International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 06th Feb 2023. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=ISSUE-02

DOI: 10.48047/IJIEMR/V12/ISSUE 02/11

Title **Automatic Human Action Recognition Model Based on Adaptive Long Short Term Memory**

Volume 12, Issue 02, Pages: 86-93

Paper Authors
Badhagouni Suresh Kumar, Dr S.Viswanadha Raju, Dr H.Venkateswara Reddy

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Automatic Human Action Recognition Model Based on Adaptive Long Short Term Memory

Badhagouni Suresh Kumar[1], Dr S.Viswanadha Raju[2], Dr H.Venkateswara Reddy[3]
Department of CSE, Vardhaman College of Engineering. Hyderabad, India[1,3]
Deparment of CSE, JNTUH College of Engineering ,Nachupally[2]

Abstract
In computer vision, human action recognition is a vital research area. Its applications include the classification of structures that include surveillance structures, patient monitoring structures, and interactions between individuals and electronic devices, for example, human-computer interfaces. Most of these applications require automated recognition of abnormal or contradictory action states created by various direct (or nuclear) actions of individuals. This paper delas with efficient human action recognition model. Mainly three approaches are proposed model namely, feature extraction, key frame selection and classification. In the initial stage key frames extraction is done from a given input video, using Structural Similarity (SSIM) measure. In the next step shape, coverage factor, and Space-Time Interest (STI) points are extracted from the key frames. In the last step all the extracted features are fed into ALSTM classifier to categorise different activities of human in given input video. The proposed ALSTM is designed based on combination of LSTM and adaptive seagull optimization (ASO) algorithm. The proposed approach is experimented on UCF 101 dataset and shows the better result based on precision, recall and F-Measure.

Keywords: Human action recognition, adaptive golden eagle optimization, ALSTM, Space-time interest, coverage factor, shape and SSIM.

Introduction

Human Activity Recognition system deals with human activities. It is integral to human-to-human interaction and personal relationships. It is not so easy to identify actions because input comprises the information of a person's identity, physical appearance, pose and etc.. Humans can recognize another person's action but how does a system can understand recognize the actions of a human. This is major challenge in computer vision applications [1]. Now-a-days, the field of human action recognition is gaining more and more attention because of opportunities in various fields like health care, surveillance, entertainment and etc.. [2]. Automatic detection of abnormalities in the surveillance environment can be used to alert potential associated power. For example, the automated report of a person travelling with a bag in surveillance environments such as airports and stations. In an entertainment environment, it's like automatically recognizing players during a game, creating an avatar on the computer for the player to play the game on. Several types of functional authentication systems are required in lot of applications like surveillance systems, human-system interaction, depiction of human activities. Different type of Events like "walking", "running" etc. arise very obviously in everyday life and are comparatively recognizable [3].

On the other hand, it is very difficult to identify more complex functions like as "apple peeling". Complex functions can be broken down into other simple functions that are generally easier to identify. This has very challenging because of some issues like background mess, incomplete obstacles, variations in scale, vision, illumination and presence, and dimensions. Classifying a person's actions with negligeable error is an exciting task [4]. To succeed these issues requires a task consisting of three components: background minus, human tracking, and human action and object detection, in which the computer can localize a human function into an image [5, 6]. Human actions are categorized according to their complexity into gestures, nuclear activities, human and object

communication and human to human communication, group actions and actions. On the basis of a comprehensive review of the literature there are two types of computer vision HAR-based approaches available those are: traditional training and classification methods developed based on features and approaches which learns the features deep learning based methods can automatically learn features from source data and follow a classifier that is generally trained for action recognition [7].

Baskaran and Saroja [8] had introduced recognition of human activity using machine learning algorithms. In this method, human action recognition was analysed using various machine learning techniques. For analysis, various wearable and non-wearable sensors are used for human activity evaluation. Moreover, Abhay Gupta *et al.* [9] had introduced post estimation and machine learning algorithm based human action recognition. In this approach, they are proposed the classification and recognition of human action using pose skeleton in the images of the person. The results showed that multiple logistic regression, SVM and FR methods achieved higher accuracies of 80.72%, 80.43%, and 80.75%. Similarly, Jaouedi *et al.* [10] had introduced recognition of human action using a new hybrid deep learning model. In this approach, the authors proposed a new hybrid deep learning model using Gated Recurrent Neural Networks, GMM and KF for human action recognition. The proposed hybrid approach achieved improved accuracy compared to other methods.

