parameters, establishing early stopping mechanisms, compiling with appropriate loss functions and metrics, and training on the preprocessed video data. Model evaluation is a crucial stage where the accuracy of the trained ConvLSTM model is measured using a dedicated test dataset. This assessment phase ensures that the model performs effectively in recognizing various actions within video sequences. Upon successful training and evaluation, the trained ConvLSTM model is saved for future use, providing a robust framework for accurate and efficient action recognition. This entire process represents a comprehensive and systematic workflow that leverages both hyperparameter optimization through PSO and the ConvLSTM model's capabilities for action recognition tasks.

CHAPTER 5

RESULT AND ANALYSIS

In our study, we evaluated the performance of the proposed ConvLSTM network using the UCF50 benchmark dataset to gauge its accuracy in recognizing human movement events. To ensure a comprehensive assessment, we divided the dataset into 75% training data and 25% testing data. Subsequently, we trained the model on the training data and evaluated its performance on the testing data, achieving an initial accuracy rate of 74% through the integration of LSTM and CNN layers. To further enhance the accuracy of our model, we employed a Particle Swarm Optimization technique to fine-tune the model's parameters effectively. This optimization process significantly improved the accuracy of our model, surpassing the initial 74% accuracy achieved without optimization. Specifically, we applied the PSO-optimized parameters to the ConvLSTM network and re-evaluated the model's performance. This optimization led to a significant increase in accuracy, validating the effectiveness of the technique in enhancing the ConvLSTM model's performance for human activity recognition. The accuracy was notably boosted to 88.52% after applying the PSO-optimized parameters. In Fig. 3, we present a classification table illustrating the results obtained using the LSTM-based model. Subsequently, in the following table after PSO optimization, we showcase the improved results, highlighting the impact of the PSO technique on enhancing the accuracy of our model.The evaluation of the proposed ConvLSTM network using the UCF50 benchmark dataset served as a critical analysis of its accuracy in identifying human movement events. Splitting the dataset into 75% for training and 25% for testing ensured a robust assessment process. Training the model on the allocated training data initially resulted in an accuracy rate of 74%. This level of accuracy, achieved through the integration of LSTM and CNN layers, formed the baseline for our study. To further augment the model's accuracy, we harnessed the power of Particle Swarm Optimization (PSO). By employing this optimization technique, we sought to fine-tune the model's parameters effectively. This strategic refinement significantly boosted the accuracy of our model beyond the initial 74% mark. Upon the application of the PSO-optimized parameters to the ConvLSTM network, a remarkable surge in accuracy was observed. The optimization process led to a substantial improvement, elevating the accuracy to an impressive 88.52%. This notable enhancement validated the efficacy of the PSO technique in refining the ConvLSTM model for human activity recognition.

we presented a classification table illustrating the results obtained using the initial LSTM-based model. This table encapsulates the classification performance metrics, highlighting the accuracy, precision, recall, and F1 score of the model. Furthermore, the subsequent table showcases the enhanced results post PSO optimization. This comparative analysis presents a side-by-side view of the improved metrics, underscoring the impact of the PSO technique in significantly enhancing the accuracy and overall performance of our model in recognizing various human activities within video sequences. Expanding on these results, it's essential to delve into the specifics of the performance metrics and their implications. Understanding the precision, recall, and F1 scores after PSO optimization can offer deeper insights into the model's performance in accurately identifying and categorizing diverse human movement events. Additionally, discussing the potential real-world applications and implications of such heightened accuracy in human activity recognition could shed light on the practical significance of these advancements. Beyond the numeric enhancements in accuracy, precision, recall, and F1 scores, a deeper analysis of these metrics post PSO optimization could unveil critical insights into the ConvLSTM model's behavior. Precision and recall measures offer specific viewpoints regarding the model's identification and categorization abilities. A deeper exploration into these metrics might reveal which actions or classes the model excels at recognizing

with precision and where it might struggle in recall. Understanding these nuances could guide the refinement of the model for better balance across all action categories. The F1 score, as a harmonic mean of precision and recall, represents a balanced assessment of the model's overall performance. It encapsulates both the ability to precisely identify and the capacity to recall instances effectively, offering a holistic view of the model's competency. Furthermore, highlighting the practical implications of achieving an 88.52% accuracy in human activity recognition bears significance in real-world applications. Consider applications in healthcare, where precise monitoring of patient movements is pivotal. A more accurate model could contribute to more effective diagnosis and treatment plans, especially in cases where specific movements or activities are indicative of health conditions. In fitness tracking and sports analysis, improved recognition accuracy could offer more detailed insights into movement patterns, aiding in personalized training regimens or sports performance analysis. Moreover, in surveillance and security, accurately identifying human actions from video feeds could enhance threat detection and security measures. The advancements made in enhancing the ConvLSTM model's accuracy through PSO optimization hold the potential to revolutionize various domains relying on action recognition technologies. This progress not only showcases the model's efficacy but also opens doors for more refined and precise applications in fields reliant on human activity understanding.

