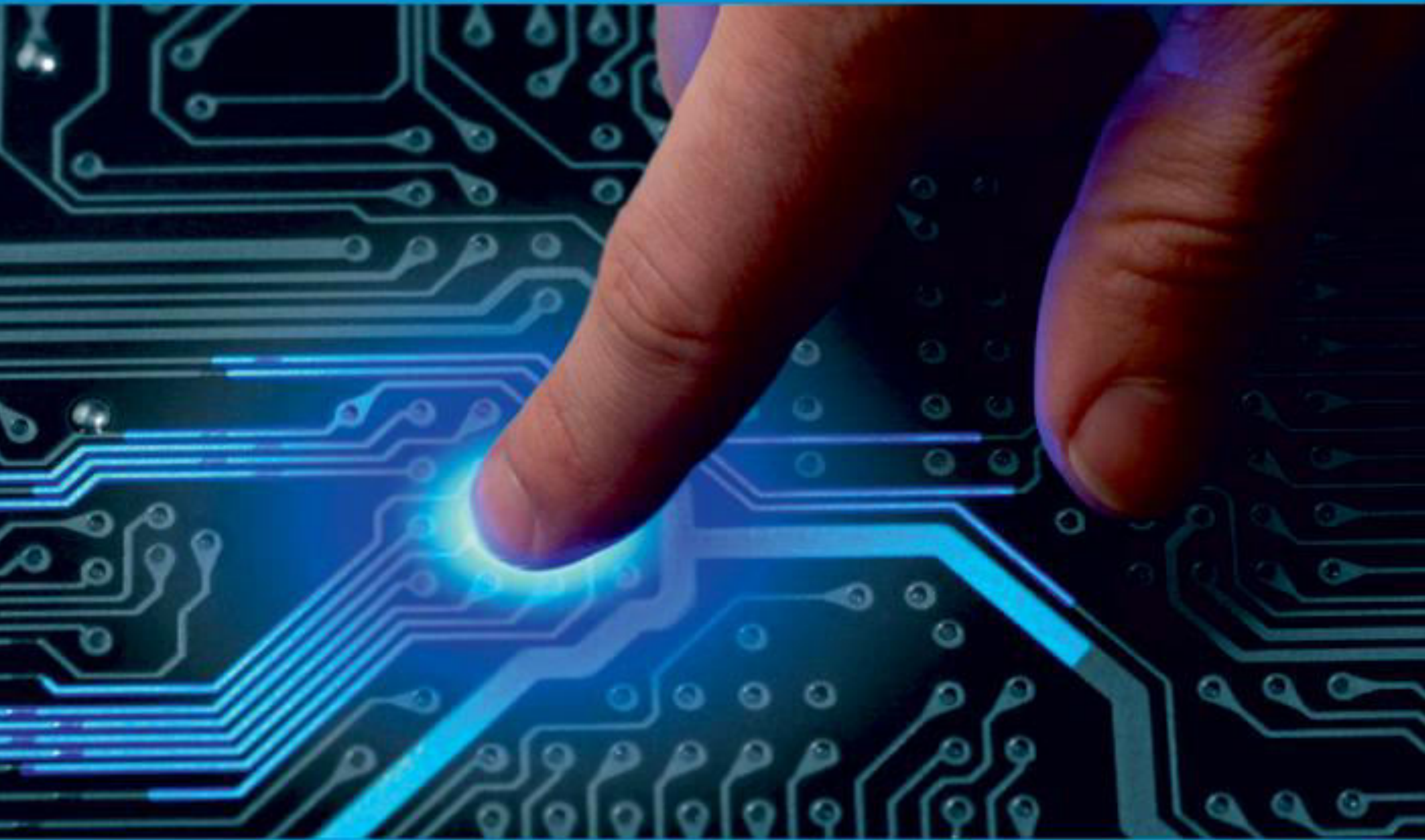
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006095 |

Building Human Action Recognition Using CNN

Mr. A T Barani Vijaya Kumar[1], R Harini [2], S Harini[3], L X Heartina Lancy[4], C Kanimozhi[5]

