POSE BASED HUMAN ACTIVITY RECOGNITION USING DEEP LEARNING CONVOLUTIONAL NEURAL NETWORK

[1]Mr.G.Venkateswarlu,[2] Ms.R.Jyothsna,
[1,2] Assistant Professor,Dept. of CSE,
Malla Reddy Engineering College (Autonomous), Secunderabad, Telangana State

Abstract: The need and cause of human activity recognition are due to various applications like health care, Elderly monitoring, Safety, Social networks analysis, monitoring the environment, transportation monitoring, surveillance systems etc. The main intention of installing CCTV is to stop the crime or damage by detecting suspicious or abnormal activities that are happening in the surveillance. There is a huge demand for the development of a smart surveillance system which not only reduces human involvement in monitoring but also alerts the respective authority on time from the future miss happening. Since people are aware of the existence of CCTV almost everywhere, in most situations, behavior of people involved in crimes may seem normal. But too many false alarms could also result in irritations or a loss of trust in the system. Thus, this research work majorly introduced to detect the various human activities from the different poses. For effective detection of human activity, the proposed method utilizes the deep learning convolutional neural network (DLCNN) for classifying the various poses. Hence, developing such a novel model with less training time and data set, with high accuracy and self-learning with time is utilized compared to the state of art approaches.

Keywords: Deep learning, Pose recognition.

1. INTRODUCTION

The most common need for the experiment is to take care of the elderly people living alone in home. Their activities has to be monitored to provide medical assistance, they stay far away from home [1]. This type of monitoring also is useful to identify if there is any deviations in the normal and routine activities of elderly people. If any abnormalities are found then necessary steps could be taken to help the elderly people. In some cases even hand gestures has to be learnt and identified to know the real meaning of hand movement the people living in a home independently.

Another major application of Human activity recognition [2] is Security and surveillance applications, even though it is beyond the scope of the paper, activity recognition application is the intention of this experiment. Traditional ways of surveillance are done by human. This continuous monitoring would lead to stress to the one who performs monitoring. Hence a vision based monitoring would be a better choice. The usage of sensors may lead some disadvantages like sensor failures, calibration of sensors etc [3]. So a visual based activity monitoring is used in this analysis in a smart home. We use a Convolution Neural network to classify the activity which was gathered from a smart home.DML smart action data set is used to test and train the input [4].

With the increase in crime all over the world, the use of visual surveillance and cameras for security applications is continuously growing, and it has become part of the modern era. Video surveillance is done by installing CCTV (Closed-Circuit Television), at places to be secured. Surveillance cameras are reasonable besides found universally present existences, but there must be someone who must monitor the activities constantly [5]. In numerous circumstances wherever surveillance cameras are utilized, it is mutual to invented proved observation pay able to humanoid issues such as boredom, tiredness and operator feeling exhausted as nothing new is happening to pay attention. Despite the effort to keep the places under surveillance 24/7, most of the time it is not possible to stop the crime in that instant. As surveillance camera can be a more useful tool if instead of passively recording the activities, it can be used to detect events that need special attention of the operator on time. There is an increasing demand for instinctive recognition of apprehensive behavior of a person in public places such as shops, parking lots, ATM centers [6], airports, railway stations, entrance and corridors of buildings, etc., to identify subjects for standoff threat analysis and detection. Identification of human activity by video stream is a stimulating job. From the past decade, human action recognition has established important consideration by the research of the CPU vision community. Analyzing of human activity by presenting patterns of movement of different parts of the body; but also, explanation of the human purpose, emotions, and opinions. Human behavior investigation and consideration are important for numerous requests like human-computer communication, observation, sports, elderly-health care, training, entertainment, and so on. In general, human activity recognition systems follow a hierarchical approach [7]. At lower-level, human objects are segmented from the video frame. Present process is tracked by feature extraction such as the characteristics of human objects such as colors, shape, silhouette, body motion and poses. The human activity act appreciation segment cascades under a mid-level method surveyed by the intellectual appliances on the high-level that interprets the situation of the activities as either normal or abnormal.

The major contributions of the paper as follows:

- Input Test video will be divided into multiple frames, Then on each frame pre-processing operation is applied to the remove the unwanted noises.
- Then background removal operation is applied to detect the humans from the particular frames along with the segmentation procedure.
- Then, to identify the various poses, Principal component analysis based pose recognition operation is performed.
- The extracted pose based features are applied to the DLCNN for both training and testing operations, the results shows that the proposed DLCNN classification gives the better performance compared to other approaches on public available dataset.

Rest of the paper is organized as follows; section 2 deals with the various literatures with their drawbacks respectively. Section 3 deals with the detailed analysis of the proposed method with its operation. Section 4 deals with the analysis of the results with the comparison analysis. Section 5 concludes the paper with possible future enhancements.

2. LITERATURE SURVEY

In recent days machine learning models [8-9] have produced highest percentage of accuracy wherever it is used, more specifically in areas like image classification. Neural networks like Deep learning classifier, Convolution Neural Network, Recurrent Neural Network, Long Short term Memory classifier are the subset of Machine learning techniques. As mentioned above theses models does not need any handcrafted inputs and could segment and normalize raw input images and could produce higher accuracy from low level to high level input images.

Smart home will play a vital role in future in providing and performing intelligent tasks. It not only controls machine activities like automatic electronics equipment switching ,sensing smoke ,making alarms, opening and closing doors. It has its own kind of action in human activity monitoring and recognizing the occupants through the deep learning techniques. Hence by recognizing the activities decision can be made in accessing the devices based on the activity performed by the occupant in the smart home. Therefore utilizing machine learning models in applications like health care, elderly monitoring etc has become more significant. This activity recognition can be made through many technical ways [10] like incorporating sensors, deploying monitoring sensors, incorporating wearable sensors to the body of the occupants. This kind of monitoring using various sensors would be annoying to the occupants and might produce disturbances like producing noise, may give wrong alarm, breaking of sensors, sensor need to calibrate sensors in regular intervals etc. This may lead to improper and inaccurate results [11].Hence to overcome the deficiencies mentioned in collecting the samples for classification, vision based methods are used to monitor the activities of human in smart home. Vision based methods are more advantageous when compared with the sensor based because of the above said deficiencies. The samples in vision based are collected using various types of diverse cameras to obtain more accurate and appropriate input than the sensor based sample collection [12].

While performing input classification in traditional methods used handcraft methods are used with the machine learning models. The obtained video frames are considered as Histogram, scale invariant images, which is of low level range and would be accepted by limited dataset only, which would be a challenging task to achieve high accuracy. In [13] authors used spatiotemporal features for activity recognition which used the same DML smart action data set [14] has used Sparse coding have been used to form visual words for extracting spatiotemporal features. on negative sparse coding method was used in [15]. Linear SVM and Intersection SVM is used to classify the data set and it achieved 58.20% of accuracy. In [16] a support vector machine classifier was proposed which used Meta-classifier and Bayesian classifier. The accuracy rate was around 72.3%.For measuring the similarities images and videos a kernel based function is used in [17],which is combined with SVM and NNSC algorithm and it achieved highest accuracy rate of 79.9%.

In [18] authors has used Deep Belief Network model to classify for fall detection images ,which was obtained from video inputs taken from a smart home. It produced accuracy rate of 79.4%.

In [19] authors used CAD120 dataset to classify their two data sets namely OA1 and OA2 with the derived CNN model designed for human activity recognition. It managed to get accuracy rate of 60.1% and 45.2% for OA1 and OA2 dataset respectively. In [20] authors have used fusion methods with the kitchen dataset along with Neural networks and SVM which used fusion methods to classify the input images. It produced accuracy rate of 73.1%.

3. PROPOSED METHOD

The major criteria to consider in case of activity recognition, is to increase the accuracy of activity recognition through a long-term period of time. The proposed method uses convolution neural network for human activity recognition. Initially the input dataset collected is preprocessed using image processing technique. The image pre-processing technique used noise removal; convert the image to gray scale, threshold the image. The goal is to detect and classify short duration tasks that compose a more complex activity. The aim of that process is to predict and classify long term action like running, jagging, jumping, and playing games. DLCNN is applied to preprocess the video in high dimensional feature space. The purpose of using the multiclass

DLCNN classifier is that our data set includes multiple activities done by different actors. The classification is then processed using multi class DLCNN to obtain long-term activity recognition. Figure 1 explains the processing steps of the proposed work.

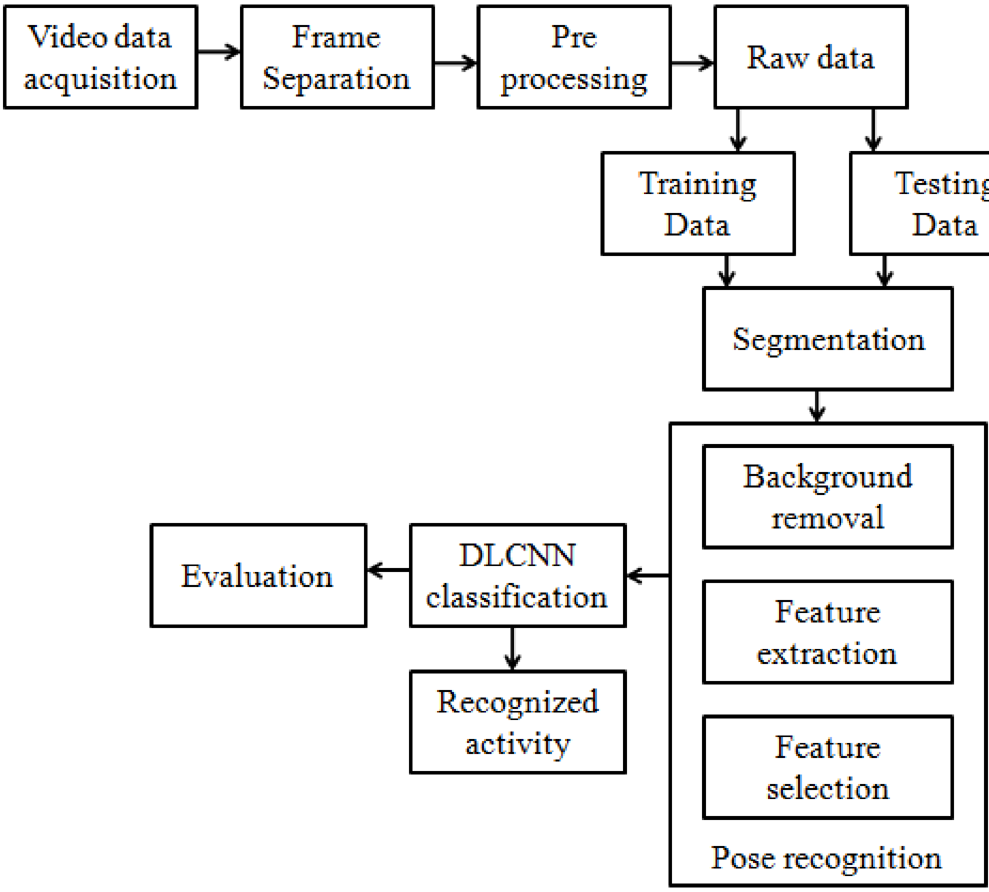

Figure 1: Human activity Eva chain of steps flow chart.

3.1. Frame Separation

In this module, we retrieve separately from the uploaded movies. Each retrieved frame can be stored research and application, pictures are normally handiest intrigued by sure parts. These parts are routinely alluded to as dreams or closer view (as different segments of the recorded past). So as to find and look at the objective inside the image, we have to confine them from the image. The photograph division alludes to the photograph is part into locales, each with attributes and to separate the objective of pastime inside the framework. Recognition is like placing a couple of prescription glasses on detection. After putting on our glasses, we will now recognize that the small blurry item in the distance is, in truth, a cat and now not a rock.

3.1. Video Pre-processing

Pre-processing and cleaning information are critical responsibilities that occur earlier than a dataset is used efficaciously for system getting to know. Raw facts are often noisy and unreliable, and may be missing values. Using records without those modeling tasks can produce misleading consequences. Because pictures are static snap shots, we can't use movement to locate the picture's items however need to rely upon different methods to parse out a scene. Edge detection strategies can help to decide the items in this type of scene. Edges outline item barriers and may be determined by using looking at how depth modifications across an picture.

Five activities from HMDB51 and UCF101 are selected. The activities are Sitting, Waving, Hitting and Kicking from HMDB51 and Walking from UCF101. These are the 5 classes that are used to train the model. Videos from each class are sent to the feature extraction module. Each frame of the videos is converted to 299*299 dimensions and passed to DLCNN networks. Videos in the datasets have frame rate of 30 fps and each video is trimmed down to 90 frames during feature extraction which gives us up to 3 seconds of features of each video. The produced feature for each class is stored in numpy arrays of size (n, 2048) where n is the total extracted features. After the extraction, numpy arrays of each class are selected such that all the rows except the last 500 are taken and concatenated for the training set. The remaining 500 rows in each class are concatenated and used for the testing set. Simultaneously, output array is created using the video labels.

3.2 Segmentation:

In computer imaginative and prescient, the time period "picture segmentation" or truly "segmentation" approach dividing the photo into businesses of pixels primarily based on a few criteria. You can do this grouping based totally on colour, texture, or some different criteria which you have decided. These agencies are every so often additionally referred to as exceptional-pixels. In instance segmentation the purpose is to detect particular gadgets in an image and create a mask around the object of interest. Instance segmentation also can be thought as item detection where

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

the output is a mask as opposed to only a bounding box. Unlike semantic segmentation, which tries to categorize every pixel in the photo, example segmentation does now not intention to label each pixel in the picture. A dynamic updating of history image via frame difference approach and make use of the strength of the history subtraction technique for detecting the moving object very efficiently and correctly. Article discovery is applicable in numerous area names going from insurance (observation), human PC interchange; apply autonomy, transportation, recovery, etc. Sensors utilized for determined observation produce peta-bytes of photograph insights in barely any hours. These data are diminished to geospatial data and included with various measurements to get perfect conviction of front line situation. This technique involves thing discovery to music substances like individuals, vehicles and suspicious items from the uncooked symbolism records. Spotting and recognizing the wild creatures in the domain of sterile zones like business place, distinguishing the vehicles left in confined locales are likewise a couple of bundles of item recognition.

3.3. Background Removal

Separating foreground from background plays an crucial function in lots of computer imaginative and prescient systems, which include motion popularity, movement capture. It differs the video compressing, teleconferencing and surveillance tracing. Image pre- processing is the primary mission in shifting object detection. The small adjustments inside the pixel lead to fake detection. Noise can be brought because of diverse reasons. Due to the noise the pixel values might be modified. So photo pre-processing is very important Noise Removing. Commotion is any element which isn't generally of preferred position to the explanation of picture handling. The affect of commotions on the photo sign abundance and segment is multifaceted nature. So the best approach to simple out commotion and keep the data of picture are the essential commitments of the image filtering. Median channel is a nonlinear procedure for putting off clamor. Its major thought is to utilize the middle of the local pixel dark expense in inclination to the dim cost of pixel

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in

point. For the atypical components, the middle alludes to the size of the inside expense in the wake of arranging.

3.4 Feature extraction-Pose recognition:

A strategy for yoga posture analysis utilizing pose recognition is implemented for effectively extracting the features. the person in the smart home is allowed to perform all common activities an occupant commonly performs. Some of them are standing, walking, eating, sleeping, siting, walk stair case using mobile phones, pick, drop something etc can be effectively classified by the pose recognition.

According to this technique initially recognizes a posture using OpenPose and a camera. At that point, it computes discrimination between the body angles between an instructor and the user. On the off chance that it is bigger than the given threshold, the strategy proposes the rectification of the part. With this proposition, it is normal that individuals can rehearse Yoga anyplace, including home. Along these lines, everybody can practice Yoga, regardless of old enough or wellbeing. For evaluations, the authors applied the proposition to 4 distinct situations, for example, different body sizes, diverse tallness, various ages, and different camera separation, with three Yoga poses. The features are extracted by using the principal component analysis.

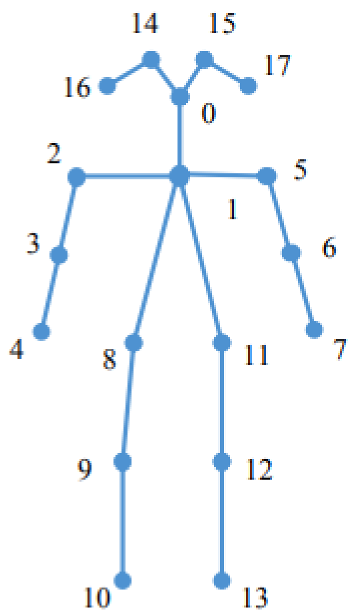

Figure 2: Key points in the human body detected by OpenPose

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in

3.

5 Classification

A deep learning model with high learning efficiency known as convolutional neural network was proposed. Different layers are involved in the basic structure of DLCNN such as output layer, fully connected layer, input layer, pooling layer, and convolution layer. Different inputs have been considered by pooling layer and convolutional layer generally. Using convolutional layer and sampling layer alternately that is one volume that build-up number of hidden layers connect to a pooling layer, and then a convolution connects after the pooling layer, and so on. The process is similar to the convolution process for why it is named as convolutional neural network as each output feature surface in the convolutional layer and its input are connected locally. To get the input value of neuron, local inputs are summed and weighted together with the offset value through the corresponding right value of connection.

Convolutional layer: Convolutional layer comprises multiple feature surfaces or Feature Map in which multiple neuron groups contain in each feature surface and each neuron connects through the upper layer of local area feature surface and the convolution kernel. The kernel of convolution is referred to a weight matrix (a 3*3 or 5*5 matrix in two dimensions). Based on the operation of convolution, the input with different features extracts by the convolution layer of DLCNN. The low-level features like corners, lines, and edges retrieve using the first layer of convolution and more advanced features can extract by higher-level volumes layers. A structure of DLCNN is obtained with a smaller convolution kernel. Some of the conclusions are described as follows:

- The accuracy rate can improve by improving the network depth;
- By increasing the number of feature faces, the accuracy can also improve;
- A higher rate of accuracy is obtained by adding a fully connected layer to the convolutional layer.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

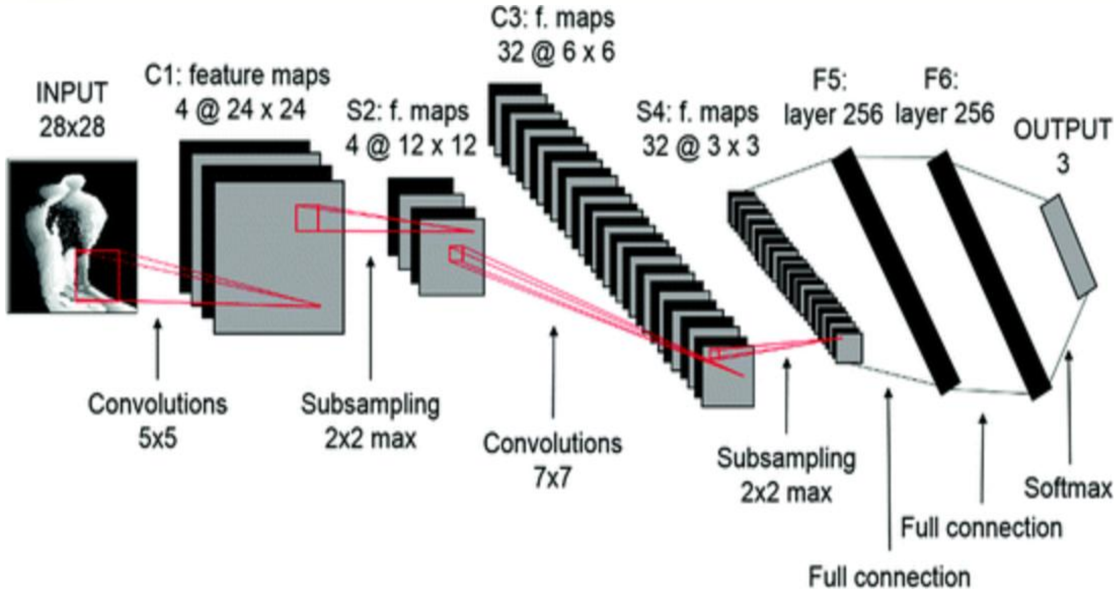

Figure 3: Human activity recognition using DLCNN

Sampling layer: After the convolutional layer, the sampling layer or pooling layer follows immediately that includes multiple feature faces, each of which has the characteristic surface that corresponds to the upper layer's characteristic surface and will not modify the number of feature faces. The sampling layer input is the convolutional layer. A feature surface of the sampling layer and convolutional layer are corresponding uniquely each other and the sampling layer neurons are connected to the input layer locally. The role of secondary feature extraction is played by the sampling layer which has neurons each of which processes the pooling operations on the local receptive field. The most widely used pooling methods is the max-pooling which involves the point with the largest value in the receptive field. In the domain of local acceptance, the random pooling and averaging of all values is done.

Fully connected layer: One or more fully connected layers connect in the structure of CNN after sampling layers and multiple convolutional layers. The convolutional layer is integrated with fully connected layer or the sampling of local information in the layer is differentiated based on the category. The ReLU function is used by each neuron's activation function for improving the performance of CNN network. The last fully connected layer's output value is passed to an output layer and it is termed as softmax layer that can utilize the softmax logistic regression (Softmax regression) for the purpose of classification. The selection of a suitable loss function is essential for a particular category.

The structure in the fully connected layer of MLP and DLCNN is similar in general. The algorithm of BP uses the DLCNN training algorithm. The retaining of test data is performed poorly if a training of large feed forward neural network is done on a small data in the collection owing to the high capacity. Although the hidden layer neuron's output value is 0.5, some hidden nodes fail, and the probability reaches to 0 through the technology, the regularization utilizes in the method of fully connected layer i.e. dropout technology to restrict the training of over-fitting. In the procedures of forward propagation and post propagation of DLCNN, these nodes are not participated for each input of sample into the network. The corresponding structure of a network is not similar but all structures share weights owing to the dropout technology randomness. The complexity of adoption between Meta-learning neurons reduces by this technology as neurons can't exist on other particular neurons and it obtains more robust features. The technologies of ReLU+dropout uses by most researches of DLCNN and good results have obtained in classification performance.

Feature surface: DLCNN includes an important parameter of number of feature faces which is set out through an actual application. Some features not conducive to the network learning in case of higher number of feature faces. If the parameter is lower, some are conducive to the network learning. The network training time and the number of trainable parameters will increase when the numbers of feature faces are higher. This is not conducive for models of learning network. When the script begins receiving frames, it starts extracting features from each frame and storing it in a numpy array. After the first 90 frames are received, it starts creating batches of 90 features from the feature array. This batch of 90 features is then sent to the DLCNN model for classification of activity observed. The advantage of the proposed method is that image classification and feature extraction is applied to improve the recognition

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in

accuracy. Another advantage is that, the recognition is made based on three different classifications that make it useful to obtain deeper analysis for the future.

4. EXPERMENTAL RESULTS

4.1 Dataset

This work adopts activity datasets from UCF101 and HMDB51. UCF101 is a dataset of 101 human activities collected from YouTube videos created by University of Central Florida. HMDB51 is a large human motion database of 51 activities created by Serre Lab research group from Brown University. To perform the investigation and to measure the accuracy of the experiment, all possible activities are considered. That is the person in the smart home is allowed to perform all common activities an occupant commonly performs. Some of them are standing, walking, eating, sleeping, siting, walk stair case using mobile phones, pick, drop something etc.

4.2 Optimization:

To a DLCNN model, data is fed using batches. Each batch of data is mapped to a certain class. In this work, batches are made of 90 frames and passed as input to the DLCNN. Training Data shape: X = (n, 90, 2048), y = (n, num class) Where 'n' is the number of batches. To improve the training of model, the number of batches is increased by repeating the features in batches multiple times. This is achieved by taking strides of 15 (selected using trial and error) while creating a batch. So a sliding window keeps sliding 15 units and adding 90 feature vectors to the input data and simultaneously adding category vector to the output data.

Two models are built for training. First is the DLCNN model which has 4 layers. The first two layers have 1024 and 400 DLCNN units respectively and the last two layers are dense layers with 500 and 5 units respectively. Activation used for final layer is softmax. Loss is calculated using the categorical cross entropy function. The trained DLCNN model is saved and exported to a python script. This script makes use of OpenCV library to read video frames from the camera. When the script begins receiving frames, it starts extracting features from each frame and storing it in a numpy array. After the first 90 frames are received, it starts creating batches of 90 features from the feature array. This batch of 90 features is then sent to the DLCNN model for classification of activity observed.

4.3 Performance Evaluation

Table 1: Experimental Results for Various Methods

	Random Forest [12]	Naïve Baiyes [17]	Logistic Regression [15]	RNN [12]	CNN [16]	ANN [18]	**DLCNN**
Precision	93.998	91.201	92.8956	93.228	94.5891	90.998	95.9887
Recall	93.001	91.989	93.112	93.005	92.008	91.996	95.1234
F1-Score	93.998	91.7748	92.112	93.479	91.003	92.778	95.1102
Accuracy	94.001	91.8887	90.448	93.963	94.999	90.998	96.9991
Recall	93.556	91.0021	91.5456	92.789	91.998	91.7752	95.1023
FPR	4.665	4.7789	3.9785	3.889	3.998	4.665	3.9875

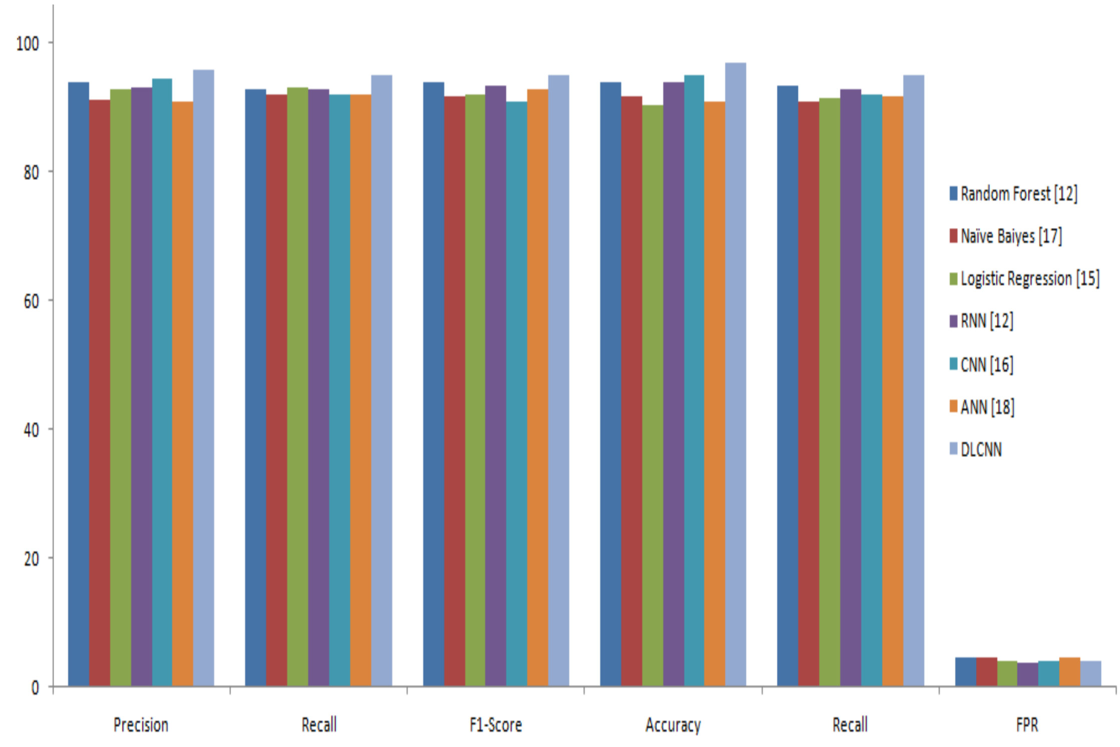

Figure 4: Experimental Results for Various Methods

Experimental results from Table 1 as well as Figure 4 demonstrate the percent of the different assessment parameters for just the credit card fraud dataset for distinct

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in

machine learning techniques. Findings indicate that DLCNN techniques demonstrate an accuracy percentage with 96.988 percent, although Random Forest 93.228 percent, LR 92.89 percent, RNN 91.2 percent, CNN 90.9 percent as well as ANN 93.99 percent demonstrate a precision percentage of human activity identification. For any machine learning or deep learning technique, greater values are shown to be accepted as just a higher performance method of precision, accuracy, recall, and F1-score. As we have seen, there are a few algorithms that have surpassed others as well quite significantly. Thus, selecting DLCNN over all other techniques could be a sensible approach in attaining a greater degree of completeness when decreasing quality just significantly.

5. CONCLUSION

A method for human movement unmistakable quality from significance maps and position estimations the utilization of significant DLCNN has been proposed. Combination activities between the yield expectations of DLCNN channels are defined to expand the score of the correct development. The accuracy rate of Human activity classification based on the various poses are compared with various Machine learning models and managed to get around 97% of accuracy which is better than the existing models discussed in related work. The highest accuracy was made possible by the automatic feature extraction from the input provided. This work can be extended to implement the for further improvement of system efficiency by applying more powerful models like AlexNet, GoogleNet etc to achieve more accuracy in the activity recognition. As a future enhancement the video images can be replaced by image sensor for sensor based activity recognition.

REFERENCES

[1]. Demrozi, Florenc, et al. "Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey." IEEE Access (2020).

[2]. Samie, Farzad, Lars Bauer, and Jörg Henkel. "Hierarchical classification for constrained IoT devices: A case study on human activity recognition." *IEEE Internet of Things Journal* 7.9 (2020): 8287-8295.

[3]. Wan, Shaohua, et al. "Deep learning models for real-time

human activity recognition with smartphones." *Mobile Networks and Applications* 25.2 (2020): 743-755.

[4]. Zhou, Xiaokang, et al. "Deep-learning-enhanced human activity recognition for Internet of healthcare things." *IEEE Internet of Things Journal* 7.7 (2020): 6429-6438.

[5]. Ferrari, Anna, et al. "On the personalization of classification models for human activity recognition." *IEEE Access* 8 (2020): 32066-32079.

[6]. Golestani, Negar, and Mahta Moghaddam. "Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks." *Nature communications* 11.1 (2020): 1-11.

[7]. Chen, Kaixuan, et al. "Deep learning for sensor-based human activity recognition: overview, challenges and opportunities." *arXiv preprint arXiv:2001.07416* (2020).

[8]. Dang, L. Minh, et al. "Sensor-based and vision-based human activity recognition: A comprehensive survey." *Pattern Recognition* 108 (2020): 107561.

[9]. Qin, Zhen, et al. "Imaging and fusing time series for wearable sensor-based human activity recognition." *Information Fusion* 53 (2020): 80-87.

[10]. Ahmed, Nadeem, Jahir Ibna Rafiq, and Md Rashedul Islam. "Enhanced human activity recognition based on smartphone sensor data using hybrid feature selection model." *Sensors* 20.1 (2020): 317.

[11]. Xia, Kun, Jianguang Huang, and Hanyu Wang. "LSTM-CNN architecture for human activity recognition." *IEEE Access* 8 (2020): 56855-56866.

[12]. Teng, Qi, et al. "The layer-wise training convolutional neural networks using local loss for sensor-based human activity recognition." *IEEE Sensors Journal* 20.13 (2020): 7265-7274.

[13]. Taylor, William, et al. "An intelligent non-invasive real-time human activity recognition system for next-generation healthcare." *Sensors* 20.9 (2020): 2653.

[14]. Kwon, Hyeokhyen, et al. "IMUTube: Automatic extraction of virtual on-body accelerometry from video for human activity recognition." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.3 (2020): 1-29.

[15]. Irvine, Naomi, et al. "Neural network ensembles for sensor-based human activity recognition within smart environments." *Sensors* 20.1 (2020): 216.

[16]. Mliki, Hazar, Fatma Bouhlel, and Mohamed Hammami. "Human activity recognition from UAV-captured video sequences." *Pattern Recognition* 100 (2020): 107140.

[17]. Mojarad, Roghayeh, et al. "Automatic classification error detection and correction for robust human activity recognition." *IEEE Robotics and Automation Letters* 5.2 (2020): 2208-2215.

[18]. Cruciani, Federico, et al. "Feature learning for human activity recognition using convolutional neural networks." *CCF Transactions on Pervasive Computing and Interaction* 2.1 (2020): 18-32.

[19]. Janarthanan, R., Srinath Doss, and S. Baskar. "Optimized unsupervised deep learning assisted reconstructed coder in the on-nodule wearable sensor for human activity recognition." *Measurement* 164 (2020): 108050.

[20]. Tanberk, Senem, et al. "A hybrid deep model using deep learning and dense optical flow approaches for human activity recognition." *IEEE Access* 8 (2020): 19799-19809.