iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51960

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Human Action Recognition using Deep Learning

Suhani Gaikwad[1], Rutuja Ghodekar[2], Nikhil Gatkal[3], Atharv Prayag[4]
[1, 2, 3, 4]*BE students, Department of Computer Engineering, TSSM'S Bhivarabai Sawant College of Engineering and Research, Narhe, Pune, Maharashtra, India*

Abstract: ***The aim of this project is to recognize human actions for monitoring and security purposes. This project is mainly focused on building a system that is helpful for doctors to monitor patients .Human Action Recognition is required to recognize a set of human activities by training a supervised learning model and displaying the activity/action result as per the input action received. It has wide range of applications such as patient monitoring system, ATM/ Bank security system, etc. Human Action Recognition model can be mainly used for security and monitoring purposes. We can use various machine learning and deep learning algorithms for this project . One of the best approach is using CNN .***
Keywords: Recognition, monitoring, machine learning, deep learning, CNN

I. INTRODUCTION

Human Action Recognition from videos and live camera is based on analysis of images / frames captured to detect the activity. Human Action Recognition is required to recognize a set of human activities by training a supervised learning model and displaying the activity/action result as per the input action received. The goal of human activity recognition is to examine activities from video sequences or still images[1].

As technology is increasing, the need of security is increasing parallely. ATM security , bank security , home security is a need of concern. To make the security more strong , this project can help.

Also, as population is increasing all around the world, it is equally important to build a system which can help doctors to observe patients.

HAR can be build using machine learning as well as deep learning. With proper training activities such as standing, sitting , drinking, sleeping, waving, bending, etc can be detected.

HAR is required to recognize a set of human activities by training a supervised learning model and displaying the activity result as per the input activity received from the camera input. It involves predicting the movement of a person based on sensor data and traditionally involves deep domain expertise and methods from signal processing to correctly engineer features from the raw data in order to fit a machine learning model. The sensor data may be remotely recorded, such as video, cameras, radar images or other wireless methods. Human action recognition basis for many applications such as video surveillance, health care, and human-computer interaction. To Analyze the activity of a person from the information collected by different devices. Discover which are the variables that determine which activity is doing a person. It is of great importance in various applications in artificial intelligence like video surveillance, computer games, robotic, and human computer interactions. Due to lack of security , there are chances of robbery in society and many places like banks,companies, etc. The main motivation is to build a strong intelligent system which would recognize human actions and thereby increase the security. Design individualized exercise tables to improve the health of a person. To monitor the health of the patient, it can be helpful for the doctors to monitor patient recovery progress[3].

II. OBJECTIVE

Human Action Recognition is the problem of predicting what a person is doing based on a trace of their movement using sensors. It is a challenging problem because there is no clear analytical way to relate the sensor data to specific actions in a general way.

Human action recognition is a field of study that deals with identifying, interpreting, and analyzing the actions specific to the movement of human beings.

1) The goal of action recognition is to recognize common human actions in real life settings. Accurate activity recognition is challenging because human activity is complex and highly diverse. Several probability-based algorithms have been used to build activity models.
2) The main goal is to build a model using deep learning which can recognize human actions and predict the output.
3) This model will thereby increase the security and help to make human life easier.

In this paper , brief explaination of all our work is depicted.

©IJRASET: All Rights are Reserved

III. RELATED WORK

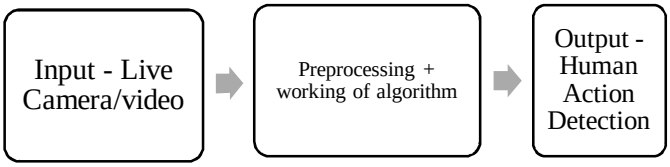
Machine learning and deep learning has gained a huge boom and demand in the recent years. There are n number of ways using deep learning to detect human activity. The most trending and popular approaches is CNN and LSTM. While building this project we particularly focused on monitoring activities such as sleeping, Standing, sitting , drinking, and many more so that it can be helpful for the doctors for analysis of the patient. Deep learning is widely used for image detection and face recognition.