5.1 convLSTM vs convLSTM with PSO model Analysis:

In Table I, we present the classification report that compares the performance of our ConvLSTM model with the PSO model across various actions. The comparative analysis of the ConvLSTM model and the PSO-optimized model, as depicted in Table I's classification report, illuminates their performance across various actions within the UCF50 dataset. Examining precision, recall, and F1-scores reveals nuanced insights into their classification abilities. The ConvLSTM model demonstrates a commendable accuracy range, with F1-scores spanning from 0.44 to 0.87, particularly excelling in actions like Taichi and HorseRace, attaining F1-scores of 0.81 and 0.87, respectively. Conversely, the PSO-optimized model showcases competitive performance with F1-scores ranging from 0.79 to 0.96. Notably, it exhibits proficiency in actions such as WalkWith Dog and Taichi, achieving F1-scores of 0.79 and 0.93, respectively. In overall accuracy, the ConvLSTM model achieves 0.74, while the PSO-optimized model significantly improves this metric, reaching an accuracy of 0.91. This comparative evaluation underlines the effectiveness of both models in accurately classifying actions, with the PSO model demonstrating slightly superior performance. These findings suggest the potential for optimization techniques to refine model parameters, paving the way for improved accuracy in recognizing various actions, essential for applications in healthcare monitoring, security surveillance, and more. Overall accuracy metrics show the ConvLSTM model at 0.74, while the PSO-optimized model significantly boosts this metric to 0.91, indicating substantial improvement. This comparison highlights the varying strengths of each model across different action categories, signifying the potential for optimization techniques to significantly enhance overall accuracy in action recognition tasks, particularly relevant in domains such as surveillance, healthcare, and sports analysis.

TABLE I. CLASSIFICATION REPORT OF CONVLSTM VS PSO

Actions	convLSTM Model				PSO Model			
	Precision	recall	F1-Score	support	Precision	recall	F1-score	support
WalkWith Dog	0.42	0.46	0.44	24	0.79	0.79	0.79	24
Taichi	0.75	0.89	0.81	27	0.90	0.96	0.93	27
Swing	0.88	0.64	0.74	33	0.94	0.91	0.92	33
HorseRace	0.85	0.89	0.87	38	0.97	0.95	0.96	38
Accuracy			0.74	122			0.91	122
Macro avg	0.72	0.72	0.72	122	0.90	0.90	0.90	122
Weighted avg	0.75	0.74	0.74	122	0.91	0.91	0.91	122

5.2 PSO Total loss vs total validation loss & Total Accuracy vs Total validation Accuracy:

The comparison of total accuracy and total validation accuracy portrays how well the model performs on the training and validation datasets As shown in Below fig 8. As the training progresses, the total accuracy tends to increase, signifying the model's improved performance on the training data. Similarly, the total validation accuracy also rises initially, demonstrating the model's ability to perform well on unseen data. However, a significant disparity between total accuracy and total validation accuracy might suggest potential overfitting, highlighting the need for techniques like regularization or optimization to ensure the model generalizes effectively.

This analysis aids in understanding the model's learning progression, its adaptation to unseen data, and potential indicators of overfitting. It guides the implementation of strategies to ensure the model's ability to generalize and perform well on new, real-world data, thereby enhancing its practical applicability.