Marinho *et al.* [11] had introduced machine learning techniques based human action recognition. In this approach, the authors evaluated the human action recognition of the feature selection. From the results, the support vector machine and MLM achieved 99.2% accuracy and the new feature selection model achieved 98.1% accuracy.Rabbi *et al.* [12] had introduced machine learning based human action analysis and recognition from smartphones. In this approach, the authors proposed machine learning models to recognize human activities and then the performance was compared using various algorithms. In result, the support vector machine performed better than other methods an average accuracy 96.33%.

Proposed Methodology

The aim of this paper is to effectively identify the human action which is used to avoid the abnormal activities. To achieve this objective ALSTM classifier is presented. The ALSTM is a combination of LSTM and ASO. Key frame extraction, feature extraction, and classification are the three stages of the proposed approach. At first, the main frames are extracted from the input video using SSIM measurements. After the key-frame extraction, the features of each frame are extracted. In this paper, three types of features are extracted namely coverage factor, shape, and STI points. Then, the extracted features are fed to the ALSTM classifier to classify human activities. The overall structure of the proposed methodology is given in figure 1.

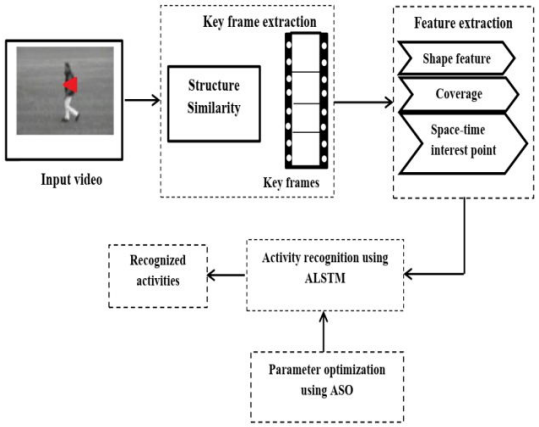


Figure 1: Overall structure of proposed methodology

Key Frame extraction

Key frame extraction is a significant process for human activity recognition. In this paper, SSIM method is used for key frame extraction process. Here, initially, the videos are changed into n-number of frames. Key-frames are frames that contain main contents of the video. Using key frame selection process, we can reduce the memory space. In order to identify human activities effectively, key frames need to be extracted from videos. By using SSIM measurements, the main

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

frame is extracted.The process involved in key-frame extraction is described below;

Assume input video V which contains n number of frames. The sample video sequences is expressed as follows;

$$V = \{F_1, F_2, F_3, \ldots, F_n\} (1)$$

After that, among the n-number of frames, the key frames are extracted using SSIM. By measuring the similarity between images, the SSIM method determines the quality of videos.

The SSIM index among two frames with windows of general dimensions j and j+1 is measured using Equation 2;

$$SSIM(F_j, F_{j+1}) = \frac{(2\mu_j\mu_{j+1}+S_1)(2\sigma_{jj+1}+S_2)}{(\mu_j^2+\mu_{j+1}^2+S_1)(\sigma_j^2+\sigma_{j+1}^2+S_1)'} \quad (2)$$

Where μ_j represent the average ofF_j, μ_{j+1} represent the average ofF_{j+1}, σ_j^2 represent the variance ofF_j, σ_{j+1}^2 represent the variance ofF_{j+1},σ_{jj+1} represent the covariance of $F_j$andF_{j+1}. S_1andS_2 are two variables defined as $S_1 = (b_1/)^2$ and $S_2 = (b_2/)^2$, where $b_1 = 0.01$, $b_2 = 0.03$, and $/$ is the dynamic range and it has the value of$2^{bitsperpixel} - 1$.
This similarity measure is done for all the pair of frames. After calculating the similarity level, there is a pre-defined threshold for selecting the main frames. If the similarity index of frame is less than the threshold means, the frame is considered as key frame.
The proposed SSIM is a reliable approach used to measure similarity by minimizing computational problems and therefore, it is possible to accurately extract key frames.

Feature extraction
After the key frame extraction process, we extract the three different features from each key frame namely, coverage factor, STI points and shape features. These three features are used for activity recognition process.

Shape feature extraction
For HAR, a shape feature is an important one. For shape feature extraction grid-based approach is utilized. A human body's shape is treated as an object to be extracted in this approach. Shapes are mapped to grids of fixed dimensions in this approach. Here, initially, we split the object into 4×4 matrix, because the size of the grid is4×4. A grid-based approach is a pixel-based technique, which maps the shape to a grid of fixed dimensions. In this, we split the object into 4×4 matrix, because the size of the grid is4×4. From left to right, and from top left corner to bottom left corner, the grid is scanned. Therefore, an array with 0 and 1 binary array can be created based on the object. The function can be authenticated based on the matrix element. Consequently a vector of magnitude 1×16 can be obtained when extracting the shape.

STI point extraction
It is necessary to identify local structures in the location and time domains of video in order to analyse spatiotemporal events. It is because the image values between the two domains differ significantly on a local basis. Image points with local variations in intensity are informative and are called 'interest points'.

To resolve the local features in the frame$K(B, C, N)$, at first, a quantitative-external representation is structured based on equation (1). The term $B, C and N$indicates the (x, y, t) coordinates of the image frame.

$$K(., \eta^2, \tau^2) = k * U(., \eta^2, \tau^2) \quad (3)$$

Where, (η, τ) represents spatial and temporal parameters and U represents as the Gaussian kernel. The Gaussian kernel U is calculated using equation (4).

$$U = \frac{exp(-(b^2+c^2)/2\eta_x^2 - n^2/2\tau_x^2)}{\sqrt{(2\pi)^3\eta_x^4\tau_x^2}} \quad (4)$$

Where, $n$represent the total number of frames. Then, based on the spatiotemporal image gradient, the second-moment matrix is calculated, i.e. within its neighbourhood,

$$\mu(., \eta^2, \tau^2) = U(., c\eta^2, c\tau^2) * (\nabla K(\nabla K)^T) \quad (5)$$

By defining the features, the local maxima are calculated.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

$$H^S = det(\mu) - trace^3(\mu) \quad (6)$$

$det(\mu)$is used to determine the matrix.

By automatically selecting $(\eta, \tau)$the size of the feature can be adjusted according to the spatio- temporal size of the frame structures. In this the size and speed are varied based on the moving object in the frame.Spatial environments of attributes contain information related to the origin of events. Following are the spatiotemporal jets used to evaluate this;

$$Y = (U_a, U_b, U_n, U_{aa}, \ldots, U_{nnnn}) \quad (7)$$

The calculation is completed at the focal point of the elements with the assistance of standardized subordinates that are estimated based on the values(η^2, τ^2). With the estimated speed, the surroundings of the features can be adjusted to calculate the variability based on the camera movements [30]. Thus, by separating the STI points from the frames, the vector of the dimension 1×40can be reached.

Coverage factor

Based on the shape or the coverage of an object, the coverage factor helps to identify the function. Material coverage is calculated by averaging the extracted interest points. With the mean value, the centre point of the object can be calculated, which indicates its location. The horizontal and vertical lengths of an object can be measured based on its location. As a result, the coverage of the object can be determined, giving a vector of dimensions 1 × 2.Therefore, from the feature extraction process, one aspect vector A dimension is 1 × N, where N = 58 can be reached. To classify the function for authentication, the proposed classifier is given as input A, which will be discussed in the following section.

Activity recognition using enhanced long short term memory

After the feature extraction process, the extracted features are fed to the ALSTM classifier to classify a different activity of key frames. LSTM is the RNN based enhancing structure. For realizing the utilization information in sentences of greater distance overcome the issue of gradient disappearance in RNN, the LSTM presents a gated mechanism and memory unit. The LSTM unit has three control gates that are input, forget and output gates. Also, memory cell state is utilized to store and update information. The mathematical expressions of the LSTM model [23] are as defined in equations (6)-(11):

The output of the input gate is defined using (7)

$$I_t = \left([h_{t-1}, v_t]W_I + b_I\right)\sigma \quad (8)$$

The output of the forget gate is defined using (8)

$$F_t = \left([h_{t-1}, v_t]W_F + b_F\right)\sigma \quad (9)$$

The output of the output gate is defined using (9)

$$O_t = \left([h_{t-1}, v_t]W_O + b_O\right)\sigma \quad (10)$$

The current state of the input vector is defined using(10)

$$\tilde{U}_t = \tanh\left([h_{t-1}, v_t]W_U + b_U\right) \quad (11)$$

The update state at time t is defined using (11)

$$U_t = \tilde{U}_t * I_t + U_{t-1} * F_t \quad (12)$$

The hidden state output at time t is defined using (12)

$$h_t = \tanh(U_t) * O_t \quad (13)$$

Where, tanh and σ denotes the hyperbolic tangent function and the sigmoid activation function respectively. v_t indicates the input vector. I_t, F_t and O_t denote the output of the input, forget and output gates at time t respectively. b and W denote the bias and weight of the control gates. $\widetilde{U_t}$ represents the input's current state. U_t and h_t represent the update state and output at time t. To improve the efficiency of LSTM, the weight values are optimally selected using ASO algorithm.

Weight optimization using ASO algorithm

For weight optimization, in this paper ASO algorithm is developed. ASO is a

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

combination of seagull optimization and oppositional based learning algorithm (OBL). The OBL strategy is used for increases the searching ability of seagulls. The seagull optimization algorithm is designed based on the behaviours of seagulls. Seagulls are a species of seabird around the world. Among these, there are a variety of species, and they feed on insects, fish, reptiles, amphibians, and earthworms. The seagull is a very intelligent bird. It uses its intelligence to discover food and attack prey. Their feet are used to mimic the sound of rain to lure earthworms that are hiding underground. To attract fish, they are dipped in breadcrumbs, and to attract fish they are dipped in breadcrumbs. Migratory behaviours are defined as a source of food for sponges and aggressive behaviours are defined as sea tigers' attack behaviours against migratory birds at sea. The migration and attack behaviour of sea tigers is a process of improving individual status. The step-by-step process of weight optimization based on the ASO algorithm is described below.

Step 1: Initialization: Initialization is a very important process, and evolution begins with approximately individuals formed from one population. Assume the size of the seagull population is n and problem space dimension is d and the seagull position. The position of each seagull updated based on the fitness function. The solutions are consisting ofthe weight values of LSTM($W_F, W_I, W_U and W_O$). At first, these weight values are generated randomly. The initial solution format is given in equation (14).

$$S_i = \{S_1, S_2, \ldots, S_n\} (14) S_1 = \{W_F, W_I W_U and W_O\} \quad (15)$$

Where;
S_i→ Set of solutions
W_F→Weight values of forget gate
W_I→Weight values of input gate
W_U→Weight values of tanH layer
W_O→Weight values of output gate

Step 2: Opposite solution creation: In the process of solution generation, opposite solutions are generated based on the initial solutions. The purpose of opposite solution generation is to increase the searching ability. Where, $S \in [a, b]$is a real number and the opposite solution $\overline{S}$ is estimated as follows

$$\overline{S} = a + b - S \quad (16)$$

Step 3: Fitness evaluation: After the solution initialization process, the fitness value is evaluated for each solution. A fitness function is considered to be the maximum accuracy. The fitness function is represented in the follow equation (17);

$$Fitness = Max(Accuracy) (17) \qquad Accuracy = \frac{TP+FN}{TP+TN+FP+FN} \quad (18)$$

*Step 4: Migration (exploration):*Seagulls migrate from one place to another as part of the process of migration. To migrate, seagulls must meet the following conditions.

An additional variable B is used to estimate the new search agent location to avoid conflicts between nearby search agents (SA). The new position of SA is calculated using equation (15).

$$S_A = D \times P_A \ (19)$$

Where, $S_A$represents the status of SA, indicates the current state of SA and D indicates the operating behavior of SA at a given search location. P_A denotes the current position of SA. The D is calculated using equation (20).

$$D = F_c - \left(y \times (F_c / Max_{iteration})\right) \quad (20)$$

In equation 14, the F_C is used to manage the frequency of use of the variable D which is linearly minimized from F_C to 0.

