Int. Jrl. of **Human Computations & Intelligence**
Vol. 02, Issue No. 01, 2023
ISSN: 2583-5696

ORIGINAL RESEARCH

# Dynamic Human Action Recognition and Classification Using Computer Vision

**Geeta C Mara . Syed Thouheed Ahmed . Vinaya R M**

School of Computing & Information Technology, REVA University, Bangalore, India

Received: 15 December 2022 / Revised: 18 January 2023 / Accepted: 14 February 2023


**Abstract –** Action Recognition is one of the time series classification problem where we analyse the data from series of time steps for classifying and predicting the action being performed accurately. Video analysis shows significant improvement as it can predict the outcome of the future state by inferring the present state. This has been possible with the help of Computer Vision, Machine Learning, and Deep Learning fields. In video-based action recognition tasks, many actions can be identified and inferred based on the movements of the actions performed. In our paper, we use UCF50 dataset which consists of fifty categories of actions performed from YouTube. In our paper, we use the video classification model to solve action recognition by analysing each frame of the videos. We aim to create a 2D Convolutional Neural Network classifier, then implement a single frame video classifier along with moving average technique on the live videos, to classify and predict the action performed. Our proposed model has obtained an accuracy of 98.56%.

**Index Terms –** Action Recognition, Video Classification, 2D Convolution Neural Network, Moving Average, Single Frame Video Classifier.

## I. INTRODUCTION

Action Recognition [AR] is most common researching area as it helps us to identify the action performed by people. Previously, various researchers have used different methods to perform the action recognition and identification tasks. It is useful and popularly used in the applications like video surveillance, identity recognition, creation of intelligent and smart systems for human-machine interactions, health monitoring, gestures identification, navigation of robots, analysis of human behaviour, etc. Action Recognition tasks can be identified and recognized by using the data collected by the sensors, images, and by inferring the videos, and some of the machine learning, deep learning, and computer vision classification methods have been implemented by the previous researchers to identify the actions performed. In sensor-based action classification, the sensors placed in the wearable devices helps us to determine the action performed by the signals collected and are mainly used for monitoring the patient's movements and fitness.

Int. Jrl. of **Human Computations & Intelligence**
Vol. 02, Issue No. 01, 2023
ISSN: 2583-5696

Action recognition in videos helps in analysing and inferring the movements and the action performed in the videos. Action recognition is a sub-domain of Computer Vision. One of the applications of vision-based action recognition is for surveillance purposes to detect abnormal or anomalous events. Video-based action recognition is a challenging task compared to the traditional recognition of activities. Computer Vision [CV] is a part of computer science that helps machines in identifying and recognizing entities in images and videos like humans do. We are using computer vision techniques as we are performing the action recognition in the videos. Computer Vision can help the computer or the machine to accurately classify the action performed in the images or videos.

2D Convolutional Neural Network [2DCNN] is a Deep Learning [DL] algorithm that is used as a classification model to classify the actions performed in the video. Deep learning algorithms use neural networks to perform like a human brain. Single frame classification divides the videos into frames and takes each frame individually and the moving average technique is used to analyse the series of frames or videos to correctly identify the action performed. A few of the challenges faced in the video-based classification are variations in the activity rhythms, viewpoints, motion of the camera, cluttered background, occlusion, rate of execution of the activity performed, etc. Each action performed by people vary and has different rhythms, the action performed is not in a fixed or static viewpoint as the location, posture, and motion patterns differ at the different camera angles and it severely affects the features of the motion and may give ambiguous information of the action performed. Cluttered background produces disturbances in the video and we may not get any useful information from it and the rate of execution is that each person performs a particular action at their own pace and it's difficult to identify the action performed.

### *A. Motivation*

There are many challenges faced in the video-based action classification and recognition as discussed above. Many researchers have performed many classification techniques to overcome the challenges faced. Action recognition in videos is an important research area and it is popular mainly in surveillance and human behaviour analysis and its necessary to accurately recognize the action performed.