Assistant Professor, Dept. of CSE, Saranathan College of Engineering, Trichy, Tamil Nadu, India[1]

UG Student, Dept .of CSE, Saranathan College of Engineering, Trichy, Tamil Nadu, India[2-5]

ABSTRACT*:* The digital world is gradually expanding with new features. Human action recognition in films and photos depends heavily on computer vision and pattern recognition. Human Action Recognition is well-developed, allowing a person's bodily action to recognize. Artificial Intelligence (AI) is used to make human action recognition easier. We used CNN, a type of deep learning, to recognize human action. Deep Learning proved to be particularly good at identifying poses in a timely and accurate manner.

KEYWORDS*:* Deep Learning,CNN, Action

I. INTRODUCTION

Nowadays, digital content is exponentially growing day by day. Image processing is regarded as one of the most rapidly involved fields in the software industry. There are many growing applications in all areas of work, so effective AI-based intelligent internet of things (IoT) systems are needed for surveillance to monitor and identify human actions and activities. It has the main responsibility of developing the ultimate machine in the future, which would be able to perform the visual functions of living beings. As such, it forms the basis of all kinds of visual automation. Human action recognition mainly focuses on identifying people, their behaviors, and suspicious activities in videos. Certain problems, such as background clutter, changes in scale, viewpoint, lighting, and appearance occur. Deep Learning methods can be used to solve the problem. There will always be people who come up with ways to break HAR. No matter how much it evolves. Our primary focus will be on a convolutional neural network (CNN). CNN works similarly to how our brain recognizes and remembers things. Standard CNN models, such as AlexNet, are used for capturing spatial information. CNN layers are unique in that they are organized in three dimensions: width, height, and depth. Currently, most researchers have developed a two-stream approach for action recognition to combine the temporal and spatial features for joint feature training to cover the current challenges and limitations of the HAR.

II. RELATED WORK

We combed through a large number of documents. In [1],Human action recognition plays an important role in modern intelligent systems, such as Yoga pose analysis and somatosensory games. When pose analysis is compared with conventional 2-D-based human action analysis using the Kinect sensor, one can obtain depth information about human actions, which is significant for human action recognition. In [2], different areas of medicine, education, entertainment, visual surveillance, and video retrieval were monitored. HAR uses characteristics of public datasets that are being used for HAR. The advantages, disadvantages, challenges, and future directions for HAR In [3], we have implemented an action recognition technique based on feature fusion. An HSI colour transformation is performed, which is the first step to improving the contrast of video frames, which are then extracted as features by an optical flow algorithm. Then the shape and texture features are fused by a new parallel approach called length control features. Finally, the features are passed to M-SVM for feature classification into relevant human actions, which is the final step. A new Weighted Entropy Variance has been implemented to combine vectors and select the best of them for classification. [4] Here, the existing action recognition techniques mainly use certain pre-trained weights of different AI architectures for the visual representation of video frames. To resolve this issue, a bi-directional long-short-term memory (BiLSTM) based attention mechanism with a dilated convolutional neural network is used.[5]Convolutional Neural Network (CNN) structure for accurate and complicated motion recognition is designed. From the world average pooling layer and fully linked (FC) layer, the elements are extracted and used by a proposed high entropy-based approach. Further, a function decision technique to identify the Poisson distribution along with other univariate measures is suggested.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |
|| Volume 10, Issue 6, June 2022 ||
| DOI: 10.15680/IJIRCCE.2022.1006095 |

Some CNN facets are improper, and a few of them are not necessary. This leads to an improper prediction of human actions.

