

Internship Project Report

# HUMAN ACTIVITY RECOGNITION WITH OPENCV

*A Project Report Submitted to the Partial fulfilment of the Requirements for the  
Award of the degree of Master of Science in Data Science*

Submitted by

P. DEEPAN  
(22PDS801)

Under the guidance of

Dr. ARUL KUMAR, M.Sc., M.Phil., Ph.D.,SET



DEPARTMENT OF DATA SCIENCE

ST. JOSEPH'S COLLEGE (AUTONOMOUS)

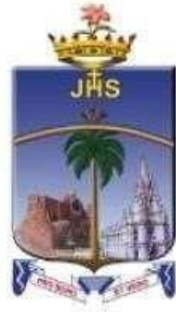
Accredited at "A++" Grade (4th Cycle) by NAAC

College with Potential for Excellence by UGC

Tiruchirappalli – 620 002.

JUNE - 2023

**DEPARTMENT OF DATA SCIENCE  
ST. JOSEPH'S COLLEGE  
(AUTONOMOUS)  
TIRUCHIRAPPALLI - 620 002**



**CERTIFICATE**

This is to certify that the **INTERNSHIP PROJECT REPORT** submitted by **DEEPAN (22PDS801)** is a record of original training undergone by him during the period from 15–MAY–2023 to 15–JUNE–2023, in partial fulfilment of the requirements for the award of the degree of **Master of Science in Data Science**.

**Internal Guide**

**Head of the department**

Submitted for Internship Project Viva-voce examination is conducted on \_\_\_\_\_

**Internal Examiner**

**External Examiner**

Date:

Date:

## **DECLARATION**

I hereby declare that the Summer Internship project report entitled “HUMAN ACTIVITY RECOGNITION WITH OPENCV” is an authentic record of my original training undergone by me during the period from 15–MAY–2023 to 1–JUNE–2023 under the supervision and guidance of Dr. V. ARUL KUMAR, M.Sc., M.Phil., Ph.D., SET, for the award of the degree of Master of Science in Data Science.

I also confirm that the report is only prepared for my academic requirement, not for any other purpose.

DEEPAN P

(22PDS801)

Department of Data Science

St. Joseph's College (Autonomous)

Tiruchirappalli

# INTERNSHIP ACCEPTANCE LETTER



Date: - May 5<sup>th</sup>, 2023

DEEPAN P  
St.joseph's college,trichy.  
22pds801

Dear DEEPAN P,

We are excited to extend an offer to you as a Trainee Software Engineer position at **AiROBOSOFT Products and Services**, Affiliated with **AICTE** India. You are requested to report on or before **8<sup>th</sup> May 2023**.

In this role, you will report directly to your assigned team leads.

During your internship with AiROBOSOFT, you may have access to trade secrets and confidential or proprietary business information belonging to AiROBOSOFT Products and Services, by accepting this offer, you acknowledge that this information must remain confidential and agree to refrain from using it for your own purposes or disclosing it to anyone outside of AiROBOSOFT Products and Services.

We look forward to having you begin your career at AiROBOSOFT Products and Services and wish a successful internship, Welcome to our team!



*Our Vision, Your Future.*

Regards,  
**CEO & Chief Ai Engineer at AiROBOSOFT.**

I accept the offer with the company on the terms and conditions set out in the letter.

DEEPAN P

**AiROBOSOFT Products and Services LLP**



No - 4, 3rd Floor, 5th A Main Rd, adjacent to  
Bangalore Baptist Hospital, Hebbal, Bengaluru -  
560024.



airobosoft.com



hr@airobosoft.com



99644 83 843  
88843 99 089

# INTERNSHIP COMPLETION CERTIFICATE

Reg. No: AAS7147



## Internship Certification

This is to certify that

**DEEPAN P**

Has completed internship at **AIROBOSOFT PRODUCTS AND SERVICES LLP**

On

**Artificial Intelligence and Machine Learning**

**May 15<sup>th</sup> 2023 – June 15<sup>th</sup> 2023**

Her / His performance was Very Good.