After avoiding a neighbour's conflict, SAs move in the direction of the best neighbours.

$$B_A = A \times \left(P_{bs}(y) - P_A(y)\right) \quad (21)$$

Where; B_A defined as position of P_A towards the P_{bs}, P_{bs}denotes the fittest seagull. A is a random responsible value for balancing between surveys and exploitative assets. Then A is calculated using equation (22).

$$A = 2 \times B^2 \times rand \quad (22)$$

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Where, rand represent the random value [0,1].
Finally, the SA update their position based on best search agent. The updating function is given in equation (23).

$$G_A = |S_A + B_A| \quad (23)$$

Where, G_A defined as distance between the SA and best fit SA.

Step 5: Attacking (Exploitation)
In migration process, the seagulls adjust the angle and speed of the attack. They utilize their wings and weight to sustain their height. While attacking the prey, vortex motion behaviour follows in the sky. This behaviour in the u, v, w direction is described as follows;

$$u' = R^r \times cos(k) \quad (24)$$
$$v' = R^r \times sin(k) \quad (25)$$
$$w' = R^r \times k \quad (26)$$
$$R^r = p \times e^{kq} \quad (27)$$

Where, R^rrepresent the radius value, k represented a random value$[0 \leq k \leq 2\pi]$, p and q are the constant value. The position updating function is given in equation (24).

$$P_A(y) = (G_A \times u' \times v' \times w') + P_{bs}(y) \quad (28)$$

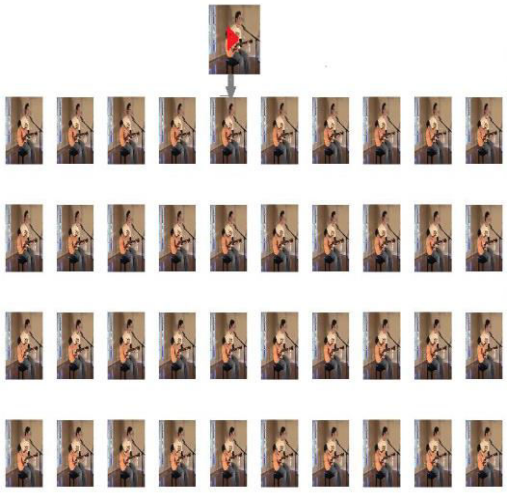
From the equation (28), P_A (y) saves the best solution and updates other SAs' position.

Step 6: Termination criteria:
The algorithm continues until the optimal seagull (weight) is selected. The selected weight value is given to the LSTM classifier.

Results and Discussion
The results obtained by proposed HAR is analysed in this section. The proposed approach is implemented using MATLAB version 20a. The system configuration is, CPU Intel® Pentium 1.9 GHz, 64-bit operating system, Microsoft® Windows 10, 4 GB of RAM. For experimental UCF 101 action recognition dataset is used. The performance of proposed approach has been analysed using different metrics namely, accuracy, sensitivity, specificity and F-Measure.

Dataset Description
For experimental analysis, in this paper UCF 101 action recognition dataset is utilized. This database contains realistic action videos collected from YouTube with 101 action types. This database is an extension of the UCF50 database containing 50 function types. The videos in the 101 action types are grouped into 25 groups, with 4-7 videos of one action in each group. Videos in the same group may share some common features, such as the same background and the same views. Some of the activity of dataset is given in figure 4.

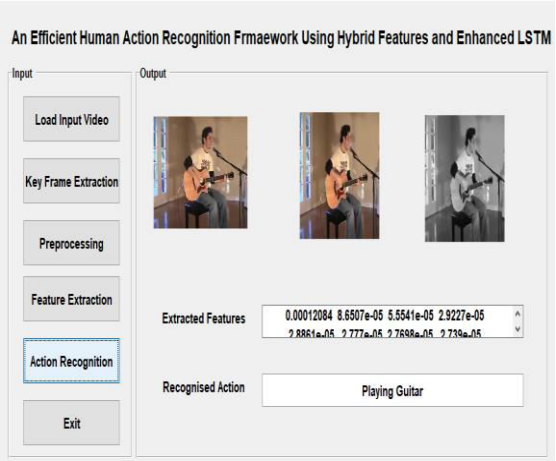
Experimental Results
The results obtained from the recommended approach are listed in this section. The proposed approach mainly consists of three stages namely, key frame extraction, feature extraction and prediction. Each stage output screenshot is presented in this section. To prove the efficiency of the suggested HAR approach, we compare our recommended approach with different state-of-art methods such as LSTM, ANN and SVM. The screen shot of proposed action recognition is given in figure 3.