III. PROPOSED ALGORITHM

The Proposed system we have used CNN-Alexnet algorithm along with Mediapipe. The first step is data obtaining process where the images are captured to define the a region of interest(ROI) in the frame. Lots of variables are present in the image which can result in unwanted outcome. So the data that needs to be processed is reduced to a large extent using Mediapipe. The Mediapipe contains only the landmark of human. TheCNN-Alexnet along with Mediapipepackagesusedin theproject torecognize the human action in real time scenario.Alexnet is part of CNN model.The model consists of five layers with a combination of max pooling followed by 3 fully connected layers and they use ReLu activation in each of these layers except the output layers. Convolutional layercontains filters and parameters. The filter is used to resize the image and reduce the dimension of the image. Max Pooling layer is used for segmentation. Dense layer is used to classify the image based on the output from convolutional layer. The output layer is the softmax layer.The softmax layerwhich take decisions.

Capturing and pre-processing

In this module, how the input is taken from the user .The input for this system will be taken from images, videos and web camera. The video input will be converted into frames and then preprocessed. During preprocessing the input will be padded and smart cropped. The input images will be preprocessed during training process.

Detection

In this module, the pre-processed image will be given as an input to a pre-trained model named “Alexnet”. The preprocessed model takes the input and marks the landmarks of human.The detection of the position of a still image or sequence of images(moving images) is known as Action detection. Machines cannot detect objects whereas human eyes can. The human pose in the image varies due to its changing position whether it be sitting, standing, clapping , drinking, running, yoga pose like plank, goddess, tree, warrior and down dog.

Classification

The detected co-ordinates along with the confidence score will be given as an input to the classifier. Our classification model is a multi-layered perceptron .It is a eight layer perceptron. This model gives how accurate a human is towards a pose after classification. Majority of the poses will be displayed out of the given poses.