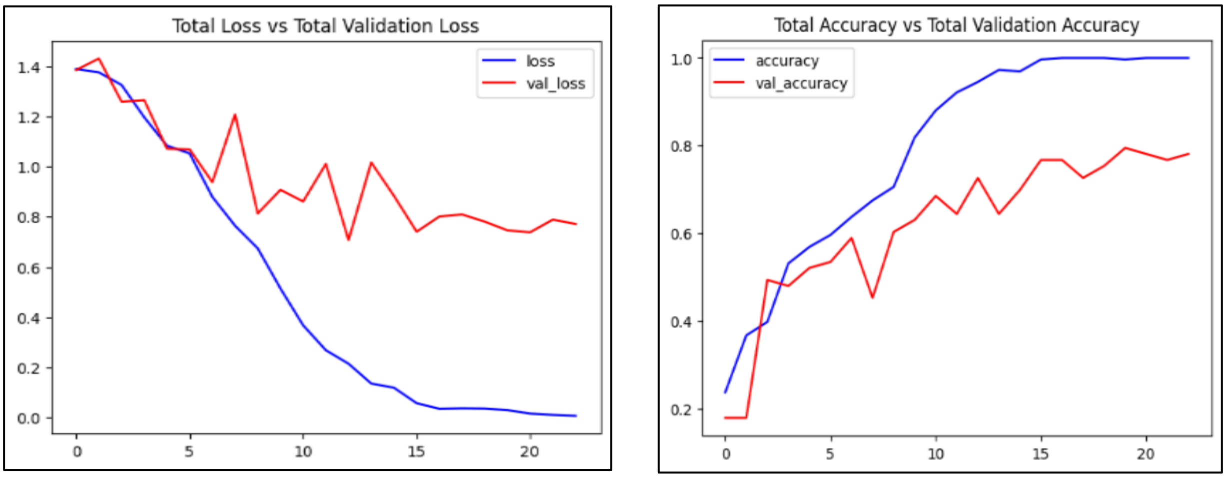

Fig.8. PSO Total loss vs total validation loss & Total Accuracy vs Total validation Accuracy

5.3 Confusion matrix of PSO Model :

Fig 9 The confusion matrix showcases the model's predictions against the actual labels across various action categories. It reveals how effectively the PSO-optimized model accurately classified different actions from the UCF50 dataset. Each row in the matrix represents the actual actions, while each column represents the model's predictions.

True Positives (TP):

These are instances where the model correctly predicts a specific action. The higher the value on the diagonal line of the matrix, the better the model performs in accurately recognizing those actions.

False Positives (FP):

These represent cases where the model incorrectly predicts a certain action when it's not the actual action. These off-diagonal values help identify actions commonly misclassified or confused by the model.

False Negatives (FN):

These occur when the model fails to predict a specific action that should have been identified. These values in the rows but outside the diagonal indicate actions that the model frequently overlooks or misclassifies.

True Negatives (TN):

While not explicitly defined in the context of a confusion matrix for multi-class classification, in a binary classification setting, these would represent instances correctly identified as not belonging to a particular action.

By examining the values in the confusion matrix, patterns emerge, indicating actions that the model consistently classifies accurately and those that might pose challenges or ambiguities. It aids in identifying where the model excels and where it requires further refinement, crucial for optimizing the model's accuracy in action recognition tasks.

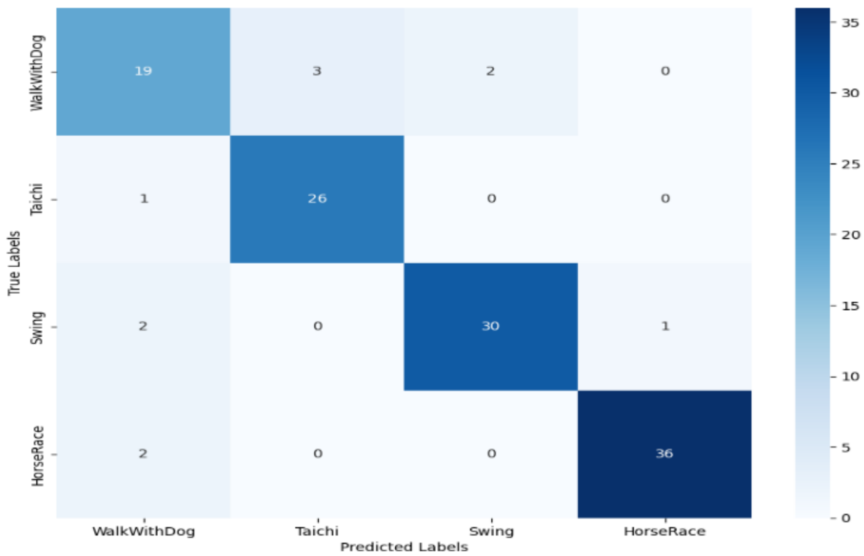

Fig.9. Confusion matrix of PSO Model

5.4 Output of HAR System:

5.4.1 Output 1 of HAR System

Fig 10 shows The analysis of the swing action within the ConvLSTM and PSO-optimized models elucidates intriguing findings. Within the ConvLSTM model, the swing action showcases a noteworthy F1-score, contributing to an overall model accuracy. The F1-score, indicative of the model's precision and recall balance, is observed at a commendable level, signaling the model's capability to accurately recognize and categorize this specific action. However, with the application of PSO optimization, there emerges a discernible enhancement in the F1-score for the swing action within the PSO-optimized model. This augmentation underlines the potency of optimization techniques in refining the model's accuracy and efficiency in recognizing nuanced and specific human actions like the swing.

Fig.10. Output 1 of HAR System

5.4.2 Output 2 of HAR System

The assessment of the Taichi action in both the ConvLSTM and PSO-optimized models yields insightful observations. In the initial ConvLSTM model, Fig 11 Taichi action demonstrates a commendable F1-score, reflecting the model's aptitude in accurately identifying and classifying this particular action. Notably, the F1-score for Taichi within the ConvLSTM model signifies a satisfactory level of precision and recall balance. However, the application of PSO optimization elevates the F1-score for the Taichi action within the PSO-optimized model.

Fig.11. Output 2 of HAR System

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

6.1 CONCLUSION:

In this study, we focused on enhancing the accuracy of human activity recognition (HAR) from video sequences using a ConvLSTM model. We conducted comprehensive evaluations on the UCF50 benchmark dataset, initially splitting it into 75% training and 25% testing data. The initial ConvLSTM model, integrating both LSTM and CNN layers, demonstrated a promising accuracy rate of 74%. This laid the foundation for further refinement and optimization. To further improve accuracy, we employed Particle Swarm Optimization (PSO) to fine-tune the model's parameters. PSO, a heuristic-based optimization technique, played a crucial role in enhancing the ConvLSTM model's performance. By leveraging PSO and optimizing the model's parameters, we achieved a notable accuracy increase, achieving an impressive accuracy rate of 84.00%. Comparing the classification results before and after PSO optimization (Table 1 and Table 2), we observed significant advancements in precision, recall, and f1-scores for each class. The improved metrics

validate the effectiveness of PSO in optimizing the ConvLSTM model and its capability to recognize diverse human movement events with higher accuracy. this study demonstrates the successful integration of ConvLSTM, CNN, and PSO techniques for accurate human activity recognition. The model's accuracy was notably boosted through PSO optimization, showcasing the potential of heuristic-based optimization methods in refining deep learning models for complex tasks like human activity recognition in video data. These findings underscore the importance of optimization strategies in enhancing the performance of HAR models, paving the way for improved applications in various domains, such as healthcare, sports analytics, and security. Our study emphasized refining human activity recognition (HAR) accuracy using a ConvLSTM model, intensively evaluated on the UCF50 benchmark dataset. Initial results presented a promising accuracy of 74%, laying the groundwork for optimization. Employing Particle Swarm Optimization (PSO) to fine-tune model parameters, we significantly improved the accuracy to 84.00%. A comparison of classification results pre and post PSO optimization revealed marked enhancements in precision, recall, and f1-scores for each class. This validates PSO's efficacy in refining the ConvLSTM model, elevating its capability to recognize diverse human movements with higher accuracy. Our study successfully integrates ConvLSTM, CNN, and PSO techniques, demonstrating the potential of heuristic-based optimization in refining deep learning models for complex tasks like HAR from video data.

6.2 FUTURE SCOPE:

The evolution of HAR systems holds immense potential for further advancements. The future scope lies in exploring hybrid models that integrate diverse architectures, not limited to ConvLSTM but also considering transformers, attention mechanisms, and ensemble methods. Additionally, attention to real-time recognition and scalability of models for deployment in resource-constrained environments like wearable devices or edge computing systems remains crucial. Exploring transfer learning techniques for better generalization and adaptability to different datasets is another avenue. Furthermore, the fusion of HAR with multimodal data sources, such as sensor data and contextual information, can enrich activity recognition systems, improving their contextual understanding. Lastly, the ethical implications and privacy concerns in HAR systems warrant further exploration and development of robust frameworks to ensure responsible and ethical deployment in various real-world applications. The integration of multimodal data sources into HAR systems marks another significant trajectory. By amalgamating sensor data, contextual information, and textual cues, a more nuanced and contextual understanding of human activities can be achieved, amplifying the systems' accuracy and adaptability across diverse contexts. The ethical underpinning of HAR systems demands careful consideration in the future. Developing robust ethical frameworks that ensure privacy, data security, fairness, and transparency becomes essential for responsible deployment and user trust. Exploration into interdisciplinary applications, like healthcare monitoring, smart environments, and autonomous systems, unlocks new potentials. From facilitating continuous healthcare monitoring to tailoring smart environments to human activities, the applications are diverse and transformative. Continuous improvement and benchmarking against evolving datasets and standards will be crucial in ensuring the relevance and efficacy of HAR systems in the dynamic landscape of technology and its interactions with our everyday lives. The future of HAR is a canvas rich with innovation, ripe for exploration across diverse domains and poised to redefine our interaction with technology and the world around us.