(Scan to Authenticate)



  
**Managing Director**



## ACKNOWLEDGEMENT

It is with immense pleasure that we present our first venture in the field of data science in the form of a project work. First of all, I am indebted to the Almighty for his choice blessing showered on me in completing this endeavor.

I extend my gratitude to **Rev. Dr. K. AMAL SJ**, our secretary, St. Joseph's college (Autonomous), Tiruchirappalli-620 002 for giving me this opportunity.

I express my sincere thanks to **Rev. Dr. M. AROCKIASAMY XAVIER SJ**, our principal, St. Joseph's college (Autonomous), Tiruchirappalli-620 002 for having given me an opportunity to study in this Institution.

I express my sincere gratitude to **Dr. P. RAJENDRAN**, our Deputy Principal, St. Joseph's college (Autonomous), Tiruchirappalli-620 002 for providing me with this opportunity to complete my project.

I would also like to acknowledge **Dr. L. AROCKIAM**, Head, Department of Data Science, St. Joseph's college (Autonomous), Tiruchirappalli-2, for this moral support and encouragement he has rendered throughout the course.

I am highly indebted to the internal guide **Dr. V. ARUL KUMAR**, Assistant Professor, Department of Data Science for his guidance and constant supervision as well as for providing necessary information regarding the project and also for his support in completing the project.

It is my radiant sentiment to place on record my best regards, and deepest sense of gratitude to **Dr. M. KRIUSHANTH**, Assistant professor, **Dr. I. PRIYA STELLA MARY**, Assistant professor, **Dr. A. BEATRICE DOROTHY**, Assistant Professor, **Mr. SYED ASIM AHMED** Managing Director of AI ROBOSOFT TECHNOLOGY for his precious guidance which was extremely valuable for my study both theoretically and practically

Sincerely  
DEEPAN P  
22PDS801

## **ABSTRACT**

Human activity recognition using OpenCV and deep learning is an emerging field that aims to automatically classify and identify human activities from video data. This project focuses on developing a system that leverages the power of deep learning models and the capabilities of OpenCV to accurately recognize and classify various human activities.

The project involves acquiring or training a deep learning model suitable for activity recognition, preprocessing video frames using OpenCV's functions, and extracting relevant features for input to the model. The trained model is then utilized to classify the activities performed in the video stream or pre-recorded videos.

The predicted activity labels are visualized on the video frames Human Action Recognition is based on a model CNN for feature extraction. Convolutional neural networks (CNN) is a technique of deep learning. Most convolutional neural networks used for recognition task are built using convolution and pooling layers followed by a few numbers of fully connected layers and identifying similar patterns in an interval to recognize the action by providing accuracy based on the task

## TABLE OF CONTENT

<b>SL.NO</b>	<b>CONTENT</b>	<b>PAGE.NO</b>
	LIST OF FIGURES	
1	INTRODUCTION	1
2	COMPANY PROFILE	3
3	PROBLEM STATEMENT	4
4	BACKGROUND STUDY	5
5	TECHNOLOGY ADOPTED	8
6	DETAILS OF TOOL USED	9
7	SOLUTION FOR THE PROBLEM	11
8	CONCLUSION	15
	REFERENCES	
	ANNEXURE 1: PROGRAM	



## LIST OF FIGURES

FIG:NO	NAME	PAGE.NO
4.1	KINETIC DATASET	8
7.1	HUMAN ACTIVITY RECOGNITION WORKFLOW	12
7.2	OUTPUT OF WEBCAM	14
7.3	OUTPUT OF INPUT VIDEO	14

# **CHAPTER 1**

## **INTRODUCTION**

Human activity recognition is a significant task in computer vision and artificial intelligence that aims to automatically identify and classify human activities from visual data, such as videos or image sequences. It plays a crucial role in various applications, including video surveillance, human-computer interaction, healthcare monitoring, and sports analysis.

In recent years, deep learning techniques have revolutionized the field of computer vision, providing powerful tools for human activity recognition. OpenCV, an open-source computer vision library, combined with deep learning models, offers a flexible and efficient framework for developing accurate human activity recognition systems.

Deep learning models, such as Convolutional Neural Networks (CNNs) can effectively capture temporal and spatial information from video data, allowing them to learn complex patterns and discriminate between different activities. OpenCV provides a rich set of functions and tools for video processing, feature extraction, and visualization, making it an ideal choice for implementing human activity recognition.

The objective of this project is to develop a human activity recognition system using OpenCV and deep learning. The system should be able to classify various human activities accurately, such as walking, running, sitting, standing, and performing specific actions. By leveraging deep learning models and OpenCV's capabilities, the system can analyse video streams or pre-recorded videos in real-time or offline scenarios.

I have used the kinetic dataset is used by the model to compare similar patterns in the input data frames that are captured in the intervals. The similar patterns can be identified by CNN through pooling layer by layer. The identified actions are categorized into classes of human activities. The recognition of the data input can be done by Resnet\_34 model by video classification of 3D kernels. Segmentation of the actions

are the classes which are identified by the model.

The project involves several key components, including acquiring or training a suitable deep learning model for activity recognition, preprocessing video frames for model input, extracting relevant features, performing classification using the trained model, and visualizing the predicted activities on the video frames.

Additionally, the system may include evaluation procedures to assess the model's performance by comparing the predicted activity labels with ground truth labels. This evaluation can help validate the system's accuracy and identify areas for improvement.

By implementing a human activity recognition system using OpenCV and deep learning, we can unlock various applications in diverse domains. The system can contribute to enhancing surveillance systems by automatically detecting suspicious or abnormal activities. It can assist in healthcare monitoring by analyse patients' movements and identifying potential risks or anomalies.

## CHAPTER 2

### COMPANY PROFILE



AIROBOSOFT is a Bangalore based IT firm Leading a team of Data Scientist, Robotics & Electronics Engineers, experts in Machine Learning and more, collaborated together to work on fascinating futuristic technologies.

Industry: IT Services and IT Consulting. Headquarters: Bengaluru, Karnataka

Specialties: Robotics, Artificial intelligence, Machine Learning, IOT, and Software development.

Website: <http://www.airobosoft>

Employees :180

Address: No – 2, 3rd Floor 5th A Main Rd, Besides Bangalore Baptist Hospital Vinayakanagar, Hebbal Bengaluru 560024

Design a human activity recognition system using OpenCV that can accurately classify and distinguish between normal and abnormal activities in real-time video footage. The system should be capable of analyzing various human actions and movements, such as walking, running, sitting and gestures, and determining whether they fall within expected normal patterns or deviate significantly from them.

The system should utilize computer vision techniques and machine learning algorithms to extract relevant features from video frames and perform classification based on those features.

The goal is to develop a reliable and efficient system that can aid in monitoring and surveillance applications, identifying potentially dangerous or suspicious behaviors, and alerting security personnel or relevant authorities in case of abnormal activities. The system should provide accurate and timely notifications, reducing false alarms and ensuring quick response to potential threats.

## **CHAPTER 4**

### **BACKGROUND WORK**

Human activity recognition using OpenCV and deep learning. It uses a model to recognize human activities in a video stream. The parsing command-line arguments, including the path to the trained model and the class labels file. It then loads the model and accesses the video stream from either a file or the webcam. The frames from the video stream are processed by resizing them and creating a blob for input to the model.

The model performs inference on the frames, and the predicted activity label is determined. If certain abnormal activities, such as smoking or drinking are detected, a sound notification is generated and also send the email message. The predicted activity label is displayed on each frame, and the processed frames are shown in a window. The background work involves loading the model, processing frames, performing inference, and visualizing the predicted activities in real-time video.

**Kinetics 400 Dataset:**The Kinetics 400 dataset is a large-scale video dataset that provides a diverse set of human actions. It consists of approximately 400 action classes, ranging from simple everyday activities to complex interactions. Each action class contains multiple video clips, resulting in a substantial amount of training data. The dataset offers a wide variety of scenarios, capturing activities from sports, household chores, to social interactions. By utilizing the Kinetics 400 dataset, we can train our models to recognize a broad range of human activities.

**OpenCV and Deep Learning:** OpenCV is a powerful computer vision library that provides a wide range of functions and algorithms for image and video processing. It offers robust tools for video capturing, frame manipulation, and feature extraction, making it an ideal choice for our project. Deep learning, a subfield of machine learning, has shown remarkable success in various computer vision tasks. By employing deep neural networks, we can learn complex patterns and representations from video data, enabling us to recognize and classify human activities more accurately.

Project flow: Our project will follow a systematic workflow to achieve our goal of human activity recognition. Firstly, we will acquire the Kinetics 400 dataset, which includes video clips annotated with corresponding activity labels. We will preprocess the dataset by splitting it into training, validation, and testing sets, ensuring that our model is trained and evaluated on different subsets of data.

Next, we will employ OpenCV to extract meaningful features from the video clips. Techniques such as optical flow, frame differences, or using pre-trained deep learning models can be applied to extract relevant information from the video frames.

With the extracted features, we will build a deep learning model using popular frameworks such as TensorFlow or PyTorch. Convolutional Neural Networks (CNNs) can be utilized as they have demonstrated excellent performance in video-based tasks. We will train the model on the training set and fine-tune it to optimize its performance.

To evaluate our model, we will assess its accuracy the validation and testing sets.

Finally, we will deploy our trained model to recognize human activities in real-time or on new video data. By integrating OpenCV for video capturing and processing, we can classify activities in a practical setting.

## Dataset

Kinetics dataset.

It consists of:

- 400 human activity recognition classes
- A total of 300,000 videos

Classes:

- |                     |                    |
|---------------------|--------------------|
| ➤ applying cream    | smoking            |
| ➤ kissing           | drinking           |
| ➤ driving car       | reading book       |
| ➤ reading newspaper | eating burger      |
| ➤ eating cake       | dunking basketball |
| ➤ eating watermelon |                    |



Fig 4.1: Kinetics Dataset

## MODEL SELECTION

Deep Learning model architecture for activity recognition, such as Convolutional Neural Networks (CNNs). Video analysis classification works by labeling a video clip. CNN (Convolutional neural network) application will be the best for you. It will tell you what activities are happening in the video, what content is included in the video.



## **CHAPTER 5**

### **TECHNOLOGY ADOPTED**

#### **5.1 Deep learning**

Deep learning is a subfield of machine learning that focuses on the development and application of artificial neural networks inspired by the structure and function of the human brain. It involves training neural networks with multiple layers of interconnected nodes, known as artificial neurons or perceptron to learn representations of data and make predictions or decisions.

#### **5.2 Image processing**

Image processing refers to the manipulation and analysis of digital images using various algorithms and techniques. Image processing finds applications in various fields, including computer vision, medical imaging and digital photography.

## **CHAPTER 6**

### **DETAILS OF TOOL USED**

#### **6.1 Python**

Python is a general-purpose interpreted, interactive, object-oriented, and high level programming language.

#### **6.2 Anaconda Navigator**

Anaconda Navigator is a graphical user interface (GUI) included with the Anaconda distribution, which is a popular Python distribution for data science and machine learning. Anaconda Navigator provides an easy way to manage and launch applications, environments, and packages related to data science projects.

#### **6.3 Visual Studio Code**

Visual Studio Code (often referred to as VS Code) is a lightweight and highly customizable source code editor developed by Microsoft. It is designed to provide a streamlined and efficient coding experience for developers.

#### **6.4 Python Libraries**

##### **Open cv**

OpenCV (Open Source Computer Vision Library) is a popular open-source computer vision and machine learning library. It provides a comprehensive set of functions and algorithms to help developers and researchers work with images and videos, perform various computer vision tasks, and build computer vision applications.

##### **NumPy**

Numpy is the fundamental package for scientific computing in Python. NumPy arrays facilitate advanced mathematical and other types of operations on large numbers of data.

## Argparse

Argparse is a module in Python's standard library that provides a convenient way to parse command-line arguments. It simplifies the process of creating command-line interfaces (CLIs) for Python programs by automatically generating help messages, handling argument validation, and providing other useful features.

## Imutils

Imutils is a Python library that provides a set of utility functions to simplify common image processing tasks. It is built on top of OpenCV (Open Source Computer Vision Library) and aims to make image processing operations more straightforward and concise.

## Sys

sys is a module in Python's standard library that provides access to some variables and functions that interact with the Python interpreter and its environment. It provides a way to interact with the runtime system and perform various system-related operations.

## Winsound

The winsound module in Python is used to play simple sound effects and tones on Windows operating systems. It provides a basic interface for controlling sound playback.

## **CHAPTER 7**

### **SOLUTION FOR THE PROBLEM**

Develop a human activity recognition system using OpenCV and deep learning to classify real-time video data into different human activities. The system should accurately recognised normal activities such as walking, running, sitting and also recognised the abnormal activities breaking glass, climbing and smoking. The goal is to create a robust and efficient solution that can be applied in different domains, such as surveillance, healthcare.

I have taken Kinetics dataset and training process

#### 1)Data Preprocessing:

- Load video clips and extract individual frames from them.
- Perform resizing, normalization, and any desired encoding or feature extraction on the frames.
- Split the dataset into training, validation, and testing sets.

#### 2)Data Augmentation:

- Implement functions that apply random transformations to the video frames.

#### 3)Model

Convolutional Neural Network (CNN) is a specialized type of artificial neural network designed for processing visual data. CNNs excel at capturing spatial hierarchies and patterns in images and videos. They learn and extract relevant features automatically from raw input data, using convolutional layers to detect local patterns, activation functions to introduce non-linearity, and pooling layers to down sample and improve robustness. Fully connected layers combine low-level features to form high-level representations, and regularization techniques like dropout prevent overfitting. CNNs are trained by optimizing parameters to minimize a loss function, and they have greatly advanced computer vision tasks such as image classification and object detection.

#### 4)Model evaluation

- Periodically evaluate the model's performance on a validation set using the trained parameters.
- Calculate relevant metrics such as accuracy.