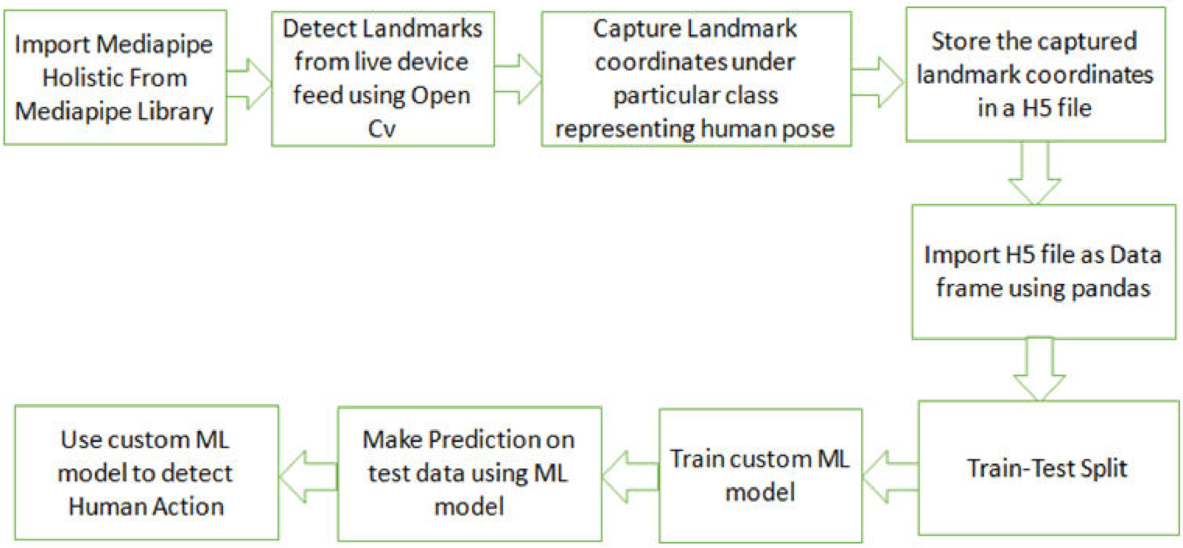


Fig.1. System Architecture

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006095 |

IV. EXPERIMENTAL ANALYSIS AND RESULTS

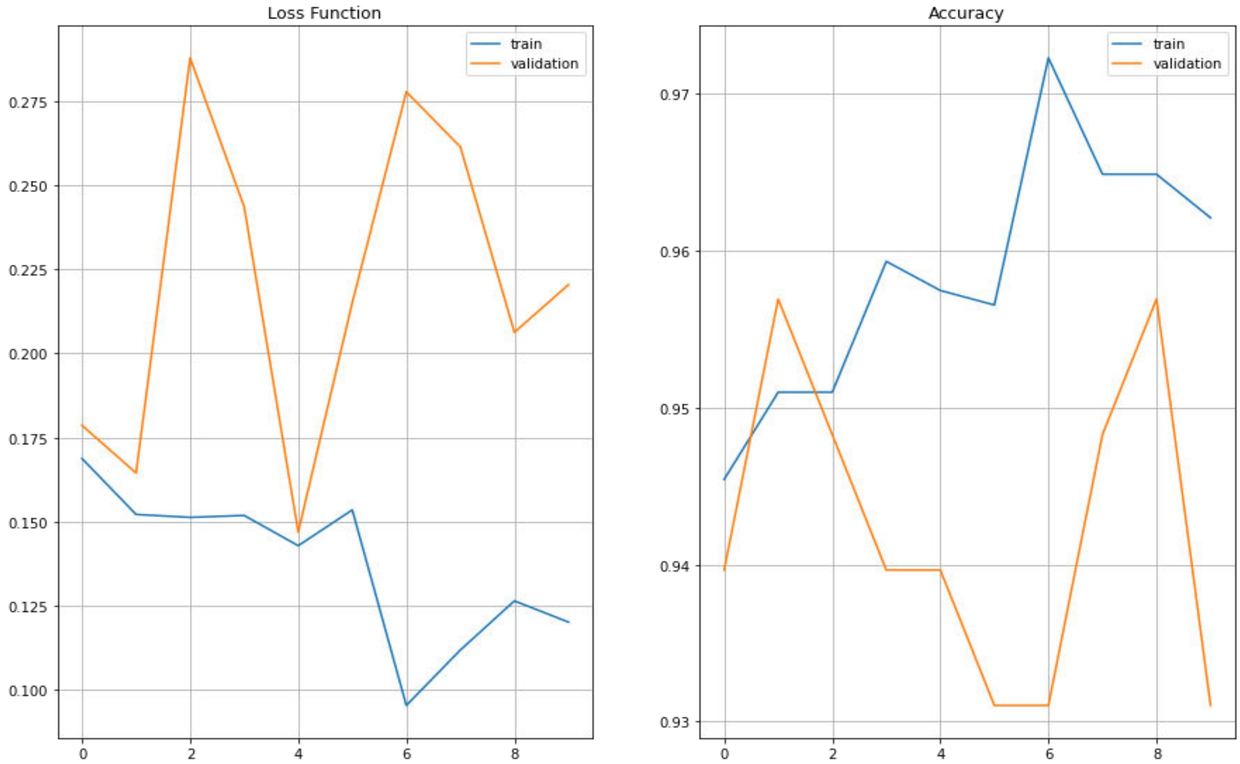


Fig.2. Model Loss Function and Accuracy (Graph)

```
+ Code  + Text
13/13 [==============================] - 35s 3s/step - loss: 0.1896 - accuracy: 0.9315
Epoch 24/50
13/13 [==============================] - 35s 3s/step - loss: 0.2604 - accuracy: 0.9218
Epoch 25/50
13/13 [==============================] - 35s 3s/step - loss: 0.2071 - accuracy: 0.9413
Epoch 26/50
13/13 [==============================] - 36s 3s/step - loss: 0.2560 - accuracy: 0.9267
Epoch 27/50
13/13 [==============================] - 34s 3s/step - loss: 0.1724 - accuracy: 0.9413
Epoch 28/50
13/13 [==============================] - 34s 3s/step - loss: 0.1409 - accuracy: 0.9609
Epoch 29/50
13/13 [==============================] - 34s 3s/step - loss: 0.1675 - accuracy: 0.9487
Epoch 30/50
13/13 [==============================] - 34s 3s/step - loss: 0.1926 - accuracy: 0.9462
Epoch 31/50
13/13 [==============================] - 34s 3s/step - loss: 0.1276 - accuracy: 0.9511
Epoch 32/50
13/13 [==============================] - 34s 3s/step - loss: 0.1625 - accuracy: 0.9462
Epoch 33/50
13/13 [==============================] - 34s 3s/step - loss: 0.1598 - accuracy: 0.9487
Epoch 34/50
13/13 [==============================] - 34s 3s/step - loss: 0.0824 - accuracy: 0.9731
Epoch 35/50
13/13 [==============================] - 35s 3s/step - loss: 0.1002 - accuracy: 0.9658
Epoch 36/50
13/13 [==============================] - 34s 3s/step - loss: 0.0811 - accuracy: 0.9756
Epoch 37/50
13/13 [==============================] - 34s 3s/step - loss: 0.0956 - accuracy: 0.9756