(a)

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

(b)



(c)

Figure 3: HAR output (a) Extracted key frames, (b)pre-processed output and (c) Prediction output

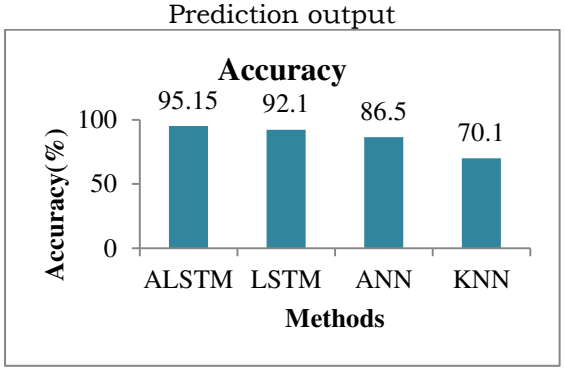


Figure 1: Comparative analysis based on Accuracy

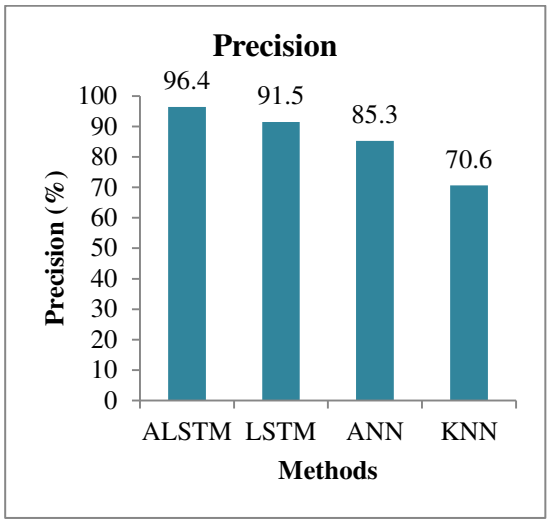


Figure 2: Comparative analysis based on Precision

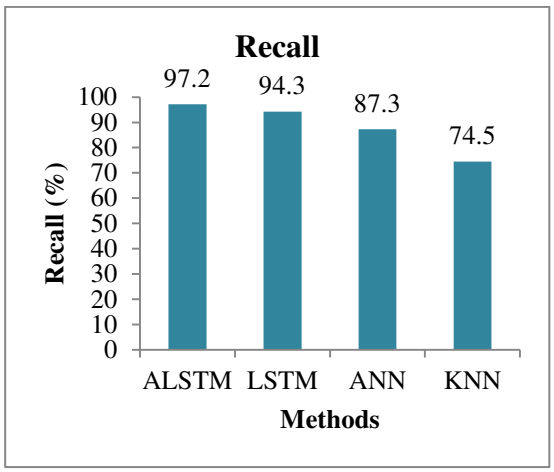


Figure 3: Comparative analysis based on Recall

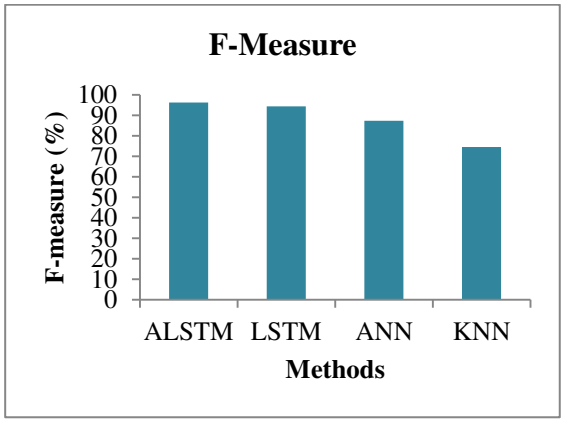


Figure 4: Comparative analysis based on F-Measure

Figure 1 shows the effectiveness analysis of suggested approach based on accuracy. The good HAR system should have the maximum accuracy. When analysing Figure 1, our suggested method achieved

the higher accuracy of 95.15% which is 92.1% for LSTM based HAR, 86.5% FOR ANN based HAR and 70.1% for KNN based HAR. This is due to SOA based parameter optimization process. Similarly, in figure 2, the effectiveness of the suggested method is analysed based on precision measure. According to Figure 2 release, our proposed approach has a higher accuracy of 96.4%, which is 91.5% for LSTM based HAR, 85.3% for ANN based HAR and 70.6% for KNN based HAR. Figure 3 presents comparative analysis based on recall measure. Figure 3 shows our suggested method achieved the higher recall of 97.2% which is high compared to the other methods. Similarly, our recommended method achieved the higher F-measure which is shows in figure 4. From the output graph, we can understand that our proposed approach has reached maximum output compared to other methods.

Conclusion

In this paper, an efficient HAR based on deep learning has been developed. The proposed approach has been designed based on three steps such as key frame extraction, feature extraction and classification. For key frame extraction, SSIM approach has been introduced to extract the key frames based on similarity. After the similarity measure, the important features namely, Shape feature, cover factor and STI points are extracted. Moreover, the ALSTM classifier has been designed for identifying different activities of human. The proposed ALSTM classifier has been enhanced by SOA. The performance of proposed approach has been evaluated and effectiveness compared with the different approaches.

References

[1] Michalis Vrigkas, Christophoros Nikou and Ioannis A. Kakadiaris, A Review of Human Activity Recognition Methods, Front. Robot. AI, 16 November 2015.

[2] Zawar Hussain, Michael Sheng, Wei Emma Zhang, Different Approaches for Human Activity Recognition– A Survey, arXiv:1906.05074v1 [cs.CV] 11 Jun 2019

[3] Andrew P. Hills, Najat Mokhtar, and Nuala M. Byrne, Assessment of Physical Activity and Energy Expenditure: An Overview of Objective Measures, FrontNutr. 2014;1: 5.

[4] Djamila Romaissa Beddiar, Brahim Nini, Mohammad Sabokrou & Abdenour Hadid , Vision-based human activity recognition: a survey, *Multimedia Tools and Applications* volume 79, pages30509–30555 (2020)

[5] Elgammal, A., Duraiswami, R.,Harwood,D., and Davis,L.S.(2002).Background and foreground modelling using non parametric kernel density for visual surveil- lance. Proc.IEEE 90, 1151–1163. doi:10.1109/JPROC.2002.801448

[6] Pirsiavash,H., and Ramanan,D.(2014). "Parsing videos of actions with segmental grammars," in Proc. IEEE Computer Society Conferenceon Computer Vision and Pattern Recognition (Columbus,OH),612–619.

[7] Cheng, G.; Wan, Y.; Saudagar, A.N.; Namuduri, K.; Buckles, B.P. Advances in human action recognition: A survey. arXiv **2015**, arXiv:1501.05964.

[8] K.R.Baskaran and M.N.Saroja, Machine Learning Algorithms for Human Activity Recognition, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-12S, October 2019

[9] Abhay Gupta, Kuldeep Gupta, Kshama Gupta and Kapil Gupta, Human Activity Recognition Using Pose Estimation and Machine Learning Algorithm, ISIC'21: International Semantic Intelligence Conference, February 25–27, 2021.

[10] Neziha Jaoued, Noureddine Boujnah, Med Salim Bouhlel, A new hybrid deep learning model for human action recognition, Journal of King Saud University – Computer and Information Sciences 32 (2020) 447–453

[11] Leandro B. Marinho, A. H. de Souza Junior, and P. P. Rebou¸cas Filho, A new approach to Human Activity Recognition using Machine Learning techniques, Conference Paper in Advances in Intelligent Systems and Computing · February 2017

[12] Jakaria Rabbi, Md. Tahmid Hasan Fuad, Md. Abdul Awal, Human Activity Analysis and Recognition from Smartphones using Machine Learning Techniques, arXiv:2103.16490v1 [cs.LG] 30 Mar 2021