### *B. Contributions*

We use the UCF50 action recognition dataset in our paper to recognize the action performed in the videos. We first build a 2DCNN model to classify the videos and then we implement the single frame classification using the moving average technique to identify and recognise the action performed in the live videos accurately.

### *C. Organisation*

The paper has been arranged as follows, Section II contains the related work previously done by the authors in the Vision-based action classification field. Section III consists of the model proposed, the algorithm used and the description of the dataset used, Section IV is about our outcomes and the results obtained by our proposed model followed by conclusion and references done in our paper.

Int. Jrl. of **Human Computations & Intelligence**
Vol. 02, Issue No. 01, 2023
ISSN: 2583-5696

## II. RELATED WORK

Popescu et al., proposed a fast trainable ML based neural search model for recognizing activities performed [1]. Information from 3D video channels is combined by giving these data as inputs individually through 2DCNN layers. The results are merged in a summarized list of scores using fusing techniques that are not computationally intensive but gives the meaningful insights from a video. The model was tested on 3 available datasets and a new dataset PRECISHAR. The model proved to be highly accurate: 98.43% on MSRDA3D, 91.41% on UTD-MHAD, 90.95% on NTU, and 94.38% on the dataset. Ajmal et al., [2] proposed a weakly supervised approach which requires only the activity labels to train and identify the complex human activity recognition from realistic videos. Restricted Boltzman is used to systematically combine multilevel contextual features. They evaluated their approach on surveillance video datasets for human interaction activity recognition. The results show improved accuracy of 97.01% on benchmark datasets.

Lu et al., developed a 3D CNN model to detect fall [3] using video data and train feature extraction model for big dataset of DL solution. 2D convolutional could not extract features of the motion so they used 3D convolution for fall detection. They trained on the Sports-1M dataset which has no fall data, and is fed to LSTM for training model with falling data. They obtained an accuracy of 100%. Zerrouki et al., proposed an HAR scheme [4], which uses AdaBoost algorithm. Their approach was verified by implementing it on the two available fall datasets. Their proposed model was compared with the other NN, KNN, SVM and NB classifiers showed better results in identifying gestures of humans with an accuracy of 93%. Tanberk et al., proposed a deep model to evaluate HAR in videos [5]. 3D-CNN, 3D-CNN with optical flow, LSTM were implemented to detect the motions. SVM was used to classify and process the videos. Their proposed architecture recognizes and classifies action in videos performed by humans.

Ayhan et al., proposed a model to recognize regular actions and used VideoGraph for monitoring the changing rhythms in the videos [6]. They evaluated their model using two available video datasets, Breakfast and VIRAT, which has of lengthy and complicated videos. The analysis showed that VideoGraph outperformed other methods with a high accuracy even when the videos have extreme changes in the rhythm. They observed that the VideoGraph is less sensitive than the other models to changing rhythms. Zhu et al., [7] proposed YoTube a new DL model to generate activity schemes in videos, where each activity relates to a spatial-temporal which helps in locating one activity performed by humans. They developed a new recursive and a static YoTube detector. The YoTube detector consecutively bounds the candidate boxes using Recursive Network by long term learning. The model outperformed the state of art models. YoTube model obtained an accuracy of 75.31%. RNN gave an accuracy of 69.08%. Tu et al., [8] proposed a model which emphasizes on the aggregation of the local descriptors of the spatio temporal vectors. The model aggregated the important features in the videos by segmenting the features and sampling the videos. Results show that their method is able to extract deep features spatiotemporally and performed better than the other models.

Rohan et al., approached to build a systematic model to analyse gait using DL technique such as convolution classifier [9]. The model proposed estimates the position of skeletal structure of humans to

identify the abnormal activities in the gait of a person. The accuracy achieved to identify and classify the normal and abnormal gait is 97.3% which proved that the model proposed was efficient. Yang et al., proposed a model for recognizing the skeletal based activity using the convolutional method [10]. The implemented depth-first traversing to represent and understand the skeletal design and to improvise the sense of skeletal images and to preserve the information of the structure. Proposed an attention network which takes the sub-images as input. They experimented on the NTU RGB+D and the SBUKI dataset which performed better than the other models. The model was also verified on noisy data of the UCF101 dataset and the Kinetics dataset.