One of the advantage of deep learning is feature extraction. The features are extracted automatically without human intervention. Deep neural network is core net in Deep learning.

Deep neural network consists of multiple layers of interconnected neurons that process the data and learn from it to make predictions or classifications. Convoluntional Neural Network is used for classification because it can automatically learn to extract relevant features from input data and map them to the correct output class. The input layer is where the model gets input that is the raw data and the processing i.e feature extraction, calculations , pattern recognition etc is done by the hidden layers. The layers closest to the output are known as output layers[4].

IV. PROPOSED SYSTEM

Various approaches were performed for human action detection. The flow of our project is as follows:



Convoluntional Neural Network is fed using images as input, it is a set of layers that transform image to set of class probabilities. Once the input video is fed , the set of images is extracted from the input and using the Neural Network(CNN,DNN) algorithm the further working is done and than the images is classified into a set of particular classes and the probability of the class is shown on the screen as the output.

The UI of the project is as follows :

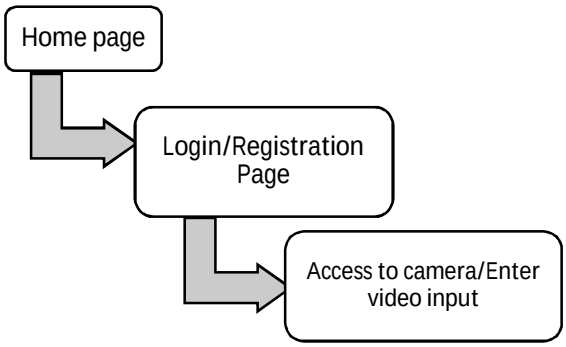
Firstly is the login and registration page of the security purposes. As keeping in consideration that we are building this project for the monitoring purposes and security purposes as well, the aim is to ensure security from the initial step itself.

For example only a particular doctor can have access to the patient monitoring system. Therefore, a person needs to register first and then login to access the camera , the input to the project is the live web camera and video. The action that is performed is captured by the camera and using logistic regression the action is detected. For video, the input is fed and the images are extracted , further calculations and feature extraction is performed by the algorithm (CNN) and lastly the prediction is shown on the screen i.e the probabitlity class of the image (standing, sleeping, sitting). In this way the output is given.

The features here are face landmarks, Hand landmarks, leg landmarks i.e the landmarks of the body skeleton. In this approach, the input features are passed through multiple layers to extract particular class feature and then machine learning model such as logistic regression, multiclass classifier is used for probability classification of classes powered by deep learning algorithms.

In the home page admin has to login by valid user name and password. The aim behind this is to allow authorized users only.After successful login , the person can access the input (camera) and view the results.

The flow of the UI is :



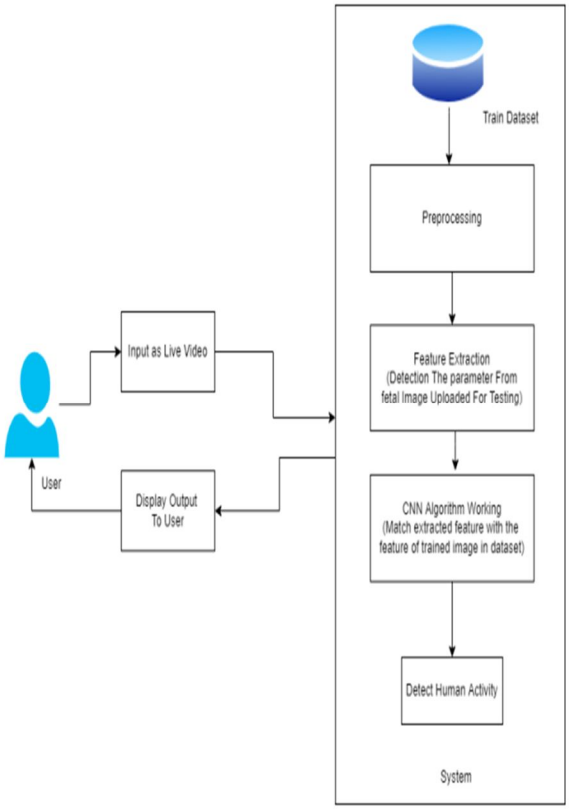
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Regression models are supervised learning algorithms. The model is firstly trained by providing input. For standing action, the standing position data is trained and Deep learning algorithm extracts the features of body skeleton and then they assign probability classes with the help of Supervised learning models. In this way output is predicted. In simple words machine learning models powered with deep learning concepts makes prediction more accurate. To extract features of a particular activity the landmarks of human skelotons pass through n number of layers , or simply a pipeline.

We used logistic regression in this project as it gives more accurate results. In this hybrid model , for live camera the classification is performed by the logistic regression.

When for detecting actions using video dataset CNN algorithm is used. Convoluntional Neural networks has capabilities to extract feature and also backtrack to get more optimized prediction/classification. This architecture can be modified and optimized based on the specific needs of the application and available sensor data from live video camera (Web camera)

Proposed system architecture:



V. WORKING

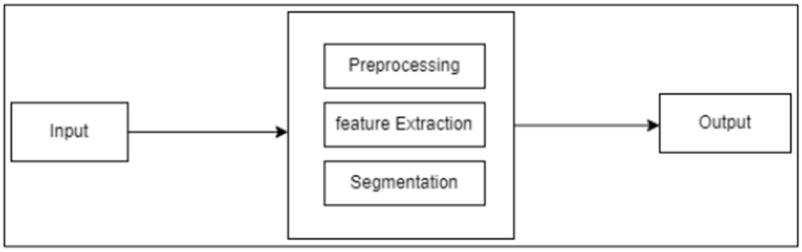
Working or practical implementation is a crucial factor in ones's project. The idea should be implemented properly for the project to get successful and used by end-users.

The Internal Working has Five layers which are input, preprocessing, feature extraction, segmentation and output.

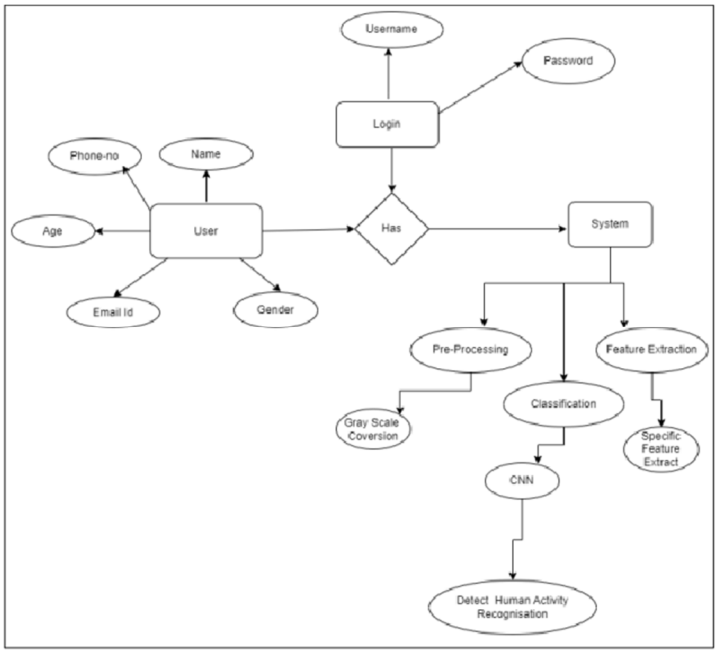
1) *Input layer*- Input layer gets the raw data from the sensors or cameras.
2) *Hidden layers*-Hidden layers perform the internal complicated and complex operations. There can be any number of hidden layers depending upon the project title. Preprocessing, feature extraction, segmentation is done by the hidden layers.

a) *Preprocessing*-It includes preparing of data useful for the model
b) *Feature extraction*-It refers to the process of transforming raw data into numerical features that can be processed while preserving the original data.
c) *Segmentation*-Segmentation is a technique of splitting the data
d) *Classification* – Classifies the images according to the activity.

3) *Output layer*- Output layer shows the output i.e the detected activity.



The detailed working of the system is depicted by the below given entity diagram. Each step along with the operations is depicted in the diagram. Preprocessing involves Gray scale conversion. Feature extraction and classification is done by the Deep learning algorithm (CNN). The result of the algorithm is shown as output, it means the activity is detected.