Epoch 38/50
13/13 [==============================] - 34s 3s/step - loss: 0.1083 - accuracy: 0.9658
Epoch 39/50
13/13 [==============================] - 36s 3s/step - loss: 0.0691 - accuracy: 0.9731
Epoch 40/50
13/13 [==============================] - 34s 3s/step - loss: 0.0837 - accuracy: 0.9756
Epoch 41/50
13/13 [==============================] - 34s 3s/step - loss: 0.1148 - accuracy: 0.9584
Epoch 42/50
13/13 [==============================] - 34s 3s/step - loss: 0.1146 - accuracy: 0.9731
✓ 12s completed at 7:04 PM
```

```
+ Code  + Text
13/13 [==============================] - 34s 3s/step - loss: 0.0837 - accuracy: 0.9756
Epoch 41/50
13/13 [==============================] - 34s 3s/step - loss: 0.1148 - accuracy: 0.9584
Epoch 42/50
13/13 [==============================] - 34s 3s/step - loss: 0.1146 - accuracy: 0.9731
Epoch 43/50
13/13 [==============================] - 34s 3s/step - loss: 0.1337 - accuracy: 0.9633
Epoch 44/50
13/13 [==============================] - 34s 3s/step - loss: 0.1247 - accuracy: 0.9584
Epoch 45/50
13/13 [==============================] - 34s 3s/step - loss: 0.1031 - accuracy: 0.9584
Epoch 46/50
13/13 [==============================] - 36s 3s/step - loss: 0.1505 - accuracy: 0.9633
Epoch 47/50
13/13 [==============================] - 34s 3s/step - loss: 0.1497 - accuracy: 0.9462
Epoch 48/50
13/13 [==============================] - 34s 3s/step - loss: 0.1336 - accuracy: 0.9633
Epoch 49/50
13/13 [==============================] - 34s 3s/step - loss: 0.1327 - accuracy: 0.9511
Epoch 50/50
13/13 [==============================] - 34s 3s/step - loss: 0.0766 - accuracy: 0.9756
CPU times: user 54min 45s, sys: 27.5 s, total: 55min 13s
Wall time: 29min 23s
<keras.callbacks.History at 0x7fea1db42250>
```