Siddiqi et al., introduced a method to select the important features by normalization [11]. The proposed method is an extension of the maximum applicability and minimum redundant method. The advantage of this method is that it fuses the strengths of various extracting methods. They used the HMM to recognize action by selecting the important features. They compared the model with standard datasets such as KTH and Weizmann datasets. The model performs better than the other model and obtained an accuracy of 98%. Xu et al., proposed a combination model of temporal and spatial Q-networking architecture (ST-DQN) [12], to optimize the searching technique. They experimented on the UCF-101, Sports, and JHMDB datasets and the outcome shows that model obtained better performance locally with less proposals and it exploits the information to detect the actions accurately. Dhiman et al., proposed an architecture that is computed by R-transformation and Zernike Average Energy Silhouette Images (AESIs) [13]. The architecture was verified on a AbHA, URFD, KARD, and NUCLA dataset. The model exhibited better results from other methods based on the accuracy. The model obtained an accuracy of 96.5%, 96.64%, 95.9% and 86.4% on URFD, KARD, AbHA, and NUCLA dataset, respectively.

In paper [14], Ye et al., proposed a probabilistic algorithm for encoding activities patterns for HAR. The model recognizes different activities performed by humans by learning the patterns of the temporal structures. They introduced an approach to generate the features from the patterns for encoding the activities in videos, and the temporal structure is measured by the Hamming distance. Experiments on HAR dataset shows that the model introduced is efficient and has better accuracy. Nawaratne et al., introduced an architecture that addresses the standard DL, hierarchy and multi learning by self organizes growing map method [15] which helps the model learn from the unlabelled video data and addresses the issues of over fitting and extract the information on neural network architecture. The model proposed was experimented on the three standard available video datasets and the results confirm the validation and usability for recognizing the activities performed by humans. Geeta et al., [16] [17] [18] have performed extensive survey on data auditing and security cloud computing. The aim of this work is to classify and predict the human activity in the videos. Various researchers have approached the classification of actions in the videos using various deep learning algorithms on different datasets. In this work, we approach this problem by using the 2DCNN video classifier and use single frame and moving average technique on the YouTube videos to test the model.

## III. METHODOLOGY

To perform our proposed model [Fig. 1] we follow the steps:

1. Data extraction

Int. Jrl. of **Human Computations & Intelligence**
Vol. 02, Issue No. 01, 2023
ISSN: 2583-5696

2. Data Pre-processing
    a. Reading the Video
    b. Resizing the Frame
    c. Normalization
3. Data Visualizing
4. Splitting the data
    a. Training the dataset
    b. Validating the dataset
    c. Testing the dataset
5. Constructing and Training 2D Convolutional Neural Network
6. Loss function and Optimization – Adam optimizer
7. Evaluate the model
8. Predicting the model using Single Frame with moving average technique.

Data Extraction: The data is downloaded and extracted directly from the website and is loaded to the google colab to access it. The dataset is in the video form and we pre-process the dataset to work on it. We use video_reader.read() to read frames from the video file.

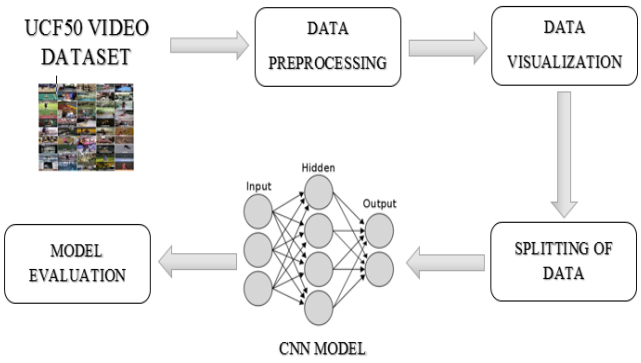


**Fig. 1. Proposed model**

Data Pre-processing: The dataset needs to be pre-processed as we are using the classification model to train on a video dataset. In this pre-processing step, path of the video file is taken as an input, it then reads the video file by individual frames, resizes each frame, normalizes the resized frame, appends the normalized frame into an array and returns the array or a list of classes.

.

The data pre-processing is done in the following ways:

i. Extracting the frames- A function is created that will extract frames from each video while performing other pre-processing operations like resizing and normalizing the images.
ii. Resize- cv2.resize() function is used to resize all the frames of the videos to a fixed dimension to avoid unnecessary computation.
iii. Normalization: Normalizing the data as 0 and 1. Normalize the resized frame by dividing it 255, so that each pixel value lies at 0 and 1.

Int. Jrl. of **Human Computations & Intelligence**
Vol. 02, Issue No. 01, 2023
ISSN: 2583-5696

We create an array or a list which consists of classes of the action we are training. In our work, we are training on 6 classes and iterate them return the frames. After the processing all the videos, labels are added to the selected videos. The processed videos returns the extracted feature and labels in an array. We make use of One Hot Encoding Labels to convert class labels to one hot encoded vectors. With the help of Kera's to_categorical (labels) we can label each action classes with a string. Data Visualization: The videos are randomly choosen from each class of the dataset and is displayed to see how the dataset looks like with their labels. A path of the directory is created for the extracted data and choose the maximum number of the images to be trained for each class.

Splitting the dataset: Two arrays are created after data pre-processing step, the first array contains the images and the second array contains the labels of the classes in the one hot encoded vector. 75% of the data is used for training our classifier and validation checks the accuracy of the model on the data it has not seen before. Rest 25% of the dataset is tested on our model. Constructing and Training 2D Convolutional Neural Network model: The classification model consists of 2 convolutional layers for classifying the videos. We trained the 2DCNN model with 50 epochs to get accurate result and avoid over fitting of the model. We used ReLu activation function in our model. ReLu is an activation function, when the inputs are positive it gives the outputs as 1, else it gives the output as 0.

Loss function and optimizer: Early stopping callback regularizer is used to avoid the over-fitting of the model as it monitors the validation loss for each epoch, if the model's loss does not decrease every 15 epochs then the training will be stopped. A model needs a loss function and an optimizer for training. Categorical_cross entropy loss function is used for classifying multiple classes. Here, we are using Adam optimizer to optimize our model to acquire good accuracy rate. Evaluate the model: We evaluate the performance of the classifier based on the training and testing data. The model is evaluated based on the losses and accuracies obtained and plot the graphs for the same. Loss shows the error we got while training the model, so lower the error value better is the performance, and accuracy shows how well our model can be utilized in classifying the videos and identifying the action performed. The graphs are plotted in the results section.

Predicting the model using Single Frame with moving average technique: After the model is constructed and trained we try to test its performance on some YouTube videos. With the help of pafy library we can access any YouTube video and return its title by passing the link of the video. We set the window size as 25 in moving average which helps in predicting on each frame independently and returns the name of the action performed.

Algorithms used:

### *A. 2D Convolutional Neural Network[2DCNN]:*

2 Dimensional Convolutional Neural Network [CNN] has 2 dimensional convolutional layers [Fig. 2] of the neural network which helps in the analysis and classification of the images or videos and has the following layers-

Int. Jrl. of **Human Computations & Intelligence**
Vol. 02, Issue No. 01, 2023
ISSN: 2583-5696

- Conv2D – 2D convolutional layer helps in determining the number of output filters in the convolution.
- Batch Normalization – Standardizes the input to a network, it accelerates the training and reduces the generalization error.
- Max Pooling – Pooling operation that is used to perform the calculation in finding the largest value in the network and it also reduces the parameters to be learnt by the model.
- Dense Layer – It feeds all output from the previous layers to all its neurons, each neuron providing one output to next layer.

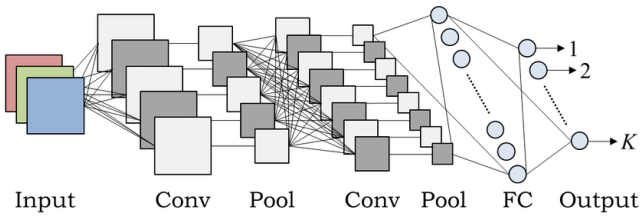


**Fig 2. Convolutional neural network**

### *B. Single Frame Classifier [SFC] :*

A video is a stack or collection of images. Images are a group or collection of pixels. A frame is one of the many still images which is composed of complete moving picture or a video. Single frame [Fig. 3] gives frames as input to the CNN. The SFC model is an example of classifying videos which aggregates the predictions across single frames/images. Video classification is implemented using an image network, we run an image classifier on all individual frames of the video and then combine all the individual probabilities to get the probability of the action performed in the video. SFC model performs well and all the frames is not considered to run on the classifier, but few frames are taken from the entire video. This single frame function will predict single output for the entire video as it takes 'n' number for frames from the complete video and prediction is made by aggregating the predictions of the 'n' number of frames to output the final action class of that video. This SFC technique is useful when a video consists of one action and can check the score of the action performed in the videos along with its name.

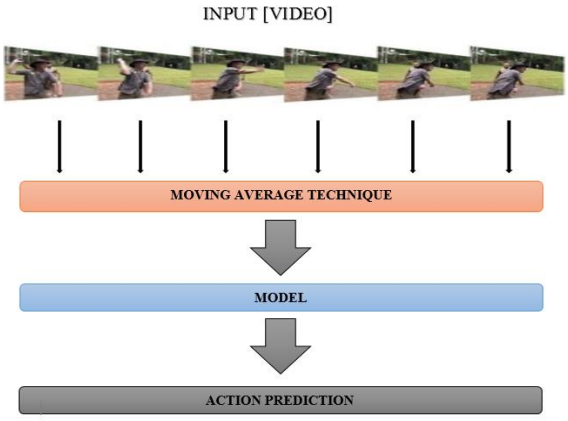


**Fig 3. SFC**

Int. Jrl. of **Human Computations & Intelligence**
Vol. 02, Issue No. 01, 2023
ISSN: 2583-5696

***Dataset Description***

UCF50 is a video based action recognition dataset which consists of fifty categories and the videos are grouped into twenty five groups where each group consists of four action clips [Fig. 4] of realistic action videos taken from YouTube. This dataset has realistic videos unlike the other action recognition dataset which are staged by actors. This dataset is very challenging because of the variations in the motion of the camera, position, location, cluttered background and the other challenges as mentioned.

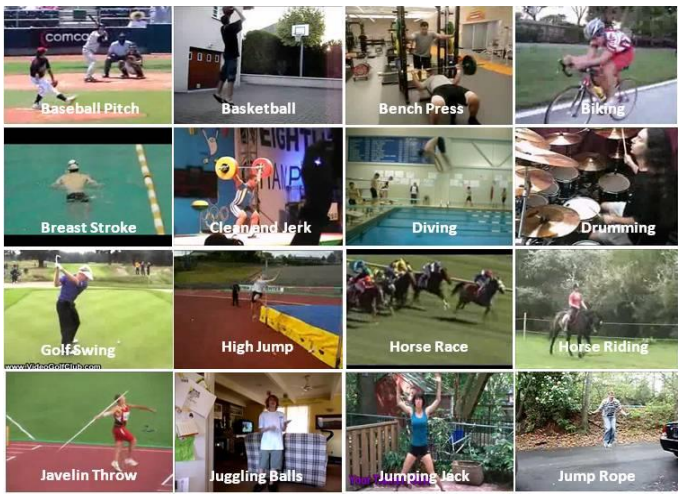


**Fig 4. UCF50 dataset**

## IV. RESULTS

In our model, we have implemented a simple CNN video classifier and a Single Frame technique with moving average to predict the action performed in the video. We have used these above methods on UCF50 dataset which contains fifty classes of actions in which we trained our model on 6 classes. Various techniques and methods were used to classify the classes of the action performed, here, we implemented the moving average with single frame method of 50 epochs to identify the action performed in the videos and obtained an accuracy of 98.55% and a loss of 0.0475%. [Fig. 5] shows the accuracy obtained from our model. The graph is plotted on number of epochs in the x-axis versus the accuracy on the y-axis. The training and the testing accuracy obtained during the training of the model is plotted. We can observe the variations in the accuracy obtained during the training and got an accuracy of 98.56% at the end of 50$^{th}$ epoch.