Here is the process for HUMAN ACTIVITY RECOGNITION

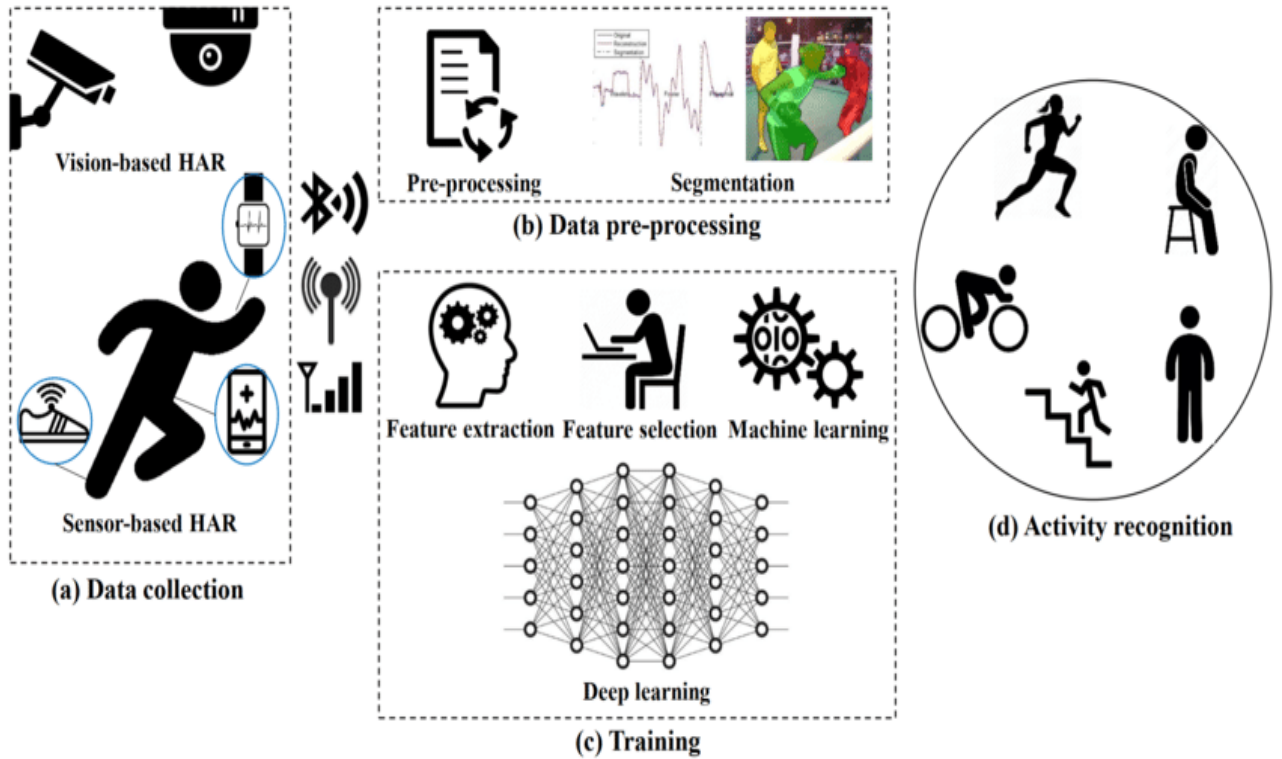


Fig7.1: Human activity recognition workflow

#### Pre-Processing:

The process leads on with the first step of importing necessary packages of numpy, argparse, imutils, sys, opencv2 and winsound after which the construction of the argument parser to parse the arguments.

#### Video Grabbing:

The video data from the dataset or recorded surveillance videos is taken into consideration. It refers to the process of obtaining or extracting a video file from a dataset or source. It involves accessing the video file and saving it to a local location for further processing, analysis, or use in applications.

#### Construction of Frames/blobs:

A list called frames is initialized to store the resized frames and each frame is read from the video stream and resized to a width of 400 pixels using `imutils.resize`. The collected frames are then used to create a blob, which is a formatted representation of the frames suitable for input to the human activity recognition model.

#### Testing the Frames:

After loading the contents of the class label the processes frames from a video stream, applies a human activity recognition model to predict activities in real-time, and provides visual feedback by displaying the frames with the predicted labels. It also includes a feature to detect abnormal activities and trigger a sound alert.

#### Deployment and Application:

The kinetic dataset model to compare similar patterns in the input data frames that are captured in the intervals. Deploy the activity recognition system in the desired environment, such as surveillance systems, smart homes, or healthcare applications, where real-time human activity recognition is required.