This system accepts input in the form of Live camera and video dataset. We know that we're doing data processing and trained dataset on the system, therefore we're employing modules: Pre-processing, Feature extraction, and classification, all of which use our CNN algorithm[2].

For video dataset:

So First Input as a video dataset, then pre-processed the dataset (pre-processing step is clean the Image and Remove Blur part) After that, then the system extracts the parameters or features of the in the extraction section. Then, in classification, where we utilize our CNN algorithm to Recognise Sign Language.

These are the steps used to training the CNN.

• Fetch the trained model from Layer.

• Loading the image of the same size as the one used in the training images. Convert the image into an array.

• Transform the numbers in the array to be between 0 and 1 by dividing by 255. These are the steps used to training the CNN.

• Steps:

Step 1: Upload Dataset.

Step 2: The Input layer.

Step 3: Convolutional layer.

Step 4: Pooling layer.

Step 5: Fully Connected

For live camera:

Logistic regression is used for detecting output. Logistic regression is supervised learning algorithm which finds relation between two data factors.

The below pictures depicts the output . The pink lines are the parameters or simply body skeleton landmarks.

When we do the action in front of camera, the images of the poses get sxtracted and the algorithm works and thus, we get the output as the action recognized.



VI. FUTURE SCOPE

HAR is rapidly growing field with many exciting future possibilities . From raw sensor data Deep learning models learns and predict probable outcomes. Models can be use for wider range of activity recognition with high level accuracy . This will help in particular domain to automate the human tasks for ex:- Hospitals, Banks, Schools, etc

To monitor the patients activity , it has high demand. Sensor(Camera) can be placed in patient rooms which can be useful for the doctor to monitor the patient's activities including movements, vital signs, sleep patterns.It will help healthcare providers to detect changes in patients health .

The future scope for HAR is casr,and new application are being developed all the time .With the advent of new sensors and machine learning techniques ,HAR has the potential to transform many different industries and improve our daily lives in numerous ways.

VII. CONCLUSION

Thus, in this project we have build a hybrid system for human action recognition which includes to types of input fields , one is the live camera and other is the input field. For live camera, logistic regression algorithm is used and for input video CNN algorithm is used.

Deep learning techniques have great potential for human activity recognition. In this project, a neural network (NN) based approach for classification and evaluation of human activities has been explored.

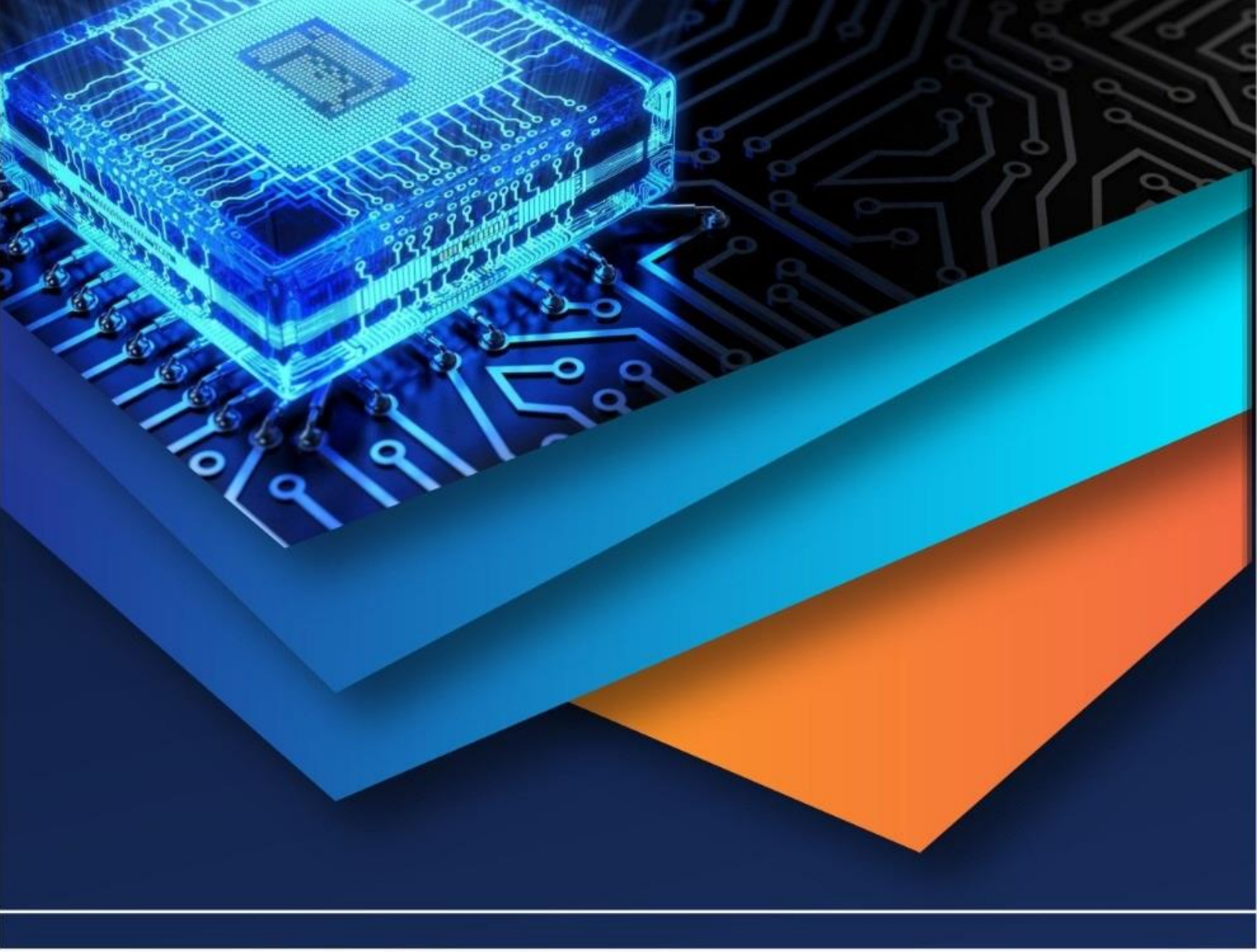
The model will be able to identify the human actions accordingly.

In this paper, we presented a hybrid method to effectively perform human action recognition.(Machine learning + Deep learning). Our method first identifies the user. For action recognition we used CNN (Convolutional Neural Network) that is Deep learning algorithm. We have proposed a Human Activity Identification system based on pose estimation and convolutional neural network. This system will combine the results of the 3D pose estimation model with the 1D convolutional neural network for better and more accurate detailed result generation.

REFERENCES

[1] Deep learning for Human Action Recognition- R. U. Shekokar, S. N. Kale, 2021 6th International Conference for Convergence in Technology (I2CT),IEEE,2021
[2] Learn Computer Vision Using OpenCV - With Deep Learning CNNs and RNNs | Sunila Gollapudi | Apress. .
[3] W. Sultani, C. Chen, and M. Shah, "Real-world Anomaly Detection in Surveillance Videos," ArXiv180104264 Cs, Feb. 2019.
[4] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, "Learning realistic human actions from movies," in 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8, doi: 10.1109/CVPR.2008.4587756.
[5] M. Jain, MrinalJain17/Human-Activity-Recognition. 2019.
[6] "Keras vs TensorFlow vs PyTorch | Deep Learning Frameworks," Edureka, 05-Dec-2018. https://www.edureka.co/blog/keras-vstensorflow-vs-pytorch/.
[7] Chen C, Jafari R, Kehtarnavaz N. Improving Human Action Recognition Using Fusion of Depth Camera and Inertial Sensors[J]. IEEE Transactions on Human-Machine Systems, 2015, 45(1):51-61
[8] Chen D, Yang J, Wactlar H D. Towards automatic analysis of social interaction patterns in a nursing home environment from video[C]
[9] Chen L, Nugent C D, Wang H. A Knowledge-Driven Approach to Activity Recognition in Smart Homes[J]. IEEE Transactions on Knowledge Data Engineering, 2012, 24(6):961-974.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)