[Fig. 6] shows the loss obtained from our model. The graph is plotted on number of epochs in the x-axis versus the loss on the y-axis. The training and the testing loss obtained during the training of the model is plotted. We can observe the decrease in loss throughout the training and got less loss of 0.0475% at the end of 50$^{th}$ epoch. The Table I shows the comparison with previous work. We observe that our proposed model has more accuracy than that of the related [4] work. In the related work they have implemented Adaptive Boosting Machine Learning algorithm on URFD dataset and obtained 96.56% accuracy but in our proposed model we have obtained 98.55% for 2D-CNN model and by using the moving average technique and single frame on the UCF50 dataset.

Int. Jrl. of **Human Computations & Intelligence**
Vol. 02, Issue No. 01, 2023
ISSN: 2583-5696

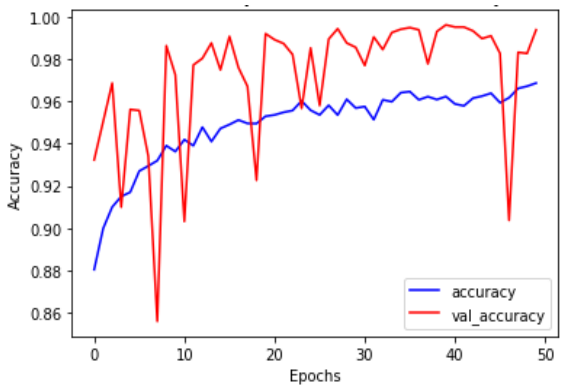


**Fig 5. Accuracy model on CNN**

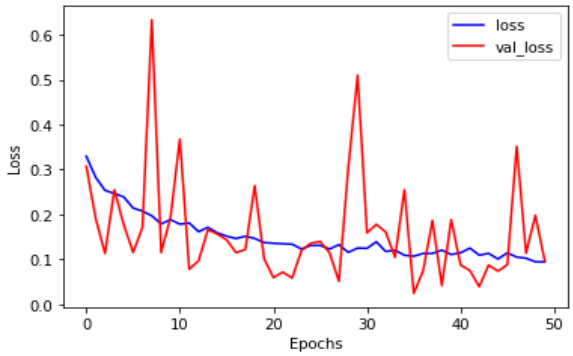


**Fig 6. Loss model on CNN**

**Table I: Comparison Table**

| Reference Paper | Algorithm | Accuracy% | Loss% |
|---|---|---|---|
| Vision-Based Human Action Classification Using Adaptive Boosting Algorithm [4] | Adaptive Boosting | 96.56 | 0.2831 |
| Proposed Model | Single Frame 2DCNN | 98.56 | 0.0475 |

[Fig. 7] shows the accuracy comparison of our work with the related work and we can observe that our model performs better than the previous work. [Fig. 8] shows the loss comparison of the previous work with our work and our model has obtained less accuracy compared to the previous work.