## Human activity recognized by webcam

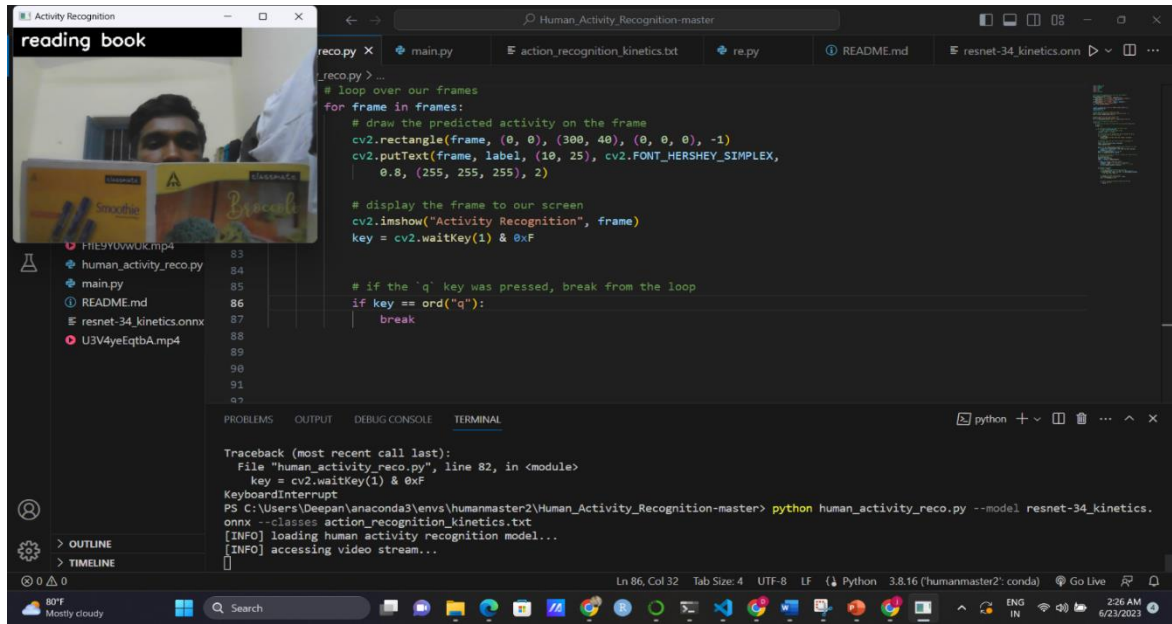


Fig 7.2: webcam recognized reading book

## Human activity recognized with input videos

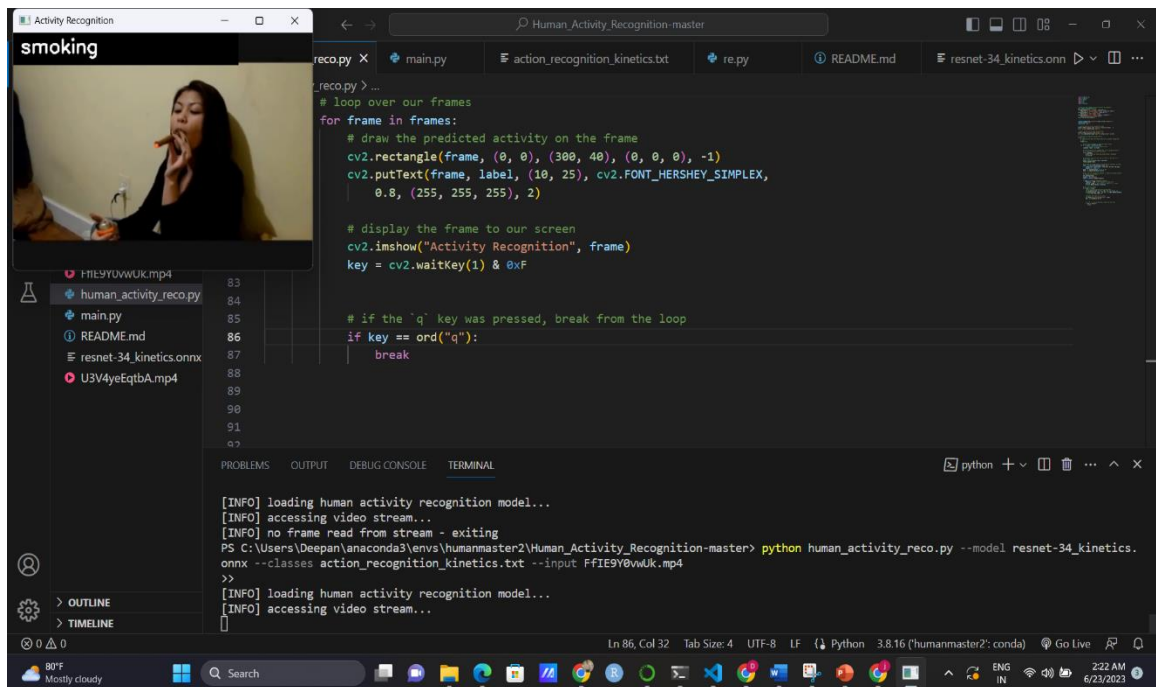


Fig 7.3: From input video it recognized smoking

Smoking is the abnormal activity so it will give the beep sound notification.

## **CHAPTER 8**

### **CONCLUSION**

In this project OpenCV and deep learning techniques to develop a human activity recognition system using the Kinetics 400 dataset. By extracting meaningful features from video data and training a deep learning model, I aim to accurately classify a wide range of implementing real-time video processing, detecting abnormal activities, and deploying the system. By following this approach, the system can accurately classify and distinguish between normal and abnormal activities in real-time video footage, providing valuable insights for surveillance, security, and safety applications.

The outcomes of this project can have significant implications in various domains, including video surveillance, healthcare monitoring, and interactive systems.



