Project Report On

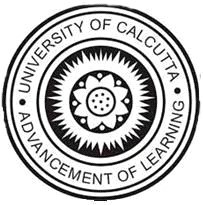
GESTURE DETECTION USING PYTHON

Name: Indranil Kundu & Argha Dutta Roll No.- 193115-21-0044 & 193115-21-0051

Reg.No. - 115-1111-0431-19 & 115-1114-0425-19

A report submitted in partial fulfilment of the degree of

**B.Sc (Hons) in ComputerScience Supervisor:** Ms. Subarna Sen



Department of Computer Science Surendranath College,University of Calcutta

# PROJECT CERTIFICATE

This is to certify that the Project entitled **“Gesture Detection”** submitted on fulfillment of the requirement of 6th semester of B.Sc in Computer Science under University of Calcutta; has been carried out by Indranil Kundu(Registration No.: 115-1111-0431-19 & Roll No.: 193115-21-0044) and Argha Dutta(Registration No.:115-1114-0425-19 & Roll No.: 193115-21-0051) under the supervision of Ms. Subarna Sen, Department of Computer Science, Surendranath College, University of Calcutta.

\_ Project Supervisor

\_

External Examiners Date:

Place: Kolkata

**Abstract**

Computer is a part and parcel in our day to day life and used in various fields. The interaction of human and computer is accomplished by traditional input devices like mouse, keyboard etc. Hand gestures can be a useful medium of human-computer interaction and can make the interaction easier. Gestures vary in orientation and shape from person to person. So, non-linearity exists in this problem. Recent research has proved the supremacy of Convolutional Neural Network (CNN) for image representation and classification. Since, CNN can learn complex and non-linear relationships among images, in this paper, a static hand gesture recognition method using CNN was proposed.

**Keywords** : Machine Learning, Python, Computer Vision, OpenCV, Tensorflow, Keras.

**Table of Contents**

**Chapter Name Page No**

1. Introduction 1
   1. [State of Problem 1](#_TOC_250012)
   2. [Domain Description 1](#_TOC_250011)
   3. Motivation 2
   4. Application 3
2. Review of Related Work 4
3. Methodology 8
   1. [Problem Formulation 8](#_TOC_250010)
      1. Convolutional Layer 8
      2. Pooling Layer 8
      3. Dropout Layer 9
      4. Non-Linear Layer 9
      5. [Fully-Connected Layer 10](#_TOC_250009)
      6. Linear Bottlenecks Layer 10
   2. [Algorithm Description 11](#_TOC_250008)
   3. [Design Description 11](#_TOC_250007)
4. Implementation 13
   1. [Dataset 13](#_TOC_250006)
   2. [Pre-processing 13](#_TOC_250005)
   3. [Training 13](#_TOC_250004)
   4. [Optimizer Hyperparameter 14](#_TOC_250003)
   5. [Loss and Accuracy Function 14](#_TOC_250002)
5. Result and Discussion 15
6. Conclusion 17
   1. [Constraints 17](#_TOC_250001)
   2. [Future Scope 17](#_TOC_250000)
7. Reference 18

# Acknowledgement

We would like to express our sincere gratitude to several individuals and organizations for supporting us throughout our Graduate study. First, we wish to express our sincere gratitude to our supervisor, Ms. Subarna Sen , for her enthusiasm, patience, insightful comments, helpful information, practical advice and unceasing ideas that have helped us tremendously at all times in our research and writing of this report. Her immense knowledge has enabled us to complete this project successfully. Without her support and guidance, this project would not have been possible. We would also like to thank our faculty members for standing by ourselves and helping us out every way possible.

Thanks!

# Chapter 1 - Introduction:

Bobick and Wilson have given a definition of gesture. According to them, the motion of the body that is intended to communicate with other agents can be defined as gesture. The sender and receiver must have same sort of information for a successful gesture.

# State of problem:

In our endeavour, we have developed a Gesture Detection System to determine which

Gesture the person is showing. Gesture Detection Platform uses the concepts of Artificial Intelligence (AI) to detect a finger/palm in the image or video of a person. To detect the existence of a fingers/palm, we try to detect the shape of the hand.

The key problem is how to make a computer able to understand the hand gestures. Hand gestures vary in orientation of fingers and shape of hands. So, non-linearity is one of the characteristics of hand gestures that has to be dealt.

# Domain Description:

In this project, we have to develop a model using object detection technique. We develop the gesture detector model for detecting what signs the person is trying to make.

**Object detection:**

This is a common Computer Vision problem which deals with identifying and locating object of certain classes in an image or video. In object detection, it draws bounding boxes around detected objects, which allow us to come across where said objects are in (or how they move through) a given scene.