Fig.3. Model Loss Function and Accuracy

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006095 |

V. RESULTS

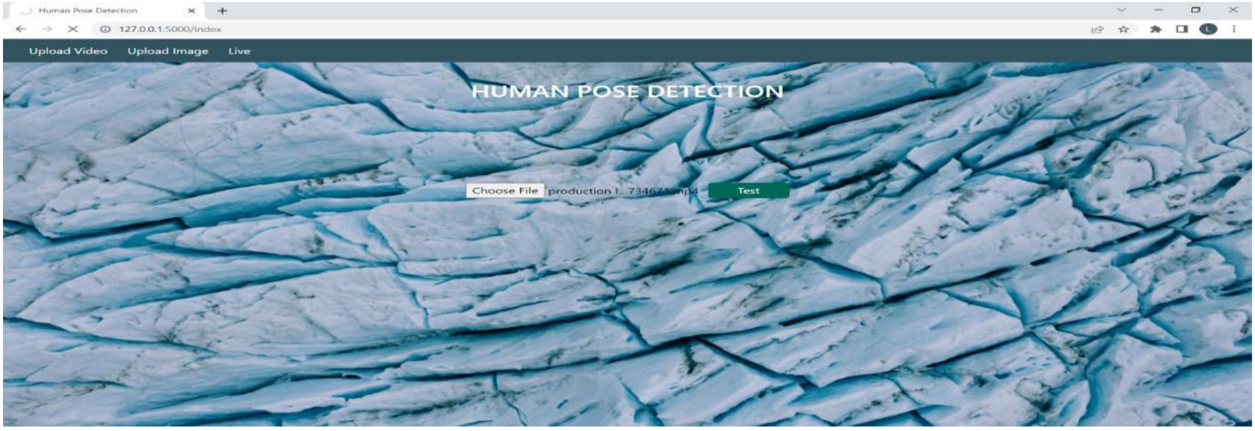


Fig.4 Home Page

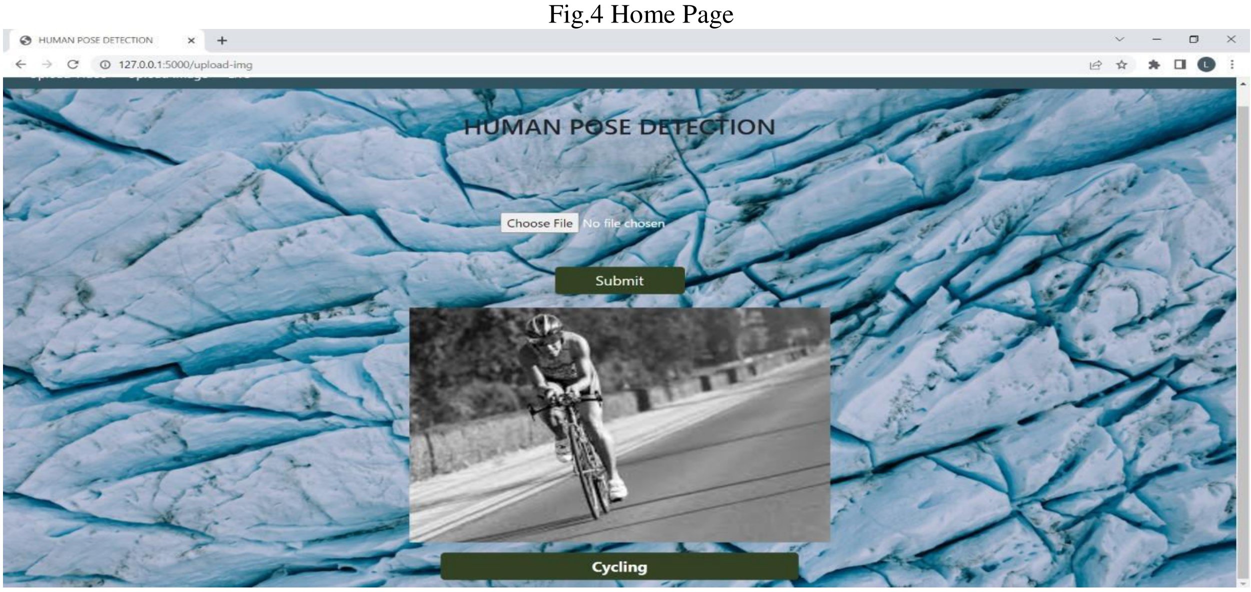


Fig.5. Recognize Cycling Action

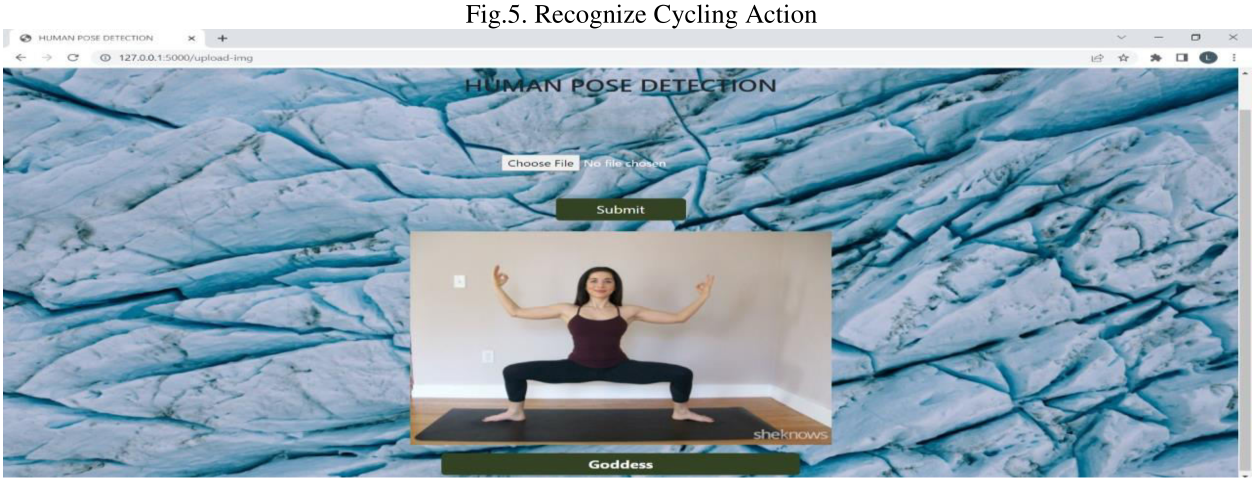


Fig.6.Recognize Goddess Action

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006095 |

VI. CONCLUSION AND FUTURE WORK

In this project we have planned, designed and implemented the system for Human activity recognition system for controlling UI which is a standalone application for controlling the various user interface controls and/or programs like Flak. In the analysis phase we gathered four major actions done by a human on his daily basis whether it be standing, sitting, sleeping and bending and the techniques and algorithms they employ and the success/failure rate of these systems. Accordingly, we made a detailed comparison of these systems and analyzed their efficiency. In the design phase we designed the system architecture diagrams and also the data flow diagram of the system. We studied and analyzed the different phases involved and accordingly designed and studied the algorithms to be used for the same. The image we used is jpg extension to keep things straight forward and code to be clean and easily usable because jpg/png images are easily available.The future scope of our project can be achieved through CCTV cameras fixed on the pillars of a bank which can be a human tracking device that tracks the movements of each person in the bank and captures their actions and if something goes way past the limits it can send data to the manager. Here we gift unique insights on present works and the methodologies utilized by researchers for spotting the human sports.