## REFERENCES

- 1)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-
- 2) P. A. Dhulekar, S. T. Gandhe, A. Shewale, S. Sonawane and V. Yelname, "Motion estimation for human activity surveillance," 2017 International Conference on Emerging Trends & Innovation in ICT (ICEI), 2017, pp. 82-85, doi: 10.1109/ETIICT.2017.7977015.
- 3)S. Zaidi, B. Jagadeesh, K. V. Sudheesh and A. A. Audre, "Video Anomaly Detection and Classification for Human Activity Recognition," 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), 2017, pp. 544- 548, doi: 10.1109/CTCEEC.2017.8455012.
- 4) P. Yadav, P. Regundwar, A. Wyawahare, P. Pawar and J. Madake, "An Intelligent System to Detect Violent Mob Activities," 2020 IEEE 17th India Council International Conference (INDICON), 2020, pp. 1-10, doi: 10.1109/INDICON49873.2020.9342270.
- 5)U. M. Kamthe and C. G. Patil, "Suspicious Activity Recognition in Video Surveillance System," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697408. [5] Taha, Ahmed "Human Activity Recognition for Surveillance Applications." International Conference on Information Technology, 2015
- 6) Neha Gaba, Neelam Barak and Shipra Aggarwal, "Motion Detection, Tracking and Classification for Automated Video Surveillance", IEEE 1 st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), pp. 1-5, 2016

## ANNEXURE 1

### PROGRAM

```
import numpy as np
import argparse
import imutils
import sys
import cv2
import winsound

# construct the argument parser and parse the arguments
ap = argparse.ArgumentParser()
ap.add_argument("-m", "--model", required=True,
    help="path to trained human activity recognition model")
ap.add_argument("-c", "--classes", required=True,
    help="path to class labels file")
ap.add_argument("-i", "--input", type=str, default="",
    help="optional path to video file")
args = vars(ap.parse_args())

CLASSES = open(args["classes"]).read().strip().split("\n")
SAMPLE_DURATION = 16
SAMPLE_SIZE = 112

# load the human activity recognition model
print("[INFO] loading human activity recognition model...")
net = cv2.dnn.readNet(args["model"])
```

```

# grab a pointer to the input video stream
print("[INFO] accessing video stream...")
vs = cv2.VideoCapture(args["input"] if args["input"] else 0)

# loop until we explicitly break from it
while True:
    # initialize the batch of frames that will be passed through the
    # model
    frames = []

    # loop over the number of required sample frames
    for i in range(0, SAMPLE_DURATION):
        # read a frame from the video stream
        (grabbed, frame) = vs.read()

        # if the frame was not grabbed then we've reached the end of
        # the video stream so exit the script
        if not grabbed:
            print("[INFO] no frame read from stream - exiting")
            sys.exit(0)

        # otherwise, the frame was read so resize it and add it to
        # our frames list
        frame = imutils.resize(frame, width=400)
        frames.append(frame)

```

```

# now that our frames array is filled we can construct our blob
blob = cv2.dnn.blobFromImages(frames, 1.0,

    (SAMPLE_SIZE, SAMPLE_SIZE), (114.7748, 107.7354, 99.4750),
    swapRB=True, crop=True)
blob = np.transpose(blob, (1, 0, 2, 3))
blob = np.expand_dims(blob, axis=0)

# pass the blob through the network to obtain our human activity
# recognition predictions
net.setInput(blob)
outputs = net.forward()
label = CLASSES[np.argmax(outputs)]

    if label in["smoking","applying cream"]:
        frequency = 2500 # Set Frequency To 2500 Hertz
        duration = 1000 # Set Duration To 1000 ms == 1 second
        winsound.Beep(frequency, duration)
# loop over our frames
for frame in frames:
    # draw the predicted activity on the frame
    cv2.rectangle(frame, (0, 0), (300, 40), (0, 0, 0), -1)
    cv2.putText(frame, label, (10, 25), cv2.FONT_HERSHEY_SIMPLEX,
        0.8, (255, 255, 255), 2)
    # display the frame to our screen
    cv2.imshow("Activity Recognition", frame)
    key = cv2.waitKey(1) & 0xFF
    # if the `q` key was pressed, break from the loop
    if key == ord("q"):
        break

```