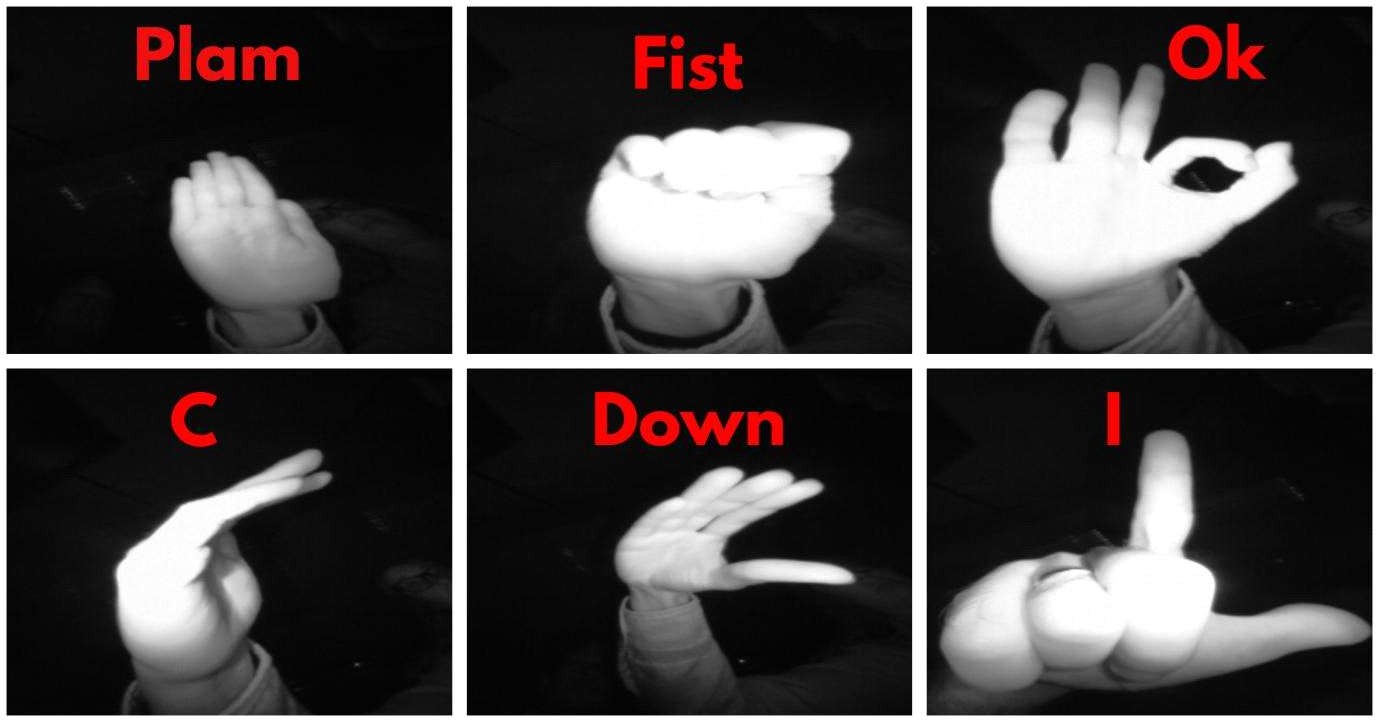
Object detection is completely connected to other similar computer vision techniques like image recognition and image segmentation, in that it helps us understand and study scenes in images or video.

we’ll examine how object detection can be used in the following areas:

* + - Video surveillance
    - Crowd counting
    - Anomaly detection (i.e. in industries like agriculture, health care)
    - Face detection
    - Self-driving cars
    - Hand detection

# Motivation:

The motivation for gesture recognition researchers is to develop a system which can detect the gestures and these gestures are widely used for conveying the information or to control the devices. Camera based solutions for gesture recognition has been widely used in numerous applications and capability to communicate through Human Computer Interaction. A new device called Leap Motion Controller is discussed which provides the complete information of hands helps to track hand movements and gestures through API (Application Programming Interface). An idea of real time hand gesture recognition process through this device is explained along with insight of existing machine learning models. Finally an attempt is made to explain the complexities with the device and the models along with its features.



# Applications:

Hand Gesture Detection system is required for ensuring that people who cant speak can communicate in public places or anywhere and do many other things such as-

### Taking to computer

Imagine a world in which a person putting together a presentation can add a quote or move an image with a flick of the wrist instead of a click of a mouse. A future in which we can easily interact in virtual reality much as we do in actual reality, using our hands for small, sophisticated movements like picking up a tool, pushing a button or squeezing a soft object in front of us.

### Medical Operation

Gestures can be used to control the distribution of resources in hospitals, interact with medical instrumentation, control visualization displays, and help handicapped users as part of their rehabilitation therapy. Some of these concepts have been exploited to improve medical procedures and systems; for example, a technology which satisfied the “come as you are” requirement, where surgeons control the motion of a laparoscope by making appropriate facial gestures without hand or foot switches or voice input

### Gesture Based Gamming Control

Computer games are a particularly technologically promising and commercially rewarding arena for innovative interfaces due to the entertaining nature of the interaction. Users are eager to try new interface paradigms since they are likely immersed in a challenging game-like environment. In a multi-touch device, control is delivered through the user’s fingertips.

### Home Appliance

Most electronic devices focus on the hand gesture recognition algorithm and the corresponding user interface. Hand Gesture Based Remote is a device to replace all other remotes used in households and perform all their functions.

### Communication

This may include the senses of sight, hearing, touch and even smell. The interaction of a user with a VR environment is limited to the use of various devices or VR head-mounted displays which often require the use of pointing devices. However, for virtual reality, commanding devices which can be manipulated unseen are much preferred for example voice commands, lip-reading, interpretation of facial expression and recognition of hand gestures.