Int. Jrl. of **Human Computations & Intelligence**
Vol. 02, Issue No. 01, 2023
ISSN: 2583-5696

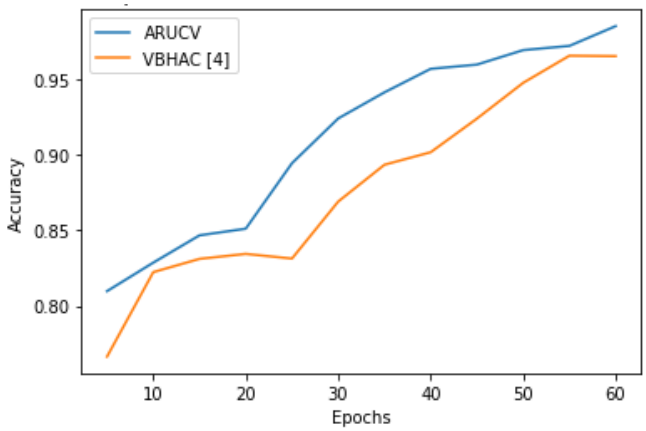


**Fig 7. Accuracy comparison graph**

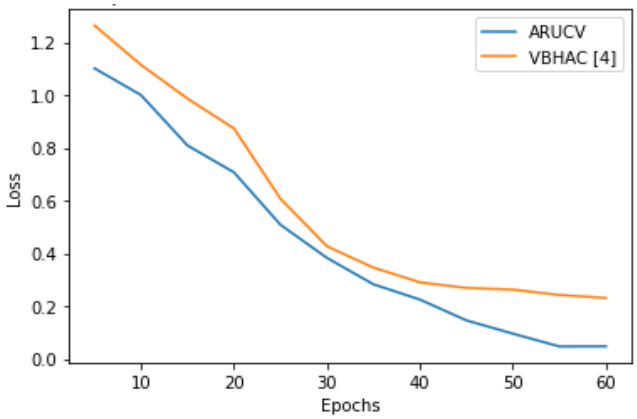


**Fig 8. Loss comparison graph**

## V. CONCLUSION

Action recognition by analysing video is a vast researching area and also a challenging task as it is not easy for a model to detect on its own with a small dataset and requires a lot of training and huge dataset. Various algorithms of ML and DL have been implemented on various already available datasets to recognize and predict the actions performed by humans. Our aim of the paper was to build a system or model which can detect the action or an activity performed from a large dataset and classify with a good accuracy. So, we have used UCF50 dataset which consists 50 action categories, consisting of realistic videos taken from YouTube and 25 groups of video per action category. We first built a CNN video classifier to classify and detect the actions performed in the videos. Later we built a single frame classifier using moving average to test how accurately our model can perform on the YouTube video by passing the link and obtained an accuracy of 98.56% and a loss of 0.0475%.

## REFERENCES

1. Popescu, A. C., Mocanu, I., & Cramariuc, B. (2020). Fusion mechanisms for human activity recognition using automated machine learning. *IEEE Access*, *8*, 143996-144014.
2. Ajmal, M., Ahmad, F., Naseer, M., & Jamjoom, M. (2019). Recognizing human activities from video using weakly supervised contextual features. *IEEE Access*, *7*, 98420-98435.