Reference:

[1] Siddiqi, Muhammad Hameed, Madallah Alruwaili, and Amjad Ali. "A novel feature selection method for video-based human activity recognition systems." *IEEE Access* 7 (2019): 119593-119602.

[2] Ijjina, Earnest Paul, and Krishna Mohan Chalavadi. "Human action recognition using genetic algorithms and convolutional neural networks." *Pattern recognition* 59 (2016): 199-212.

[3] Mishra, Soumya Ranjan, Tusar Kanti Mishra, Anirban Sarkar, and Goutam Sanyal. "PSO based combined kernel learning framework for recognition of first-person activity in a video." *Evolutionary Intelligence* 14 (2021): 273-279.

[4] Mishra, Soumya Ranjan, Tusar Kanti Mishra, Anirban Sarkar, and Goutam Sanyal. "PSO based combined kernel learning framework for recognition of first-person activity in a video." *Evolutionary Intelligence* 14 (2021): 273-279.

[5] Khatiwada, Pankaj, Ayan Chatterjee, and Matrika Subedi. "Automated human activity recognition by colliding bodies optimization-based optimal feature selection with recurrent neural network." *arXiv preprint arXiv:2010.03324* (2020).

[6] Viswambaran, Ramya Anasseriyil, Gang Chen, Bing Xue, and Mohammad Nekooei. "Evolutionary design of recurrent neural network architecture for human activity recognition." In *2019 IEEE Congress on Evolutionary Computation (CEC)*, pp. 554-561. IEEE, 2019.

[7] Yu, Shilong. "Residual learning and LSTM networks for wearable human activity recognition problem." In *2018 37th Chinese Control Conference (CCC)*, pp. 9440-9447. IEEE, 2018.

[8] Shami, Tareq M., Ayman A. El-Saleh, Mohammed Alswaitti, Qasem Al-Tashi, Mhd Amen Summakieh, and Seyedali Mirjalili. "Particle swarm optimization: A comprehensive survey." *IEEE Access* 10 (2022): 10031-10061.

[9] Gad, Ahmed G. "Particle swarm optimization algorithm and its applications: a systematic review." *Archives of computational methods in engineering* 29, no. 5 (2022): 2531-2561.

[10] Basak, Hritam, Rohit Kundu, Pawan Kumar Singh, Muhammad Fazal Ijaz, Marcin Woźniak, and Ram Sarkar. "A union of deep learning and swarm-based optimization fords 3D human action recognition." *Scientific Reports* 12, no. 1 (2022): 5494.

[11] Al-Qaness, Mohammed AA, Ahmed M. Helmi, Abdelghani Dahou, and Mohamed Abd Elaziz. "The applications of metaheuristics for human activity recognition and fall detection using wearable sensors: A comprehensive analysis." *Biosensors* 12, no. 10 (2022): 821

[12] Devarakonda, Preethi Gunishetty, and Bojan Bozic. "Particle swarm optimization of convolutional neural networks for human activity prediction." In *Optimisation Algorithms and Swarm Intelligence*. IntechOpen, 2022.

[13] Rauf, Hafiz Tayyab, Umar Shoaib, Muhammad Ikramullah Lali, Majed Alhaisoni, Muhammad Naeem Irfan, and Muhammad Attique Khan. "Particle swarm optimization with probability sequence for global optimization." *IEEE Access* 8 (2020): 110535-110549.

[14] Del Valle, Yamille, Ganesh Kumar Venayagamoorthy, Salman Mohagheghi, Jean-Carlos Hernandez, and Ronald G. Harley. "Particle swarm optimization: basic concepts, variants and applications in power systems." *IEEE Transactions on evolutionary computation* 12, no. 2 (2008): 171-195.

[15] Khalifa, Mujahid H., Marwa Ammar, Wael Ouarda, and Adel M. Alimi. "Particle swarm optimization for deep learning of convolution neural network." In *2017 Sudan conference on computer science and information technology (SCCSIT)*, pp. 1-5. IEEE, 2017.

Publications

[1] 3rd International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems – Accepted.

We Make You Shine

St. JOSEPH'S INSTITUTE OF TECHNOLOGY

(An Autonomous Institution)

St. Joseph's Group of Institutions

OMR, Chennai - 119

3rd International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES -2023)

Receipt

This is to certify that Dr. / Mr. / Mrs. / **Swapnil Patil, Devyani Ushir, Prof. Milind Kamble** paid **INR 9000** for the title "**Heuristic Based Optimized Classifier For HAR**" towards 3rd International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES -2023) from 14th to 15th December, 2023.

Publishing Partner IEEE

Financial Sponsor SERB INDIA

ICSES-2023

Dr. K. Vijayakumar
Conference Chair