# Chapter 2 - Review of related work:

Various works have been done on hand gesture recognition and some notable research in this topic are mentioned.

* Hand gesture recognition addresses a fault in interaction systems[1]. Controlling things by hand is more natural, easier, more flexible and cheaper, and there is no need to fix problems caused by hardware devices, since none is required. From previous sections, it was clear to need

to put much effort into developing reliable and robust algorithms with the help of using a camera sensor has a certain characteristic to encounter common issues and achieve a reliable result. Each technique mentioned above, however, has its advantages and disadvantages and may perform well in some challenges while being inferior in others. The proposed system gives recognition rate of 98.02%.

* The paper will discuss the gesture acquisition methods, [2]the feature extraction process, the classification of hand gestures, the applications that were recently proposed, the challenges that face researchers in the hand gesture recognition process, and the future of hand gesture recognition. We shall also introduce the most recent research from the year 2016 to the year 2018 in the field of hand gesture recognition for the first time.
* Gestures are a common form of human communication and important for human computer

interfaces (HCI).[3] Recent approaches to gesture recognition use deep learning methods, including multi-channel methods. We show that when spatial channels are focused on the hands, gesture recognition improves significantly, particularly when the channels are fused using a sparse network. Using this technique, we improve performance on the ChaLearn IsoGD dataset from a previous best of 67.71% to 82.07%, and on the NVIDIA dataset from 83.8% to 91.28%.

* Gesture recognition is an important issue in computer vision[4]. Recognizing gestures with videos remains a challenging task due to the barriers of gesture-irrelevant factors. In this paper,

we propose a multimodal gesture recognition method based on a ResC3D network. One key idea is to find a compact and effective representation of video sequences. Therefore, the video enhancement techniques, such as Retinex and median filter are applied to eliminate the illumination variation and noise in the input video, and a weighted frame unification strategy is utilized to sample key frames. Upon these representations, a ResC3D network, which leverages the advantages of both residual and C3D model, is developed to extract features, together with a canonical correlation analysis based fusion scheme for blending features. The performance of our method is evaluated in the Chalearn LAP isolated gesture recognition challenge. It reaches 67.71% accuracy.

* The dynamic gesture recognition based on HMM and D-S evidence theory proposed in this paper achieves better recognition effect and has important significance for human-computer interaction under the Internet of Things technology. With the continuous development of Internet of Things technology, non-contact human-computer interaction has become a hot spot. In the fields of smart cars, smart homes, robots, etc.[5] Accuracy was 91.6%.
* We propose DeepGRU, a novel end-to-end deep network model informed by recent developments in deep learning for gesture and action recognition, that is streamlined and

device-agnostic. DeepGRU, [6]which uses only raw skeleton, pose or vector data is quickly understood, implemented, and trained, and yet achieves state-of-the-art results on challenging datasets. At the heart of our method lies a set of stacked gated recurrent units (GRU), two fully-connected layers and a novel global attention model. We evaluate our method on seven publicly available datasets, containing various number of samples and spanning over a broad range of interactions (full-body, multi-actor, hand gestures, *etc*.). In all but one case we outperform the state-of-the-art pose-based methods. For instance, we achieve a recognition accuracy of 84.9% and 92.3%.

* With the rapid development of computer vision,[7] the demand for interaction between human and machine is becoming more and more extensive. Since hand gestures are able to express enriched information, the hand gesture recognition is widely used in robot control, intelligent furniture and other aspects. The paper realizes the segmentation of hand gestures by establishing the skin color model and AdaBoost classifier based on haar according to the particularity of skin color for hand gestures, as well as the denaturation of hand gestures with one frame of video being cut for analysis. In this regard, the human hand is segmentd from the complicated background, the real-time hand gesture tracking is also realized by CamShift algorithm. Then, the area of hand gestures which has been detected in real time is recognized by convolutional neural network so as to realize the recognition of 10 common digits. Experiments show 98.3% accuracy.
* Using gestures can help people with certain disabilities in communicating with other people. This paper proposes a lightweight model based on YOLO (You Only Look Once) v3 and DarkNet-53 convolutional neural networks for gesture recognition without additional preprocessing, image filtering, and enhancement of images.[8] The proposed model achieved high accuracy even in a complex environment, and it successfully detected gestures even in

low-resolution picture mode. The proposed model was evaluated on a labeled dataset of hand gestures in both Pascal VOC and YOLO format. We achieved better results by extracting features from the hand and recognized hand gestures of our proposed YOLOv3 based model with accuracy, precision, recall, and an F-1 score of 97.68, 94.88, 98.66, and 96.70%,

respectively. Further, we compared our model with Single Shot Detector (SSD) and Visual Geometry Group (VGG16), which achieved an accuracy between 82 and 85%. The trained model can be used for real-time detection, both for static hand images and dynamic gestures recorded on a video.

