Sparkling Light Publisher

Sparklinglight Transsactions on Arificial Intelligence and Quantum Computing

Website: **https://sparklinglightpublisher.com/** **ISSN (Online):2583-0732**

Optimized Skeleton graph based CNN for Human Abnormal Detection in Video Streams

Bhagya Jyothi K [a*], Vasudeva [b]

[a] *Associate Professor, Computer Science and Engineering, KVG College of Engineering, Sullia, Karnataka,India,*
[b] *Professor, Department of IS, NMAMIT, Nitte, Nitte, Karnataka,India*

Abstract

Human Action Recognition (HAR) is the process of understanding human actions and behavior. HAR has a broad range of applications, and it has been focused on increasing the attention in various domain of computed vision. Abnormal detection from video stream is vigorous to guarantee the security in both outside spaces with the internal. Furthermore, the abnormal actions are really infrequent and rare, which makes the supervision process more challenging and difficult. In this research, skeleton graph-based Convolutional Neural Network (CNN) is devised for human abnormal activity detection. Here, the skeleton graph-based CNN (Skeleton graph_CNN) is devised based on the concept of classical convolution and skeleton graph generation. The human action recognition classifies the human actions into normal and abnormal class. The abnormal actions from the recognized outcome are detected with Skeleton graph_CNN, which provides the various actions of human as an output. The Skeleton graph_CNNgenerates the skeleton shaped human structure by connecting the joints within the frame to consecutive frames. Moreover, the HAR is carried out using IITB-Corridor Dataset based on metrics, such as testing accuracy of 0.961, sensitivity of 0.956 and specificity of 0.960, correspondingly.

*Keywords:*Convolutional Neural Network, Classical convolution, Skeleton graph, Human Action Recognition, Human activity recognition.

1. Introduction

Video is a major source of information with multiple applications for enhancing the superiority of life. The processing of video with high quality information is very crucial. HAR is the process of identifying human actions from video [6]. Video-based action recognition is a challenging domain in detecting and recognizing the human action from video. The major applications of video-based action recognition are surveillance systems [15], content-based video retrieval [14], activity recognition and human–computer-interaction [16][4][1]. In real world environments, the HAR appears as a major role in various domains [4][9]. HAR is a challenging multiclass classification issue due to the high variability of intraclass [8]. The motive of action recognition is to detect and recognize the people, their performance, distrustful activities in the videos, and promote the suitable information to assist interactive programs as well as IoT based applications [4].

Email address of Authors: *kl.bhagya@gmail.com*
vasudevanitk@gmail.com

Please cite this article as: Bhagya Jyothi, K. and Vasudeva (2022). Optimized Skeleton graph based CNN for Human Abnormal Detection in Video Streams. *Sparklinglight Transactions on Artificial Intelligence and Quantum Computing (STAIQC)*, 2(1), 7-16. **ISSN (Online):2583-0732.** Received Date: 25/05/2022, Reviewed Date: 02/06/2022, Published Date: 10/06/2022.

HAR is the major research area in computer vision, and it acts as an important function in various intelligent systems containing human-machine interaction, video surveillance, robot vision, self-driving cars and so on.

Thus, an important goal of HAR is used for determining and recognizing the human actions [9]. Recently, various techniques were introduced for human action recognition. In an advanced video surveillance system, the researchers primarily concentrate on video stitching and virtual reality and abnormal detection in which the abnormal recognition is insignificant application domain and it progressed the major attention in before twenty years. Abnormal detection from video is vigorous in ensuring the security for both internal and outside spaces. Presently, deep neural networks are the favored progressive algorithms subjected to numerous application areas [17] including action recognition. These networks mainly concentrate on knowing the optimal features of information [18]. A deep learning scheme in action recognition from videos aims to recognize the spatial as well as temporal aspects. Numerous deep structural design have been recommended, which have one or more of the subsequent components. The convolution layer of deep learning model is used to excavate the spatial features, and LSTM layers are used to excavate temporal dependencies of features [19][20][6]. The process of detecting abnormal events necessitates the building of complex visual patterns, and some patterns are learned with long-term temporal relationship as well as causal inference [22]. Although a video delivers excess information, it is non-trivial to excavate the information owing to a number of difficulties, like camera motions, viewpoint changes and scale variations [7].

The main intention of this research is the development of Skeleton graph_CNN. Here, the frames are extracted from the video in frame extraction process in which the extracted frames are processed under feature extraction such that the relevant features are extracted. The Skeleton graph_CNN is used to perform two functions, such as human abnormal action recognition and human abnormal activity detection. The abnormal action recognition process classifies the extracted features as either normal or abnormal patients. By the abnormal action, the abnormal activity is detected using Skeleton graph_CNN.

The main contribution of this research is,

Devised Skeleton graph_CNN for human abnormal activity detection: In this research, the Skeleton graph_CNN recognizes and detects the human abnormal activity. Here, the Skeleton graph_CNN is modelled by adapting the concept of classical convolution and skeleton graph generation. Here, the abnormal activity is detected from the recognized abnormal patients as playing with ball, chasing, hiding face, cycling and fighting.

The organization of this research is briefly explained here. Section 2 portrays the motivation, literature survey and challenges of human abnormal activity detection, section 3 portrays the developed methodology for abnormality detection, section 4 explains the results and discussion of abnormality detection and section 5 affords the conclusion of this paper.

2. Motivation

HAR acts as a major function in the human-to-human interaction as well as an interpersonal relation since it offers the information regarding the individuality of a person, psychological state and their personality, which is problematic to excavate. This is the major motivation of this research topic.

2.1 Literature survey

The literature survey of various conventional human abnormal detection methods are explained below. Jaouedi, N. *et al*. [1] devised the hybrid deep learning scheme for recognizing the human action. Here, the Gated Recurrent Neural Network (GRNN) was considered as the combined deep learning approach. Here, the GRNN was devised by incorporating the Gated Recurrent Unit (GRU) and RNN. Although, it provided the better performance, it did not process well with complex datasets. For providing better result with complex datasets, Wei Peng*et al.* [2] modeled the Graph Convolutional Network (GCN) for recognizing the human action through the neural searching. Here, the devised scheme utilized the multiple-hop functions for breaking the drawbacks of representational capacity. Although, it performed effectively in using large scale skeleton-based datasets, it had high computational complexity. In order to diminish the high computational complexity, Dai, C. *et al.* [3] modeled the Two-stream attention based LSTM network (Two-stream attention_LSTM) for recognizing the human action. Here, the devised scheme was not appropriate for dissimilar scenarios, but it produced the error due to the smaller context information. For avoiding the occurrence of error, Muhammad, K. *et al*. [4] modeled the bi-directional long short-term memory (BiLSTM) with a dilated convolutional neural network (DCNN) for identifying the human action. Here, the devised scheme was provided the improved recognition performance, but it failed to utilize the two stream learning strategy.

2.2. Challenges

The challenges of various conventional human abnormal detection techniques are explained below.

- In [1], hybrid deep learning scheme was developed for recognizing the human action. However, the classification time achieved by the devised approach was high. Thus, the challenge relies on reducing the video classification time.
- For reducing the classification time, GCN was devised to recognize the abnormal behaviour in [2]. The challenge of devised method in [2] did not suitable in real time applications.
- In order to implement the action recognition in real time scenarios, BiLSTM was devised in [4]. However, it did not consider the multiple discriminative features for recognizing the complex movements in large-scale datasets.
- The major challenges of human activity recognition are parallel activity recognition at similar time and overlapped activity recognition.

3. Proposed Skeleton graph assisted CNN for human abnormal activity detection

The primary intention of investigation is to progress a new process for human abnormal activity detection using devised Skeleton graph_CNN. The input video is collected from the dataset and the frames connected with the input video are effectively extracted. Based on the video frames, features are mined that includes Hierarchical Skeleton Features (HSF) [10], Histogram of Gradient (HOG) [12], and Shape Local Binary Texture (SLBT) [11]. After excavating the features from input frames, the abnormal action recognition is done based on the skeleton graph based CNN in order to classify the action into normal or abnormal classes. Finally, the process of human abnormal activity detection is carried out using skeleton graph based CNN. Figure 1 shows the block diagram of human abnormality detection.

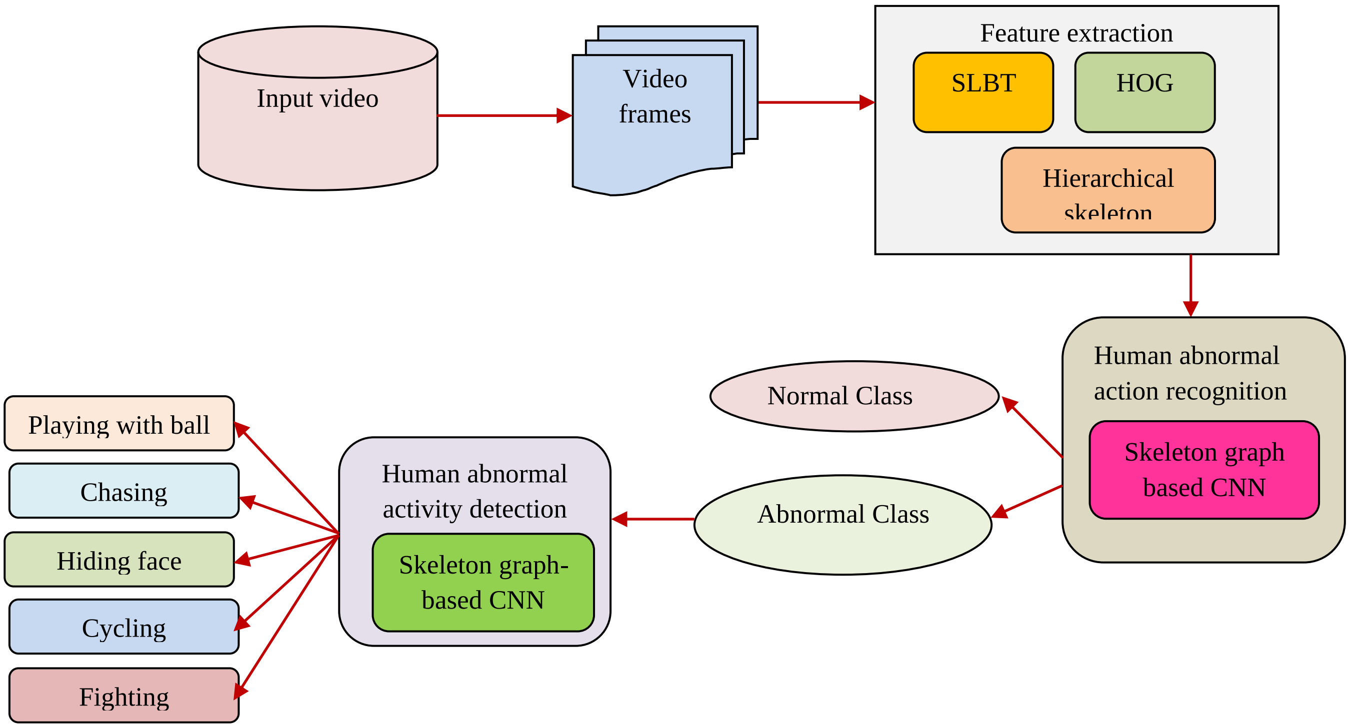


Figure 1. Block diagram of devised Skeleton graph assisted CNN for human abnormality detection

3.1 Video acquisition

Let us assume the dataset B, which contains multiple number of videos, and is stated as,

$$B = \{J_1, J_2, ..., J_b, ..., J_n\} \quad (1)$$

where, J_b denotes the b^{th} video from dataset B and n specifies the overall count of videos. Here, the video J_1 is passed to the frame extraction phase in order to partition the video into multiple frames.

ISSN (Online): 2583-0732

3.2 Frame extraction

After the acquisition of videos from dataset B, the input video J_1 is used in the frame extraction phase. In this step, the number of frames is excavated from the video J_1, which is represented as,

$$N_m = \{m_1, m_2, ..., m_b, ..., m_v\} \quad (2)$$

where, m_b denotes the b^{th} frame from video J_1, N_m signifies the extracted frame set from video J_1 and v specifies the overall count of extracted frame. Once the extraction of frame is completed, any one of the frame is selected to extract the significant features. Hence, this research selects the frame m_b for excavating the significant features.

3.3 Feature extraction

Here, the frame m_b is used for extracting the relevant features. Generally, each of the frames contain multiple meaningful and meaningless features however all of them are not crucial for activity recognition process, but it requires only the powerful and meaningful features. Thus, the significant features are mined in the feature extraction step. Hence, this research computes the features, such as HSF, HOG and SLBT for extracting it from the frame.

HSF:The process of extracting HSF [10] is termed as skeleton pruning approach. Here, the HSF are extracted by eliminating insignificant skeleton branches using boundary abstraction approach, namely Discrete Curve Evolution (DCE). Thus, the HSF features are extracted by eliminating the insignificants parts using the following expression.

$$A(c_1, c_2) = \frac{\eta(c_1, c_2) w(c_1) w(c_2)}{w(c_1) + w(c_2)} \quad (3)$$

Here, $\eta(c_1, c_2)$ signifies the angle of corner having c_1 and c_2, w shows the parameter spanbasedon A, if $A(c_1, c_2)$ ismaximum, then contribution of $c_1 \cup c_2$ to polygon is also maximum. Therefore, the extracted HSF feature from extracted frame m_b is signified as P_1.

HOG: HOG [12] is a feature, which is used to extract the significant region. The HOG descriptor is calculated by partitioning the frames into smaller blocks, termed as spatial image cells. Here, the gradients are deliberated in both vertical as well as horizontal orientations in every pixel of the frame. Thus, the direction ϕ and magnitude C is calculated as,

$$C = \sqrt{C_t^{\,2} + C_u^{\,2}} \quad (4)$$

$$\phi = \tan^{-1}\left(\frac{C_u}{C_t}\right) \quad (5)$$

where, C_t and C_u specifies the horizontal as well as vertical directions of every pixels within the frame. Here, the extracted HOG feature from extracted frame m_b is denoted as P_2.

SLBT: SLBT [11] is the grouping of texture as well as shape information. SLBT feature is mined by seeing the LBP texture feature. Consequently, the amalgamation of texture as well as shape vector is statistically stated as,

$$R_{rs} = \begin{pmatrix} Y_r Z_r \\ Z_s \end{pmatrix} \quad (6)$$

where, Z_s represents the texture information, Y_r signifies the diagonal matrix, and Z_r denotes the shape information. The extracted SLBT feature from extracted frame m_b is indicated as P_3.

Thus, the final feature vector is attained by joining three features, for instances, HOG and SLBT, which is expressed as,

$$P = \{P_1, P_2, P_3\} \quad (7)$$

where, P_1 denotes the HSF feature, P_2 denotes the HOG feature, and P_3 depicts the SLBT feature.

3.4 Human abnormal action recognition using skeleton graph-based CNN

The extracted feature P is considered as the input of human abnormal action recognition using Skeleton graph_CNN[5][13]. The process of transforming from images to graph structured information is adapted in this region. Generally, the Graph convolutional network comprises of two subdomains, such as spectral perspective and spatial perspective. The spectral perspective is employed to transform the graph data into spectrum data by subjecting the CNN on spectral domain. Likewise, the spatial perspective is a process, which is used to get the spatial data by applying the CNN on spatial domain. Generally, the skeleton-based recognition system provides the skeleton data as output for recognizing the human motion.

3.4.1 Skeleton graph_CNN

The skeleton-based recognition system operates on the basis of two streams, such as hand crafted features and deep learning techniques. Here, the Skeleton graph_CNN is used to recognize the human abnormal action, which is designed to generate the skeleton shaped human as output. The Skeleton graph_CNNis modeled by applying classical convolution on image pixels [13], which is expressed as,

$$d_{jk}^{out} = \sum_{d_{jk} \in M(d_{jk})} d_{jk} w\left(f(d_{jk})\right) \tag{8}$$

where, d_{jk}^{out} specifies the convolution process, $M(d_{jk})$ specifies the neighboring pixel set, d_{jk} denotes the pixel in $M(d_{jk})$. After the application of CNN in image pixels, the skeleton gram is generated [13]. Generally, the 2D and 3D co-ordinates of human joint is used to deliberate the skeleton system. In this research, the skeleton graph is generated with the use of Skeleton graph_CNN. Let us assume the graph $G = (K, M)$ on a skeleton sequence having P joints and V frames. In a graph G, the node set K is represented as $K = \{k_{vx} | v = 1,...,V, x = 1,...,P\}$, which contains every joint in a skeleton sequence. Moreover, the skeleton graph is generated with two processes. For that, the joints within one frame are linked with the edges of human structure. After that, every joint in the same frame is linked with the similar joint in the successive frame. Finally, the combination of classical convolution and the generated skeleton graph produces the skeleton structure of human. In addition, the edge set M composed of two subsets in which the first subset indicates the intra skeleton link of entire frame, whereas the second subset contains the edges of inter frame, which links the similar points in consecutive frames. Thus, the outcome of Skeleton graph_CNN is illustrated as W, which is either normal class or abnormal class. If the recognized outcome is normal class, then the process is terminated, else if the recognized outcome is abnormal class, then the recognized outcome is sent to the abnormal activity detection phase where the detection is done using Skeleton graph_CNN. Figure 2 shows the process of Skeleton graph_CNN.

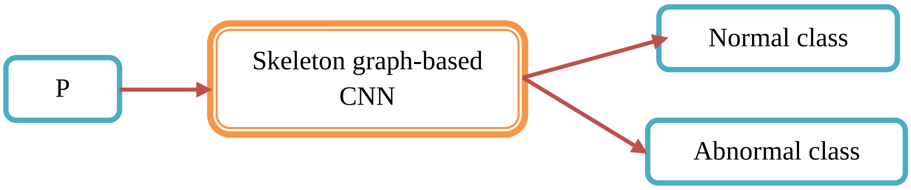


Figure 2. Human abnormal action recognition using Skeleton graph_CNN

3.5 Human abnormal activity detection using Skeleton graph_CNN

The outcome of human abnormal action recognition with abnormal class is indicated as W, which is used to perform the abnormal activity detection using Skeleton graph_CNN. This process detects the abnormal class, which includes playing with ball, chasing, hiding face, cycling and fighting. Moreover, the process of Skeleton graph_CNNis explained in section 3.4. Figure 3 shows the process of abnormal activity detection.

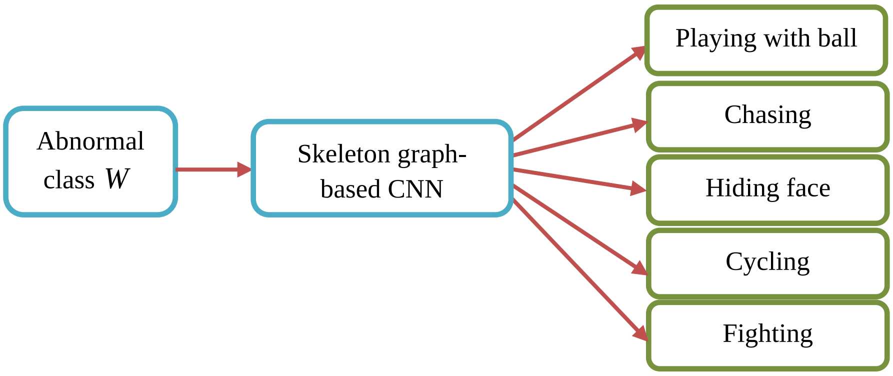


Figure 3.Human abnormal activity detection using Skeleton graph_CNN

4. Results and discussion

This section explains the results and discussion of devised scheme for human abnormal activity detection.

4.1 Experimental setup

The developed skeleton graph_CNN scheme is executed using MATLAB tool using PC with intel i5 core processor.

4.2 Dataset description

The devised Skeleton graph_CNN scheme utilized the IITB-Corridor Dataset [21]. The utilized dataset is gathered in IIT Bombay campus with a single-camera system. The gathered scene contains a corridor with normal activities, such as walking, standing and so on. Moreover, the dataset have various normal and anomalous activities of single human.

4.3 Evaluation metrics

The evaluation metrics considered for the implementation of devised skeleton graph_CNN scheme is explained as below.

Testing accuracy: Testing accuracy is an efficiency measurement metric, which is used to evaluate the closeness of detected outcome with the expected outcome, and is expressed as,

$$W = \frac{d+a}{d+e+a+p} \tag{9}$$

where, d specifies the true positive, a denotes the true negative, e denotes the false positive and p depicts the false negative.

Sensitivity: Sensitivity is a measurement metric, which is used to evaluate the ratio of true positives that are accurately predicted by the devised model, and is expressed as,

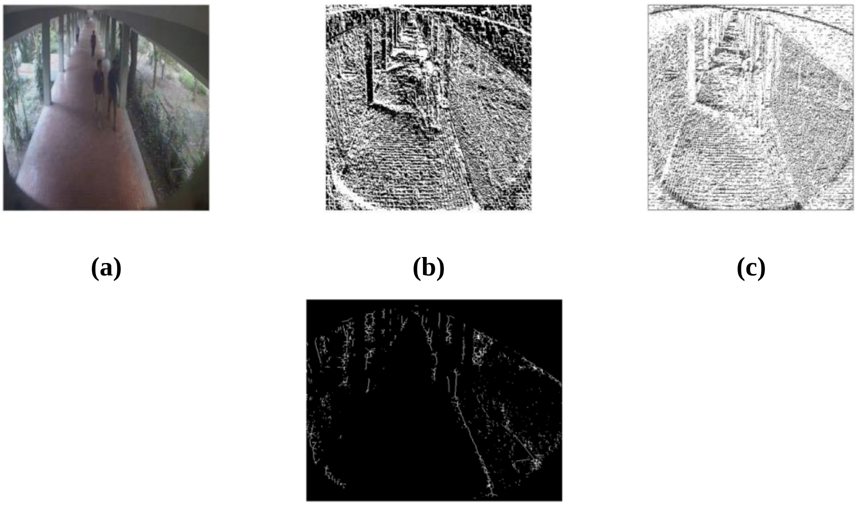
$$X = \frac{d}{d+p} \tag{10}$$

Specificity: Specificity is an evaluation metric, which identifies the ratio of true negatives that are accurately recognized by the devised model, and is deliberated as,

$$Y = \frac{a}{e+a} \tag{11}$$

4.4 Experimental images

Figure 4 shows the experimental outcome of skeleton graph_CNN approach in which the figure 4 a) shows the input frame extracted from video, figure 4 b) demonstrates the extracted HOG feature, figure 4 c) indicates the extracted SLBT feature, and figure 4 d) illustrates the extracted HSF feature from the corresponding input frame.



(a) (b) (c)

(d)

Figure 4. Experimental outcome of devised scheme, a) Extracted input frame from video, b) HOG feature, c) SLBT feature, d) HSF feature

4.5 Comparative methods

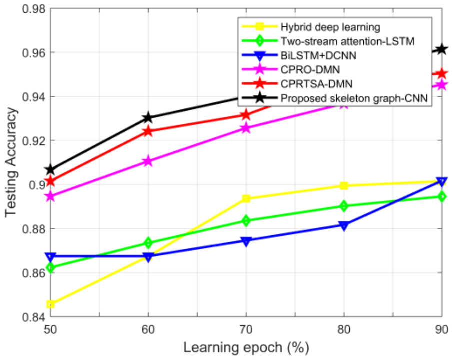
The comparative methods used for analyzing the performance of devised model are Hybrid deep learning [1], Two-stream attention_LSTM [3], BiLSTM+DCNN [4], Chronological Poor and Rich Optimization (CPRO)-based DMN (CPRO_DMN)and Chronological Poor and Rich Tunicate Swarm Algorithm-based DMN (CPRTSA_DMN).

4.6 Comparative analysis

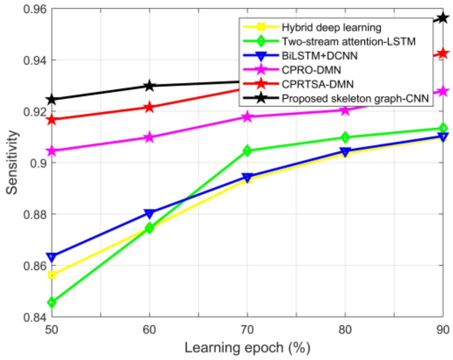
The comparative assessment of devised skeleton graph_CNN scheme is done by adjusting the learning epoch with respect to evaluation metrics.

4.6.1 Analysis based on human abnormal action recognition

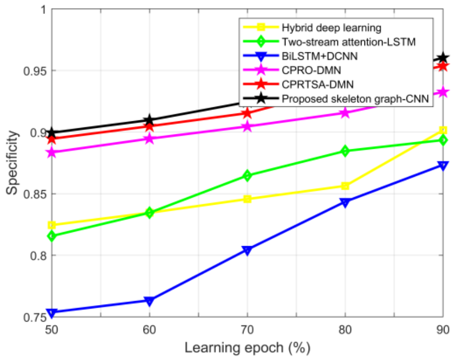
Figure 5a) shows the comparison of projected skeletongraph_CNN scheme for human abnormal action recognition based on testing accuracywith respect tolearning epoch. When the learning epoch is 90%, then the testing accuracy of Hybrid deep learning is 0.901,Two-stream attention_LSTM is 0.894, BiLSTM+DCNN is 0.901, CPRO_DMN is 0.945 and CPRTSA_DMN is 0.950, correspondingly. Moreover, the improved percentage of recognition with existing approaches is 6.23%, 6.95%, 6.21%, 1.68% and 1.15%, correspondingly. Figure 5 b) shows the comparison of devised scheme for sensitivity. The sensitivity of projected skeletongraph_CNN and existing approaches are 0.942, and 0.903, 0.909, 0.904, 0.920 and 0.932 when the learning epoch is 80%. Moreover, the performance upgrading of projected skeletongraph_CNN scheme for sensitivity is 4.13%, 3.47%, 4.03%, 2.34% and 1.07%, correspondingly. The specificity of devised scheme is explained in figure 5 c). The specificity of 0.845, 0.864, 0.804, 0.904, 0.915 and 0.924 are achieved by the existing and projected techniques when the percentage of learning epoch is 70. Furthermore, the improved percentage of devised method with previously modeled techniques are 8.51%, 6.45%, 12.94%, 2.14% and 0.97%, correspondingly.



(a)



(b)

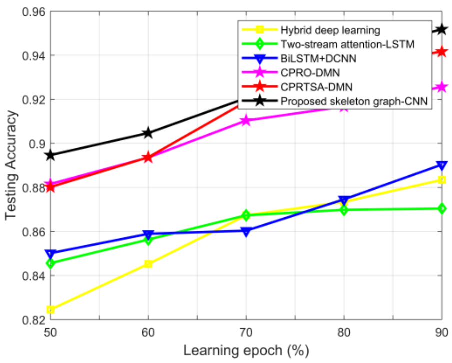


(c)

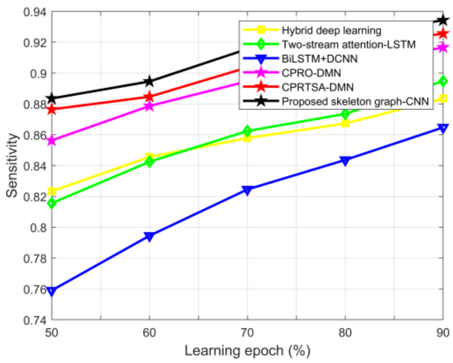
Figure 5. Analysis of human abnormal action recognition based ona) Testing accuracy, b) Sensitivity, c) Specificity

4.6.2 Analysis based on human abnormal activity detection

The graphical outcome of projected skeletongraph_CNN based on testing accuracy is demonstrated as in figure 6 a). The testing accuracy of skeletongraph_CNN is 0.951, whereas the testing accuracy of Hybrid deep learning is 0.883,Two-stream attention_LSTM is 0.870, BiLSTM+DCNN is 0.890, CPRO_DMN is 0.925 and CPRTSA_DMN is 0.941, correspondingly, when the learning epoch is 90%. The improved percentage of projected scheme for testing accuracy is 7.18%, 8.54%, 6.45%, 2.74% and 1.06%, correspondingly. The sensitivity of projected skeletongraph_CNN is given in figure 6 b). For the learning epoch is 50%, then the specificity got by the projected skeletongraph_CNN is 0.883, whereas the specificity of techniques, such as Hybrid deep learning, Two-stream attention_LSTM, BiLSTM+DCNN, CPRO_DMN and CPRTSA_DMN are 6.81%, 7.68%, 14.09%, 3.07% and 0.8%.Figure 6 c) exhibits the comparison of devised scheme for specificity. The specificity of projected skeletongraph_CNN and existing approaches are 0.915, and 0.831, 0.843, 0.803, 0.883 and 0.899 when the learning epoch is 80%. Moreover, the performance upgrading of projected skeletongraph_CNN scheme for specificity is 9.13%, 7.82%, 12.20%, 3.45% and 1.73%, correspondingly.

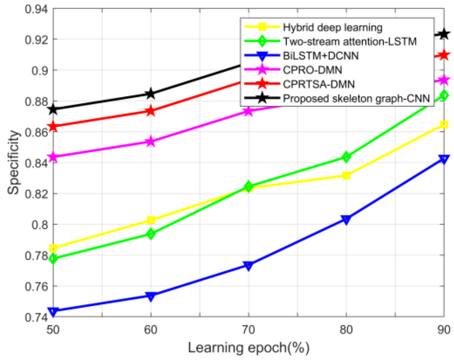


(a)



(b)

ISSN (Online):2583-0732



(c)

Figure 6. Analysis of human abnormal activity detection based on Testing accuracy, b) Sensitivity, c) Specificity

4.7 Comparative discussion

The comparative discussion of devised skeletongraph_CNN is explained in figure 1. Here, the developed approach is compared with the performance of previously devised approaches. Moreover, the comparison of devised scheme is done by comparing it with several traditional approaches such as Hybrid deep learning, Two-stream attention_LSTM, BiLSTM+DCNN, CPRO_DMN and CPRTSA_DMN. In this research, the comparison is done based on human abnormal action recognition and human abnormal activity detection. From the table, it clearly showed that the devised approach outperformed with traditional schemes in human abnormal action recognition. The developed skeletongraph_CNN acquired the better testing accuracy, sensitivity and specificity of 0.961, 0.956 and 0.960. Likewise, the traditional techniques perceived the testing accuracy of 0.901, 0.894, 0.901, 0.945 and 0.950, sensitivity of 0.909, 0.913, 0.910, 0.927 and 0.942 and the specificity of 0.901, 0.893, 0.873, 0.932 and 0.953, correspondingly.

Table 1. Comparative discussion

Variation	Metrics	Hybrid deep learning	Two-stream attention_LSTM	BiLSTM+DCNN	CPRO_DMN	CPRTSA_DMN	Proposed skeleton graph_CNN
Human abnormal action recognition	Testing accuracy	0.901	0.894	0.901	0.945	0.950	**0.961**
	Sensitivity	0.909	0.913	0.910	0.927	0.942	**0.956**
	Specificity	0.901	0.893	0.873	0.932	0.953	**0.960**
Abnormal activity detection	Testing accuracy	0.883	0.870	0.890	0.925	0.941	0.951
	Sensitivity	0.883	0.894	0.864	0.916	0.925	0.934
	Specificity	0.864	0.883	0.842	0.893	0.909	0.923

From the discussion, the devised scheme outperformed with conventional approaches because of the efficacy of projected approach. The projected skeletongraph_CNN scheme is devised by applying the concept of classical convolution and generated skeleton graph on the input features such that the effectiveness of detection based on testing accuracy, sensitivity and specificity of devised scheme is improved.

5. Conclusion

HAR is the process of identifying the abnormal activity from videos, images or text. Generally, HAR is utilized in various applications, such as anti-terrorists, surveillance, life logging and anti-crime securities. This paper devises the abnormality activity detection process, namely skeletongraph_CNN scheme. Here, the devised approach extracts the multiple frames from video using video frame extraction approach. In order to improve the effectiveness of detection, the more significant and meaningful features, such as HSF, SLBT and HOG features are extracted, which aims to improve the effectiveness of recognition and detection. The abnormal action recognition based on skeletongraph_CNN scheme classifies the extracted features into abnormal and normal class. The activity detection based on skeletongraph_CNN scheme is applied on the abnormal class for detecting the abnormal activities of humans including playing with ball, chasing, hiding face, cycling and fighting and so on. The experimental outcome evaluates that the devised approach attained the better performance based on the testing accuracy, sensitivity and specificity of 0.961, 0.956 and 0.960, correspondingly. In addition, the devised approach was failed to process with real world applications. In future, the performance of devised approach can be improved by applying it with real time applications.

References

[1] Jaouedi, N., Boujnah, N., & Bouhlel, M.S. (2020). A new hybrid deep learning model for human action recognition. *Journal of King Saud University-Computer and Information Sciences,* 32(4), 447-453.

[2] Peng, W., Hong, X., Chen, H., & Zhao, G. (2020). Learning graph convolutional network for skeleton-based human action recognition by neural searching. *In Proceedings of the AAAI Conference on Artificial Intelligence,* 34(03), 2669-2676.

[3] Dai, C., Liu, X., & Lai, J. (2020). Human action recognition using two-stream attention based LSTM networks. *Applied soft computing,* 86(1), 105820.

[4] Muhammad, K., Ullah, A., Imran, A.S., Sajjad, M., Kiran, M.S., Sannino, G., & de Albuquerque, V.H.C. (2021). Human action recognition using attention based LSTM network with dilated CNN features. *Future Generation Computer Systems,* 125(1), 820-830.

[5] Yan, S., Xiong, Y., & Lin, D. (2018). Spatial temporal graph convolutional networks for skeleton-based action recognition. *In proceedings of Thirty-second AAAI conference on artificial intelligence.*

[6] Majd, M., & Safabakhsh, R. (2020). Correlational convolutional LSTM for human action recognition. *Neurocomputing,* 396(1), 224-229.

[7] Cho, S., Maqbool, M., Liu, F., & Foroosh, H. (2020). Self-attention network for skeleton-based human action recognition. *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision,* 635-644.

[8] Patel, C.I., Labana, D., Pandya, S., Modi, K., Ghayvat, H., & Awais, M. (2020). Histogram of oriented gradient-based fusion of features for human action recognition in action video sequences. *Sensors,* 20(24), 7299.

[9] Pham, H.H., Khoudour, L., Crouzil, A., Zegers, P., & Velastin, S.A. (2018). Exploiting deep residual networks for human action recognition from skeletal data. *Computer Vision and Image Understanding,* 170, 51-66.

[10] Sheeba, P.T., & Murugan, S. (2019). Fuzzy dragon deep belief neural network for activity recognition using hierarchical skeleton features. *Evolutionary Intelligence,* 1-18.

[11] Lakshmiprabha, N.S., & Majumder, S. (2012). Face recognition system invariant to plastic surgery. *In proceedings of 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA),* 258-263.

[12] Lakshmi, N.D., Latha, Y.M., & Damodaram, A. (2013). Silhouette extraction of a human body based on fusion of HOG and graph-cut segmentation in dynamic backgrounds.

[13] Zhang, X., Xu, C., Tian, X., & Tao, D. (2019). Graph edge convolutional neural networks for skeleton-based action recognition. *IEEE transactions on neural networks and learning systems,* 31(8), 3047-3060.

[14] Spolaor, N., Lee, H.D., Takaki, W.S.R., Ensina, L.A., Coy, C.S.R., & Wu, F.C. (2020). A systematic review on content-based video retrieval. *Engineering Applications of Artificial Intelligence,* 90, 103557.

[15] Keshavarzian, A., Sharifian, S., & Seyedin, S. (2019). Modified deep residual network architecture deployed on serverless framework of IoT platform based on human activity recognition application. *Future Generation Computer Systems,* 101, 14-28.

[16] Antar, Anindya Das, Masud Ahmed, and MdAtiqurRahmanAhad. (2019). Challenges in sensor-based human activity recognition and a comparative analysis of benchmark datasets: a review. *In IEEE Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR),* 134-139.

[17] Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M.S. (2016). Deep learning for visual understanding: A review. *Neuro computing,* 187, 27-48.

[18] Budnik, M., Gutierrez-Gomez, E.L., Safadi, B., & Quénot, G. (2015). Learned features versus engineered features for semantic video indexing. *In IEEE 13th International Workshop on Content-Based Multimedia Indexing (CBMI),* 1-6.

[19] Srivastava, N., Mansimov, E., & Salakhudinov, R. (2015). Unsupervised learning of video representations using lstms. *In International conference on machine learning,* 843-852.

[20] Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., & Darrell, T. (2015). Long-term recurrent convolutional networks for visual recognition and description. *In Proceedings of the IEEE conference on computer vision and pattern recognition,* 2625-2634.

[21]IITBCorridor dataset, "https://rodrigues-royston.github.io/Multi-timescale_Trajectory_Prediction/" accessed on March 2022.

[22] Huiyu Mu, Ruizhi Sun, Gang Yuan, and Yun Wang. (2021). Abnormal Human Behavior Detection in Videos: A Review. *Information technology and control,* 50(3), 522-545.

******