3. Lu, N., Wu, Y., Feng, L., & Song, J. (2018). Deep learning for fall detection: Three-dimensional CNN combined with LSTM on video kinematic data. *IEEE journal of biomedical and health informatics*, *23*(1), 314-323.
4. Zerrouki, N., Harrou, F., Sun, Y., & Houacine, A. (2018). Vision-based human action classification using adaptive boosting algorithm. *IEEE Sensors Journal*, *18*(12), 5115-5121.
5. Ahmed, S. T., Guptha, N. S., Lavanya, N. L., Basha, S. M., & Fathima, A. S. (2022). IMPROVING MEDICAL IMAGE PIXEL QUALITY USING MICQ UNSUPERVISED MACHINE LEARNING TECHNIQUE. *Malaysian Journal of Computer Science*, 53-64.
6. Tanberk, S., Kilimci, Z. H., Tükel, D. B., Uysal, M., & Akyokuş, S. (2020). A hybrid deep model using deep learning and dense optical flow approaches for human activity recognition. *IEEE Access*, *8*, 19799-19809.
7. Ayhan, B., Kwan, C., Budavari, B., Larkin, J., Gribben, D., & Li, B. (2020). Video activity recognition with varying rhythms. *IEEE Access*, *8*, 191997-192008.
8. Zhu, H., Vial, R., Lu, S., Peng, X., Fu, H., Tian, Y., & Cao, X. (2018). YoTube: Searching action proposal via recurrent and static regression networks. *IEEE Transactions on Image Processing*, *27*(6), 2609-2622.
9. Tu, Z., Li, H., Zhang, D., Dauwels, J., Li, B., & Yuan, J. (2019). Action-stage emphasized spatiotemporal VLAD for video action recognition. *IEEE Transactions on Image Processing*, *28*(6), 2799-2812.
10. Rohan, A., Rabah, M., Hosny, T., & Kim, S. H. (2020). Human pose estimation-based real-time gait analysis using convolutional neural network. *IEEE Access*, *8*, 191542-191550.
11. Yang, Z., Li, Y., Yang, J., & Luo, J. (2018). Action recognition with spatio–temporal visual attention on skeleton image sequences. *IEEE Transactions on Circuits and Systems for Video Technology*, *29*(8), 2405-2415.
12. Guptha, N. S., & Patil, K. K. (2017). Earth mover's distance-based CBIR using adaptive regularised Kernel fuzzy C-means method of liver cirrhosis histopathological segmentation. *International Journal of Signal and Imaging Systems Engineering*, *10*(1-2), 39-46.
13. Siddiqi, M. H., Alruwaili, M., & Ali, A. (2019). A novel feature selection method for video-based human activity recognition systems. *IEEE Access*, *7*, 119593-119602.
14. Xu, W., Yu, J., Miao, Z., Wan, L., & Ji, Q. (2019). Spatio-temporal deep Q-networks for human activity localization. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*(9), 2984-2999.
15. Dhiman, C., & Vishwakarma, D. K. (2019). A Robust Framework for Abnormal Human Action Recognition Using $\boldsymbol {\mathcal {R}} $-Transform and Zernike Moments in Depth Videos. *IEEE Sensors Journal*, *19*(13), 5195-5203.
16. Ye, J., Qi, G. J., Zhuang, N., Hu, H., & Hua, K. A. (2018). Learning compact features for human activity recognition via probabilistic first-take-all. *IEEE transactions on pattern analysis and machine intelligence*, *42*(1), 126-139.
17. Nawaratne, R., Alahakoon, D., De Silva, D., Kumara, H., & Yu, X. (2019). Hierarchical two-stream growing self-organizing maps with transience for human activity recognition. *IEEE Transactions on Industrial Informatics*, *16*(12), 7756-7764.
18. Geeta, C. M., Raghavendra, S., Buyya, R., Venugopal, K. R., Iyenga, S. S., & Patnaik, L. M. (2018). Data auditing and security in cloud computing: issues, challenges and future directions. *International Journal of Computer (IJC)*, *28*(1), 8-57.
19. Mara, G. C., Rathod, U., RG, S. R., Buyya, R., Iyengar, S. S., & Patnaik, L. M. (2020). CRUPA: collusion resistant user revocable public auditing of shared data in cloud. *Journal of Cloud Computing*, *9*, 1-18.
20. Sreedhar Kumar, S., Ahmed, S. T., Mercy Flora, P., Hemanth, L. S., Aishwarya, J., GopalNaik, R., & Fathima, A. (2021, January). An Improved Approach of Unstructured Text Document Classification Using Predetermined Text Model and Probability Technique. In *ICASISET 2020: Proceedings of the First International Conference on Advanced Scientific Innovation in Science, Engineering and Technology, ICASISET 2020, 16-17 May 2020, Chennai, India* (p. 378). European Alliance for Innovation.
21. Ahmed, S. T., & Basha, S. M. (2022). *Information and Communication Theory-Source Coding Techniques-Part II*. MileStone Research Publications.
22. Guptha, N. S., Balamurugan, V., Megharaj, G., Sattar, K. N. A., & Rose, J. D. (2022). Cross lingual handwritten character recognition using long short term memory network with aid of elephant herding optimization algorithm. *Pattern Recognition Letters*, *159*, 16-22.