* We linearly sampled 16 frames from each sequence with each frame containing the entire gesture space.[9] Then, end-to-end training was conducted for the C3D architecture after replacing the last two FC layers and the classification layer. The mini-batch gradient descent with a learning rate of 10−4, a weight decay of 10−6, and a momentum of 0.9 was used to fit the entire model over 100 iterations. The batch size was 16 samples. We repeated the experiment by changing the number of trainable and frozen layers each time to find the optimal level for knowledge transfer. We started by training only the last 3DCNN layer with the FC layer and the classification layer, while the remaining layers were frozen. Then, in each repetition, we incremented the number of trainable layers by activating the next nearest layer to the previously activated ones. FIGURE 7 illustrates the results of the experiment in terms of evaluation loss and recognition accuracy. It shows that the performance of the model is improved as we increase the number of trainable layers as long as the first layer is frozen. That is, the best performance (80.94%) was achieved by fine-tuning all the layers except the first one.
* Recently, automatic hand gesture recognition has gained increasing importance for two principal reasons: the growth of the deaf and hearing-impaired population, and the development of vision-based applications and touchless control on ubiquitous devices. As hand gesture recognition is at the core of sign language analysis a robust hand gesture recognition system should consider both spatial and temporal features.[10] Unfortunately, finding discriminative spatiotemporal descriptors for a hand gesture sequence is not a trivial task. In this study, we proposed an efficient deep convolutional neural networks approach for hand gesture recognition. The proposed approach employed transfer learning to beat the scarcity of a large labeled hand gesture dataset. We evaluated it using three gesture datasets from color videos: 40, 23, and 10 classes were used from these datasets. The approach obtained recognition rates of 98.12%, 100%, and 76.67% on the three datasets, respectively for the signer-dependent mode. For the signer-independent mode, it obtained recognition rates of 84.38%, 34.9%, and 70% on

the three datasets, respectively.

**Chapter 3 - Methodology:**

In research, methodology is defined as a systematic approach to resolving a study topic by collecting data using various approaches, interpreting the data, and deriving conclusions from the data. A research technique is essentially the plan for a research or study.

* 1. **Problem Formulation:**

The step of problem formulation is to determine the user's characteristics and needs. The performance criteria for the chosen solvent will be defined in this step. For the developed solvents, performance requirements and goal attributes were established.

### Convolutional Layer:

The Convolutional Neural Network's basic building element is this layer. Convolution is a mathematical term that refers to the mathematical combining of two functions to produce a third function. It employs a sliding window technique to aid in the extraction of characteristics from an image. This assists in the creation of feature maps. The output C is obtained by convolution two functional matrices, one of which is the input image matrix A and the other is the convolutional kernel B.

C(T)=(A\*B)(x)= ∫∞−∞A(T)×B(T−x) dT

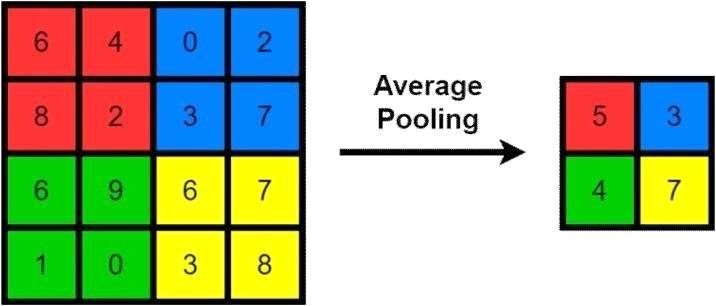
### Pooling Layer:

The use of pooling techniques can speed up calculations by reducing the size of the input matrix without sacrificing many properties. There are several types of pooling operations that can be used, some of which are described below:

* + - * **Max Pooling:** It uses the highest value in the designated region

where the kernel is now located as the value for the output of the matrix value for that cell.

* + - * **Average Pooling:** It takes the average of all the values currently in the region where the kernel is located and uses that value as the matrix value for that cell's matrix value. The average-pooling operation is depicted.



Average-Pooling Operation.

### Dropout Layer:

By removing random biased neurons from the model, this helps to reduce overfitting that can occur during training. These neurons can be found in both hidden and visible levels. The dropout ratio can be modified to alter the likelihood of a neuron being dropped.

### Non-Linear Layer:

The convolutional layers are frequently followed by these layers. Rectified

Linear Units (ReLUs), such as Leaky ReLUs, Noisy ReLUs, Exponential ReLUs, and so on, as well as sigmoid and tanh functions, are among the most often utilised non-linear functions. Non-Linear Functions of various types and their equations are shown below:

Sigmoid: σ(x)= 1/1+e−x Leaky Relu: f(x)=max(0.1x, x)

Tanh: f(x)=tanh(x)

Maxout: f(x)=max(wT1x+ b1 , wT2x+b2)

Relu: f(x)=max(0,x) ELU: f(x)= {x x≥0 α(ex−1) x<0 )

### Fully-Connected Layer:

These layers connect to the activation layers and append the model. These layers help in the multi-class or binary classification of pictures. SoftMax is an example of an activation function used in these layers, and it calculates the possibility of predicted output classes.

### Linear Bottlenecks:

Because many matrix multiplications cannot be reduced to a single numerical operation, non-linear activation functions such as ReLU6 are used in neural networks to easily remove several disparities. A multilayer neural network can be formed using this method. Because the ReLU activation function ignores values less than zero, To combat information loss, the network's dimensions expand by increasing the number of channels.

## Algorithm Description:

Using the method presented below, the suggested Gesture Detection methodology has been effectively explained. The photos were first pre-processed before being trained on the entire dataset then model that was trained in the previous step was used to accurately recognise the gesture.

#### Algorithm for pre-processing and training on dataset:

INPUT: Images. OUTPUT: Trained Model

Step 1: Load Images.

Step 2: Process the images, i.e. resizing, normalization, and conversion to a 1D array. Step 3: Load the filenames and their respective labels.

Step 4: Perform Data augmentation and then spilt data into training and testing batches. Step 5: Make model->sequential.

Step 6: Add all the necessary layers,

Train it on training batches and compile it using Adam optimizer.

Step 7: Save the model for future use.

#### Algorithm for deployment of Gesture Detector:

INPUT: Choice of deployment and files(optional) OUTPUT: Images classified into various gestures.

Step 1: Load saved classifier from disk.

Step 2: If the choice is classification on image:

load Image(s)

i. Apply gesture detection model to Detect gestures in an image ii. if gestures are detected:

Crop gestures and convert them into grayscale. Get predictions from the gesture classifier model. Show predictions and save resultant image.

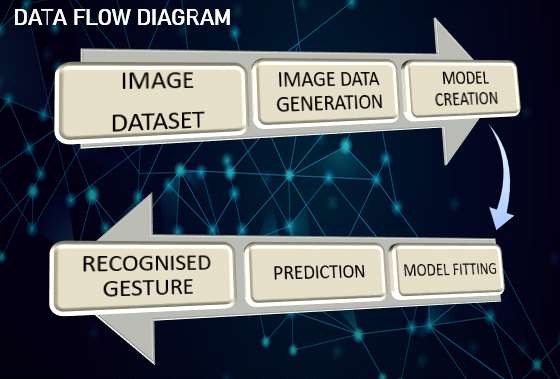
Else

Show no output

Step 3: End streamed when q is pressed.

## Design Description:

It depicts our suggested system design (image courtesy of the dataset). First the dataset is downloaded. Images are generated using image data generator. Model is created using various layers. Then the model that was created is fitted. Then image is taken from the testing imaes to predict the output. After that images from outside is added to predict the output.



# Chapter 4 - Implementation:

## Dataset:

Here we get dataset ‘Hand Gesture Recognisation Dataset’.

In our technique, we used 80% of the dataset for training and 20% for testing, resulting in a split ratio of 0.8:0.2 for train to test data. We used 20% of the training data to create a validation data set. The dataset is used for training 64 percent of the time, validation 16 percent of the time, and testing 20 percent of the time.

## Pre-processing:

The pre-processing phase occurs before the data is trained and tested. The pre- processing consists of four steps: scaling the image size, turning the image to an array, pre-processing input , and performing hot encoding on labels.

Due to the efficacy of training models, image scaling is a significant pre-processing step in computer vision. The model will perform better if the image is smaller. In this case, scaling an image means converting it to 224 224 pixels.

The next step is to create an array from all of the photographs in the dataset. The image is transformed into an array so that the loop function can call it.

## Training:

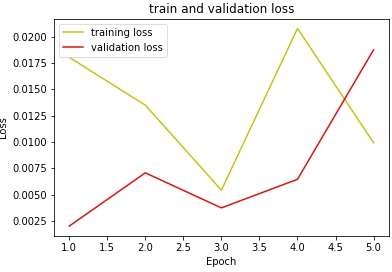
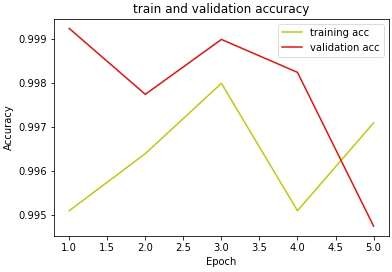
At training time, we compare ground truth boxes with default bounding boxes of various sizes and aspect ratios, and then use the Intersection over Union (IoU) approach to select the best matching box for each pixel. IoU determines how much of our anticipated box corresponds to reality. The values range from 0 to 1, and increasing IoU values influence forecast accuracies; the best value is the highest IoU value. IoU's equation and pictorial representation are as follows:

IoU(B1, B2) = B1 ∩ B2/B1 𝖴 B2

## Optimizer Hyperparameter:

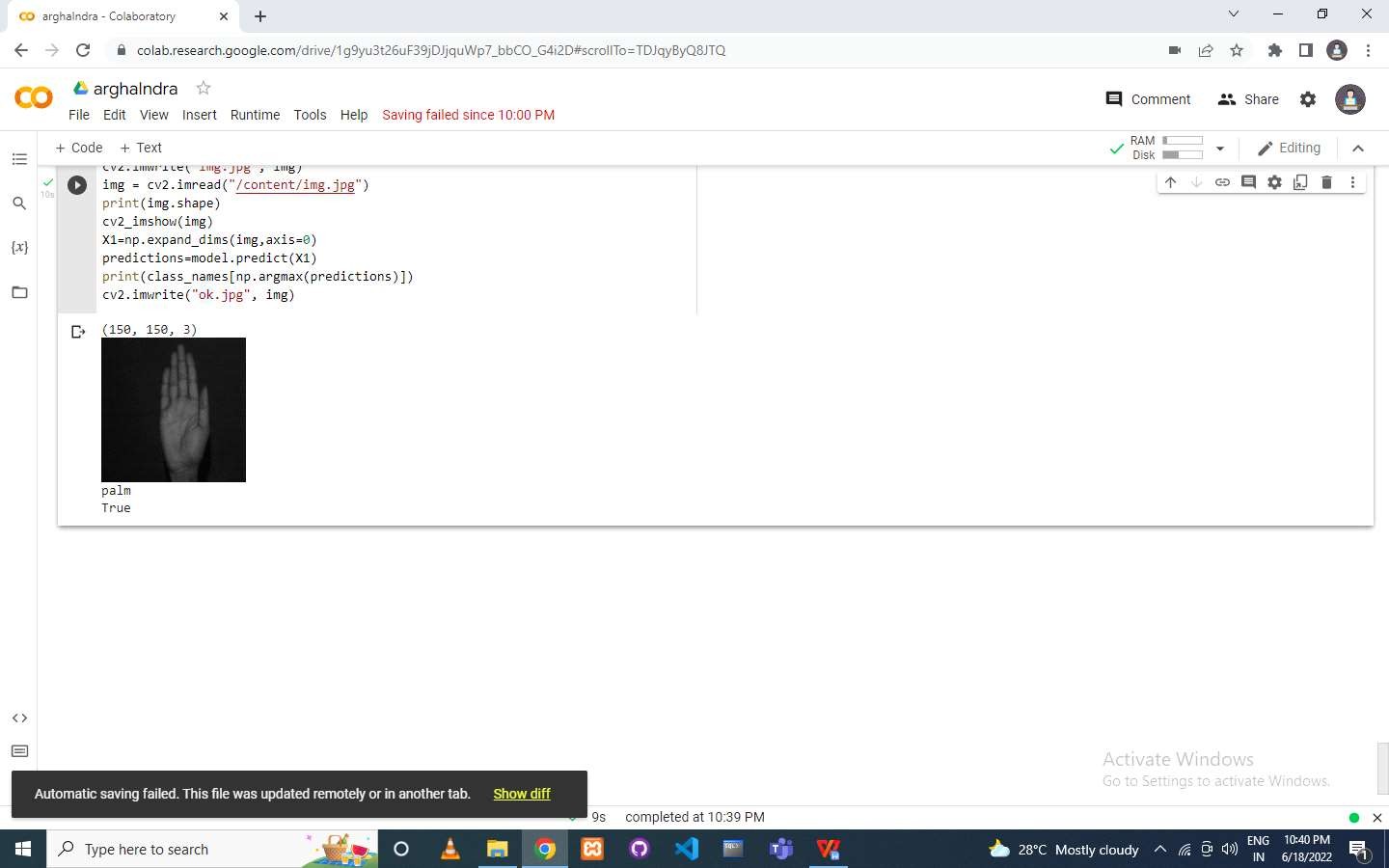
We start the actual training process, we can adjust or optimise our model using optimizer parameters. The learning rate is a hyperparameter that governs how much our neural network's weights are adjusted in relation to the gradient. Mini-batch size is a hyperparameter that affects the training's resource requirements as well as the speed and number of repeats. Epochs are the hyperparameters that govern how often the model is performed.

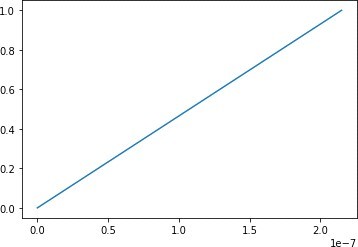
## Loss and Accuracy Function:

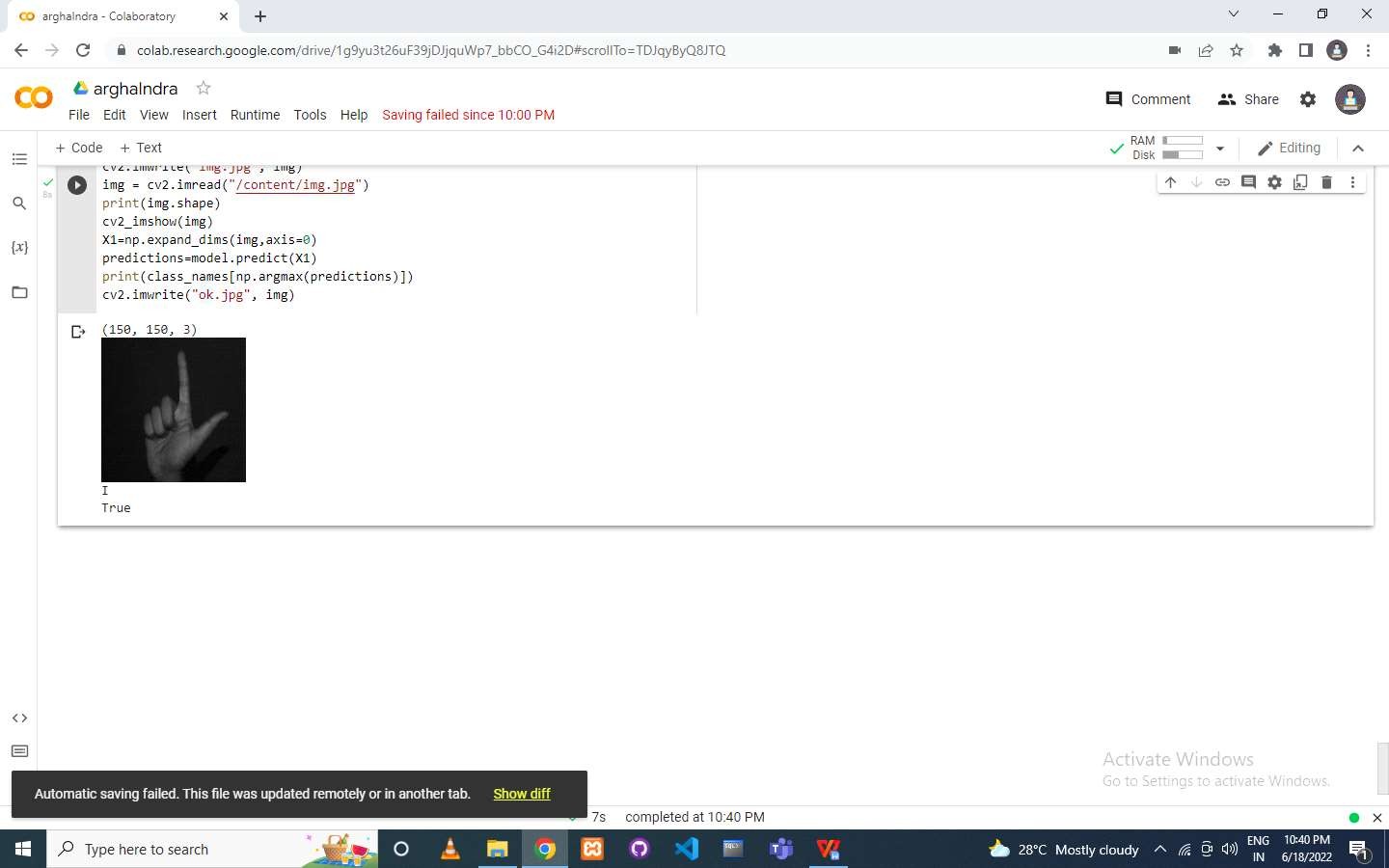
There are two types of loss in the overall detection problem: localization loss and confidence loss. The difference between the default anticipated bounding box and the ground truth bounding box is the localization loss (g). We try to change the width and height of the box for a particular centre in order to reduce the loss.

# Chapter 5 - Result and Discussion:

The outcomes of our experimental examination are presented in this section. The purpose of

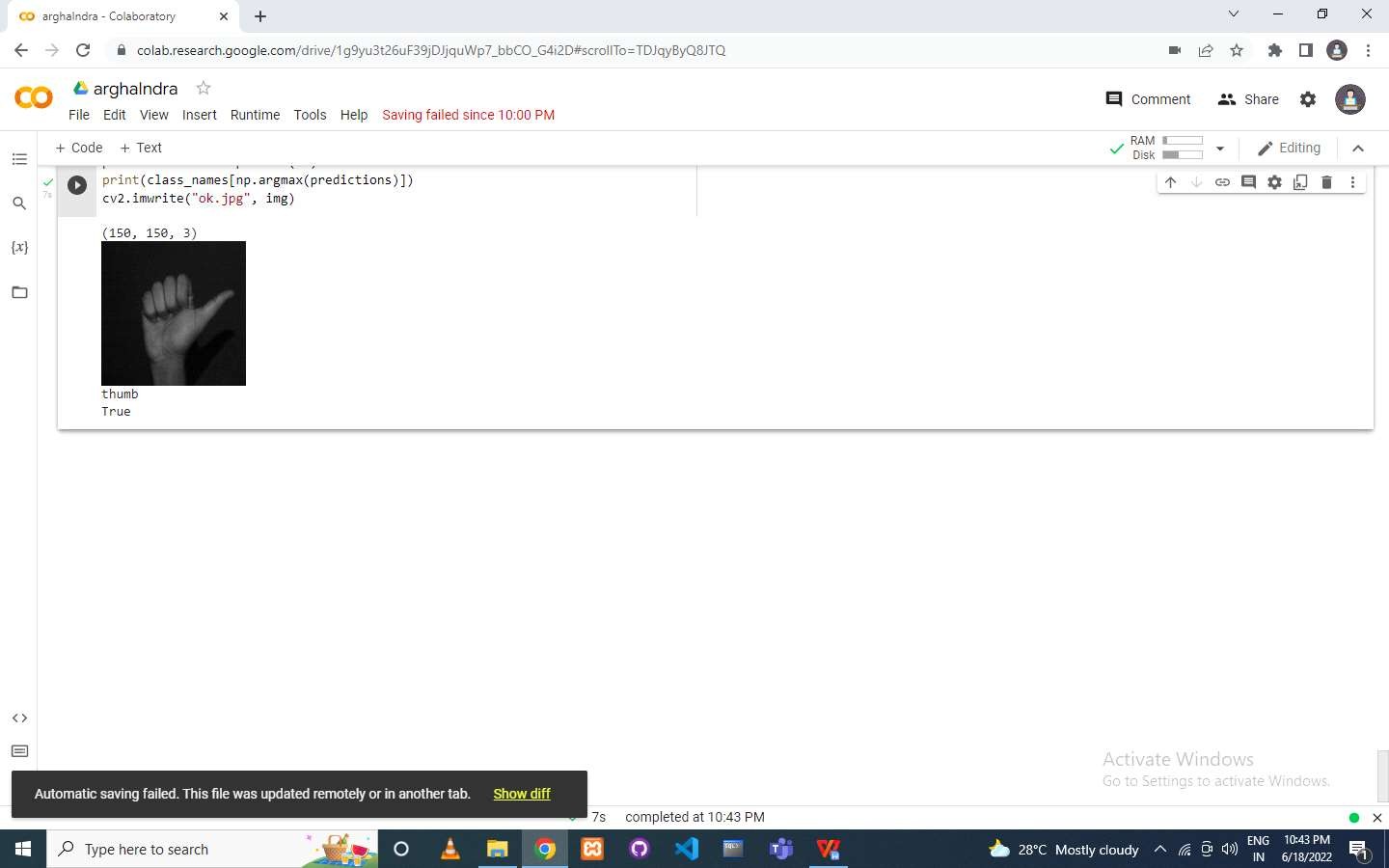
these tests is to gain insight into the performance of detection approaches with masked-face photos, as well as to assess the performance of identification models that, in addition to detecting the existence of face masks, also identify whether the masks are worn appropriately. As a result of our project, the accuracy level is 99.32%.

In the fields of image analysis, computer vision, authentication, and verification, face mask recognition is a critical strategy. The proposed system's training set includes both masked and unmasked images, and the system will then be processed via the segmentation, feature extraction, and classification stages. Three different classifiers, namely Support vector machine neural network Decision tree and Suggested Method CNN, are used to validate the accuracy of mask detections in this proposed system. By projecting on a higher dimensional space, the CNN classifier predicts an ideal hyperplane that linearly separates all the characteristics vectors.



#### Output 2 & it’s corresponding graph

This system will be processed on video as well as. And here is some screenshots of the output video in the below:



**Output of the video & it’s corresponding graph**

# Chapter 6 - Conclusion:

Gesture recognition technology is the turning point in the world of VR/AR development. It can allow seamless non-touchable control of computerized devices to create a highly interactive, yet fully immersive and flexible hybrid reality.

The inclusion of this technology in multiple applications across various sectors is further revolutionizing human-computer communication. That said, gesture recognition is no novice’s game.

It’s a fully integrated, highly advanced technology that requires specialized skills of individuals with relevant experience that can guarantee favorable results. AppReal is a development company with the resources, talents, and expertise of over 200 extremely proficient and dedicated professionals who can recognize and understand your requirements and successfully deliver to your expectations. AppReal can help you realize your VR goals and make them a reality.

## Constraints:

* + Required high resolution camera to detect the hand’s object properly.
  + In the dark place this application does not work.
  + Required high processing computer machine to execute the application.
  + Required many module and libraries to implement the application.

## Future Scope:

We will try to add some features in our system. By tracking hands we can increase/decrease the volume of our computer system. Basically we can control many

things in our PC using this. The idea is to recognise it and apply it as input to the computer. Another improvement would be to enhance the recognition capacity for various lightning conditions, which is encountered as a challenge in this project can be worked upon in future.

# Chapter 7 - References:

* Oudah, Munir, Ali Al-Naji, and Javaan Chahl. "Hand gesture recognition based on computer vision: a review of techniques." journal of Imaging 6.8 (2020): 73.
* Yasen, Mais, and Shaidah Jusoh. "A systematic review on hand gesture recognition techniques, challenges and applications." PeerJ Computer Science 5 (2019): e218.
* Narayana, Pradyumna, Ross Beveridge, and Bruce A. Draper. "Gesture recognition: Focus on the hands." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
* Miao, Qiguang, et al. "Multimodal gesture recognition based on the resc3d network." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2017.
* Li, Gongfa, et al. "Dynamic gesture recognition in the internet of things." IEEE Access 7 (2018): 23713-23724.
* Maghoumi, Mehran, and Joseph J. LaViola. "DeepGRU: Deep gesture recognition utility." International symposium on visual computing. Springer, Cham, 2019.
* Sun, Jing-Hao, et al. "Research on the hand gesture recognition based on deep learning." 2018 12th International symposium on antennas, propagation and EM theory (ISAPE). IEEE, 2018.
* Konar, Amit, and Sriparna Saha. "Gesture recognition." Principles, Techniques and Applications., Cham: Springer International Publishing (2018).
* Al-Hammadi, Muneer, et al. "Deep learning-based approach for sign language gesture recognition with efficient hand gesture representation." IEEE Access 8 (2020): 192527-192542.
* Al-Hammadi, Muneer, et al. "Hand gesture recognition for sign language using 3DCNN." IEEE Access 8 (2020): 79491-79509.
* Li, Gongfa, et al. "Hand gesture recognition based on convolution neural network." Cluster Computing 22.2 (2019): 2719-2729.

12. Mujahid, Abdullah, et al. "Real-time hand gesture recognition based on deep learning YOLOv3 model." Applied Sciences 11.9 (2021): 4164.