REFERENCES

1.GangLi,ChunyuLi, LearningskeletoninformationforhumanactionanalysisusingKinect, 2020.
2.Preksha Pareek & Ankit Thakkar ,A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications,2021.
3. FarhatAfza, Muhammad Attique Khan, Muhammad Sharif, SeifedineKadry,A framework of human action recognition using length control features fusion and weighted entropy-variances based feature selection,2021.
4.Khan Muhammad, Mustaqeem, Ami nUllah, Ali ShariqImran, Muhammad Sajjad,Human action recognition using attention based LSTM network with dilated CNN features.,2021.
5.Muhammad Attique Khan, Yu-Dong Zhang, Sajid Ali Khan, Muhammad Attique, AmjadRehman&SanghyunSeo.,A resource conscious human action recognition framework using 26-layered deep convolutional neural network.,2021.
6.A.D. Antar, M. Ahmed, M.A.R. Ahad, Challengesin sensor-based humanactivityrecognitionandacomparativeanalysisofbenchmarkdatasets:Areview,in:2019Joint8thInternationalConferenceonInformatics,Electronics&Vision(ICIEV) and 2019 3rd International Conference on Imaging, Vision & PatternRecognition,IcIVPR,IEEE,2019.
7. K.A.daCosta,etal.,Internetofthings: A survey on machine learning-based intrusion detectionapproaches, Comput.Netw. 151(2019)147-157
8.J.K.Aggarwal,M.S.Ryoo,Humanactivityanalysis:Areview,ACMComput.Surv.43(3) (2011)1-43.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165

doi crossref

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING