###### Project Report On

**Analyzing the presence of violence and a particular event of violence by using Deep learning approach.**

*by*

###### Neeraj Patil (B150214284)

###### Amruta Kumbhar (B150214268)

###### Deepak Pitambare (B150214294)

###### Prachi Sabale (B150214307)

*Under the guidance of*

###### Prof. N. S. Devekar



Department of Computer Engineering AISSMS College of Engineering, Pune

**SAVITRIBAI PHULE PUNE UNIVERSITY 2020-2021**

**Project Approval sheet**

PROJECT APPROVAL SHEET

A

**Project**

on

**Analyzing the presence of violence and a particular event of violence by using Deep learning approach.**

Is successfully completed by

Neeraj Patil (B150214284)

###### Amruta Kumbhar (B150214268)

###### Deepak Pitambare (B150214294)

###### Prachi Sabale (B150214307)

at

**Department of Computer Engineering AISSMS College of Engineering, Pune Savitribai Phule Pune University**

**2020-2021**

**Report Documentation**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Report Documentation** | | | | | | | | | |
| Report Code: CS-BE-Project 2020-2021 | | | | | | | **Report Number:** | | |
| Report Title: **Analyzing the presence of violence and a particular event of violence by using Deep learning approach.** | | | | | | | | | |
| **Address (Details):**  **AISSMS College of Engineering, Pune Pin – 411001, M.S. INDIA.** | | | | | | | | | |
| **Author 1** | |  | **Author 2** |  | **Author 3** | |  | **Author 4** |  |
| **Neeraj Patil** | | | **Amruta Kumbhar** | | **Deepak Pitambare** | | | **Prachi Sabale** |  |
| **Address:** Shivsunder Residency Bhelkenagar , Kothrud , Pune - 38 | | | **Address:** sr no 683,sukhwani baug soc, A wing, fl22, adinathnagar, bhosari,pune-411039 | | **Address:** 1618 B Ward Jasud Galli Mangalwar Peth Kolhapur- 416012 | | | **Address:** Plot.no.13,Mauli Nagar,Pipeline Road,Savedi,Ahmednagar-414003 | |
| **Email:** patilnv20@gmail.com | | | **Email:** amrutak7200@gmail.com | | **Email:**  pitambaredeepak@gmail.com | | | **Email:** [patilshivani420@gmail.co](mailto:patilshivani420@gmail.co) m | |
| **Roll no:** 17CS036 | | | **Roll no:** 17CS024 | | **Roll no:** 17CS040 | | | **Roll no:** 17CS046 | |
| **Cell no:** | 9527379391 | | **Cell no:** 8208059687 | | **Cell no:** 8796878232 | | | **Cell no:** 7774851498 | |
| **Year:** 2020-2021  **Branch:** Computer Science | | | | | | | | | |
| ***Key Words*:** Convolutional Neural Network(CNN), Dataset ,keras , Mobilenet , Deep Learning, TensorFlow,etc | | | | | | | | | |
| Type of |  | Report Checked By: | | Report Checked | | **Guides Complete Name:** | | | Total |
| Report:  Final | | Date: |  |  | Prof. N.S. Devekar | | Copies |
|  | |  |  | **N+2** |
|  | | | | | | | | | |

**Abstract**

Nowadays human security against violence is one of the major concern. Violence detection techniques analyze the surveillance camera videos. Over the last few years, these cameras and other surveillance equipment are installed at sensitive areas like ATMs, Government offices, Schools, Hospitals, etc.For the security against violence, there is a need of generating timely and automated alerts to concern officials to take further action.

The aim is to develop a system for monitoring and analyzing the video streams from surveillance cameras and making the decision about violence.

Frame level features in a video are extracted by employing a CNN after which they are compounded by utilizing a variant which in turn makes use of convolutional gates CNN used for the analysis of local motion in a video.

The Deep learning technique for violence detection is used to classify the violent recognition on the base of data set and extracted features using more convolutional layers.

After assessing contemporary circumstances in the world, it is imperative that there is existential and paramount need of exploiting automated visual surveillance for detecting weapons , which will enhance effectiveness of security operations. In future we can extend this system by adding weapon detection system using various algorithms like SSD , FasterCNN etc for weapon identifcation.

Therefore with the help of video surveillance we can analyze the presence of violence and a particular event of violence by weapon detection using Deep learning techniques.

### Index

|  |  |  |
| --- | --- | --- |
| Sr. No. | Contents | Page NO. |
|  | Introduction   * 1. Problem Statement   2. Motivation | 1 |
|  | Goals And Objectives | 3 |
|  | Literature Survey | 4 |
|  | Software And Hardware Requirements   * 1. Specifications Required   4.2 Technology Required | 7 |
|  | System Architetcture | 9 |
|  | Proposed System | 11 |
|  | Modelling | 13 |
|  | Coding   * 1. Dataset Summary   2. Coding Snippets   3. Testing Snippets   4. Final Result | 16 |
|  | Result Analysis | 21 |
|  | Conclusion | 22 |
|  | Future Scope | 23 |
|  | References | 24 |

1. **Introduction**

Security is a major in every domain , due to rise in violent activities and weapons . Earlier surveillance systems were more dependent on human operator. Now, because of surveillance cameras installed at various public places like offices, hospitals, schools ,highways, etc. it can be helpful for capturing useful actions and movements for event prediction and online monitoring.Having an automated system with the ability to recognize the occurrence of violence in videos with real time response will enable authority holders to increase safety and take appropriate decisions. Violence is an abnormal behavior and those actions can be identified through smart surveillance system using which we can prevent further fatal accidents

The fundamental goal is to collect categories and recognize the most prominent and effective methods or techniques that are used in violence and anomalous activity detection using deep learning approach.

The aim is to develop an intelligent surveillance system which detects violence or weapons in given video frame using a deep supervised learning approach.

The model incorporates a pre-trained convolution Neural Network (CNN) connected to Convolutional layer.

The model takes the raw video as an input which is subsequently converted into frames and output is a binary classification of violence or non-violence label.

Therefore with the help of video surveillance we can analyze the presence of violence and a particular event of violence.

##### Problem Statement

Analyzing the presence of violence and a particular event of violence by using Deep learning approach.

.

##### Motivation

Today, the amount of public violence has increased dramatically. Security is a major in every domain , due to rise in violent activities and weapons. As much in high schools as in the street. This has resulted in the ubiquitous use of surveillance cameras. Developing a technique for the automatic analysis of surveillance videos in order to identify the presence of violence and weapon identification. Research on violence detection in videos is scarce, compared with other video analysis problems. Existing systems face the problems of low accuracy, high false alerts and high computational cost in monitoring and analyzing the video streams from surveillance cameras and making the decision about violence in real-time.

1. **Goals And Objectives**

In recent years, violence has become a major issue across the globe. So, there is a need to detect violence automatically without human intervention and generate alerts timely to the control crew. Developing a technique for the automatic analysis of surveillance videos in order to identify the presence of violence and weapon identification. To develop a system having high accuracy, less false alerts and low computational cost in monitoring and analyzing the video streams from surveillance cameras and making the decision about violence in real-time

1. **Literature Survey**

Various strategies have been proposed by researchers dealing with the problem of detection of violence from video surveillance.All the existing techniques can be divided into classes depending on the basic idea –

1. Inter-frame changes: Frames containing violence undergo massive variations because of fast motion due to ﬁghts.
2. Local motion in videos: The motion change patterns tak-ing place in the video is analyzed.
3. Several other methods follow the techniques used in action recognition, i.e. to identify spatio-temporal interest points and extract features from these points.

**3.1 Violence Detection Using Spatiotemporal features:-**

Fath U Min Ullah, Amin Ullah, Khan Muhammad, Ijaz Ul Haq and Sung Wook Baik.The three-tiered structure is forced into this program. Detection of a person using CNN performed in the first phase, in the second phase,Frame sequence provided by 3D CNN training model again in the third phase transferred to SoftMax separator.With comparative analysis again final prediction made, slide window works better as compared to SVM.The OPENVINO toolkit was used and model modeling and growth system performance.

**3.2 Improved Anomaly Detection in Surveillance Videos Based on A Deep Learning Method**

In this paper a program is developed that finds normal and unusual video. The first data preparation step, input video separated by frame and the next pre-processing step removes the background. The removal process is done manually or automation that creates a behavioral structure of the data that modeling and feature detection is available. Items are later obtained using CNN and the final decision was made in two based classifier.

The hand-crafted feature based techniques used methods like bag of words, histogram, improved Fisher encoding, etc. for aggregating the features across the frames.Recently various models using long short term memory(LSTM) RNNs are developed for addressing problems involving sequences like MT, speech recognition, caption generation and video action recognition . The LSTM was introduced in 1997 to combat the effect of vanishing gradient problem which was plaguing the deep learning community.The LSTM incorporates a memory unit which contains in-formation about the inputs the LSTM unit has seen and is regulated employing a number of fully-connected gates.

**3.3** **An In-Depth Learning Approach Based on Unwanted Visibility in Watching Videos.**

Prof. Prakhar Singh and Prof. Vinod Pankajakshan introduced a method using standard features removed from the inclusion video in their paper The Convolutional Neural Network (CNN) stack is used to extract a feature from the video input sequence frames. The Convolutional Long Short-Term Memory (convLSTM) stack is then used to predict future sequences and later the CNN transmission stack is used to predict future video sequences. Combined error is compared to the limit and the category is determined.

**3.4 Using Convolutional Neural Network with long short term memory**

Recently, Xingjian et al. replaced the fully-connected gate layers of the LSTM with convolutional layers and used this improved model for predicting precipitation now casting from radar images with improved performance. This newer model of the LSTM is called as convolutional LSTM (convLSTM). Later, it’s been used for predicting optical ﬂow images from videos and for anomaly detection in videos.By replacing the fully-connected layers within theLSTM with convolutional layers, the convLSTM model is capable of encoding spatio-temporal information in its memory cell.

# **3.5** **Robust Sports Image Classification Using InceptionV3 and Neural Networks**

This paper by Ketan Joshi, Vikas Tripathi, Chitransh Bose, Chaitanya Bhardwaj presents a robust framework for classifying the sport images based on the environment and related surroundings. In this paper, our approach is based on the use of the Inception V3 for the extraction of features and Neural Networks for the classification of various sport categories. Six categories rugby, tennis, cricket, basketball, volleyball, and badminton have been used for analysis and classification. To validate the effectiveness of the framework and Neural Networks, comparisons have been done with other classifiers like Random Forest, K-Nearest Neighbors (KNN) and Support Vector Machine (SVM).

# **3.6** **Violence Detection in Video by Using 3D Convolutional Neural Networks by Chunhui Ding, Shouke Fan, Ming Zhu, Weiguo Feng, and Baozhi Jia**

Deep models can act directly on the raw inputs and automatically extracts features. So we developed in this paper a novel 3D ConvNets model for violence detection in video without using any prior knowledge. A typical ConvNet is usually composed of three types of layers: a filter bank layer, a non-linearity layer, and a feature pooling layer. 2D convolution is performed at the convolutional layers to extract features from local neighbourhood on feature maps in the previous layer.To evaluate our method, experimental validation conducted in the context of the Hockey dataset. The results show that the method achieves superior performance without relying on handcrafted features.

**4 Software Requirements**

**4.1 Specifications Requirements**

The requirement for a system are the description of what the system should do, the service or services that it provides and the constraints on its operation. The software requirements are description of features and functionalities of the target system.

###### Hardware Interface

* + - * Processor : Intel core i5
      * RAM : 2GB or above

###### Software Interface

* Operating System : Windows 10

**4.2 Technology Requirements**

**Python :**

Python is an interpreted , high level and  general purpose  programming language. Supports multiple paradigms including structured , Object Oriented Programming.

Python offers concise and readable code. While complex algorithms and versatile workflows stand behind machine learning and AI, Python’s simplicity allows developers to write reliable systems. Developers get to put all their effort into solving an ML problem instead of focusing on the technical nuances of the language.

**TensorFlow :**

TensorFlow is a framework created by Google for creating Deep Learning models. Deep Learning is a category of machine learning models that use multi-layer neural networks. The reason for its popularity is the ease with which developers can build and deploy applications.

TensorFlow is an open source library for numerical computation and large-scale machine learning. TensorFlow bundles together a slew of machine learning and deep learning (aka neural networking) models and algorithms and makes them useful.

**Keras :**

Keras is a minimalist Python library for deep learning that can run on top of Theano or TensorFlow.It wraps the efficient numerical computation libraries Theano and TensorFlow and allows you to define and train neural network models in just a few lines of code.

It was developed to make implementing deep learning models as fast and easy as possible for research and development.

**Kaggle:**

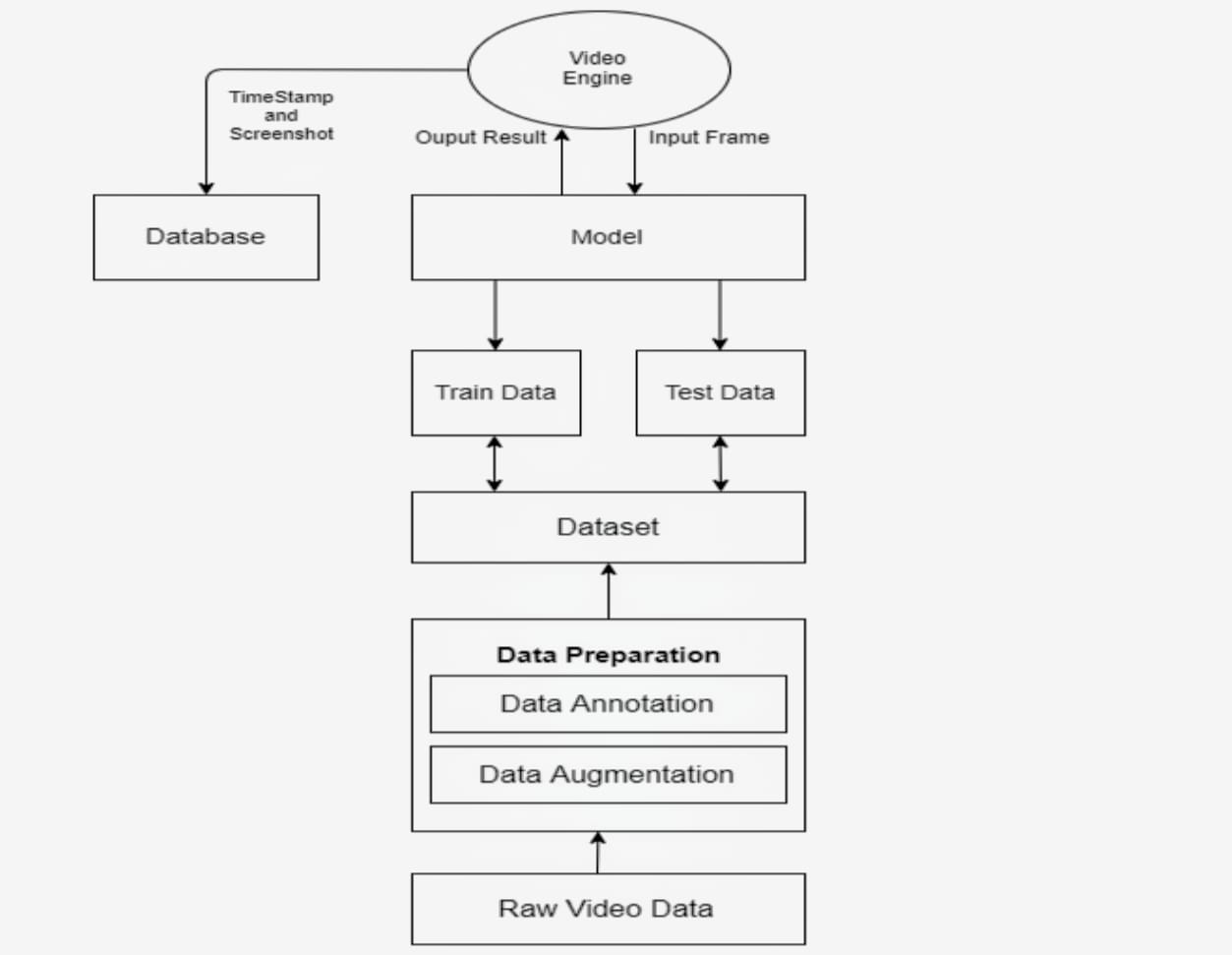
Kaggle is a source where we can get the required dataset.Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges.

**Google Collaboratory:**

It is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, deep learning ,data analysis and education.

To avoid the limitations that the hardware puts us in, we use Google Collaboratory for such advanced execution which uses Google's virtual machines to execute the code over the cloud.

**5. System Architecture**



System architecture consists of 5 modules namely Data preparation module which includes a data annotation and data augmentation , dataset, Deep Leaning model, video engine and database. This system is implemented in python and TensorFlow as a backend. User gives video file as an input and system gives output as video classification as violent or nonviolent. System supports .mp4 and video formats.

**Modules:** This system is divided into five parts according to functions performed by individuals.

**Data Preparation:** Data preparation is the act of manipulating raw data into a form that can readily and accurately be analysed, This module will be dealing with raw video data. It consists of two submodules Data Augmentation and Data Annotation.

**Data Augmentation:** It is a method of augmenting the available data.Data augmentation is a strategy that enables practitioners to significantly increase the diversity of data available for training models, without collecting new data. Data augmentation techniques such as cropping, padding, and horizontal flipping are commonly used to train large neural networks Main purpose of augmentation is to increase the size of available dataset.

**Data Annotation:** Data annotation basically is the process of adding metadata to a dataset .Since the system is based on supervised learning , annotation is an important module which labels the data.Data annotators helps us to categorize things. They can work with things like videos, advertisements, photographs and other types of material. They assess the content and then attach tags to the content.

**Dataset:** Dataset consist of data prepared by data preparation module. Dataset is further split into training and testing.

**Deep Learning Model**: Deep learning models are built using neural networks. A neural network takes in inputs, which are then processed in hidden layers using weights that are adjusted during training. Then the model spits out a prediction. The weights are adjusted to find patterns in order to make better predictions. This is the deep learning model trained using input dataset. This model will be invoked by video engine and model will classify input as violent or non-violent.

**Video Engine:** This video engine is an interface between user and deep learning model. The video engine will take the input from user and will pass it through the DL model.

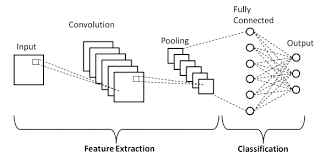
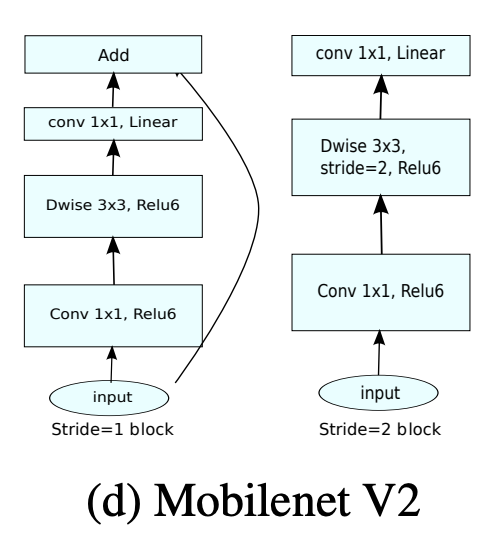
**Database:** This database contains timestamp and screenshot of already trained activities identified by system.

**6. Proposed System**

The network consists of a series of convolutional layers followed by max pooling operations for extracting discriminant features and convolutional long short memory (convLSTM) for encoding the frame level changes, that characterizes violent scenes existing in the video.

The CNN architecture which we have used is MobilenetV2.

MobileNet is a class of CNN that was open-sourced by Google, and therefore, this gives us an excellent starting point for training our classifiers that are insanely small and insanely fast.

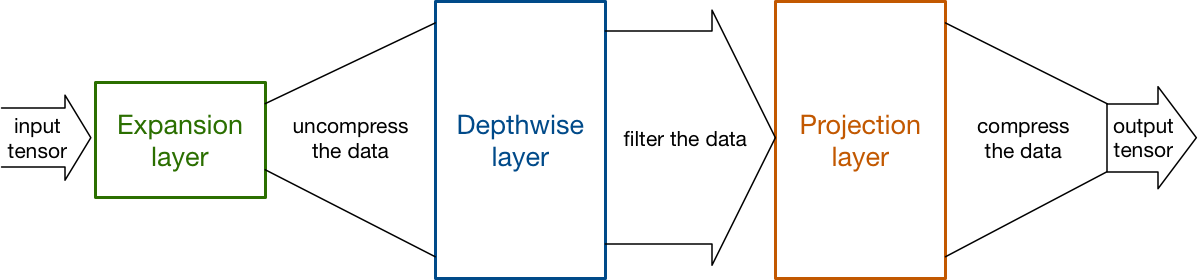
 

* **MOBILENETS :**

There are various neural network architecture like VGG16,VGG19 , Resnet50 , InceptionV3 but we have used MobilenetV2 due to its advantages over other architectures.

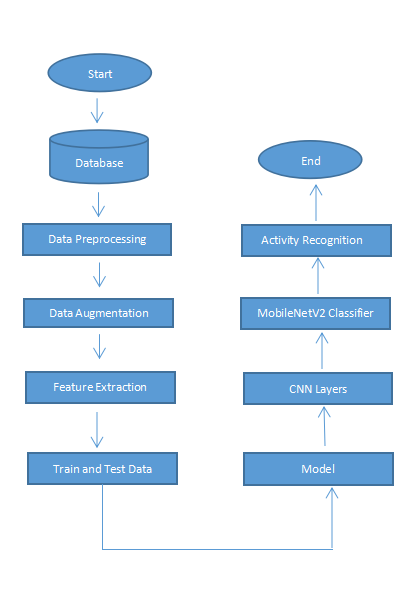
MobileNets are a family of neural network architectures released by Google to be used on machines with limited computing power like mobile devices and provides real-time classification.They strive to provide state of the art accuracy, while requiring as little memory and computing power as possible. This makes them a very fast family of networks to use for image processing.

MobileNets achieve this performance by reducing dramatically the number of learnable parameters, which also makes them faster and easier to train compared to more traditional networks.It is based on an inverted residual structure where the residual connections are between the bottleneck layers. The intermediate expansion layer uses lightweight depthwise convolutions to filter features as a source of non-linearity. As a whole, the architecture of MobileNetV2 contains the initial fully convolution layer with 32 filters, followed by residual bottleneck layers.

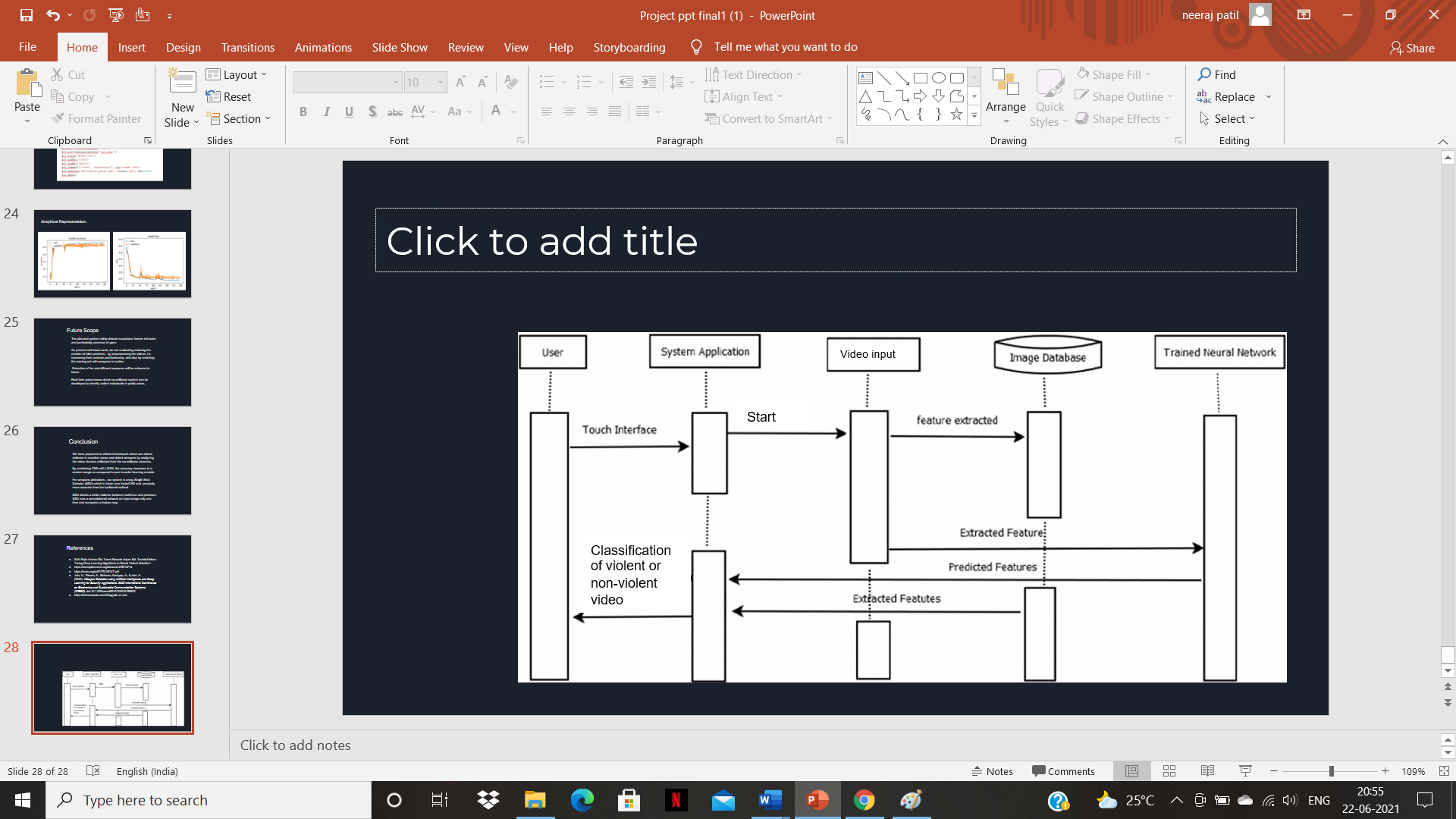


**7. Modelling**

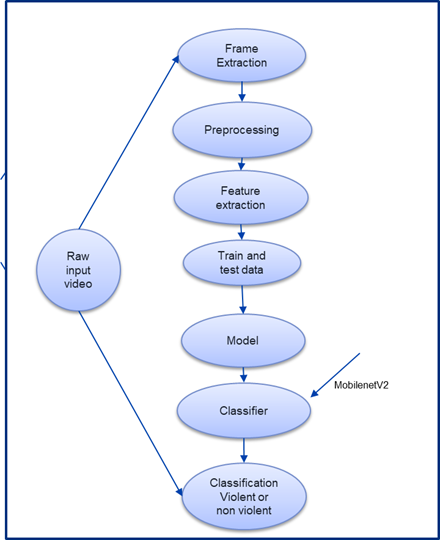
**7.1 Flow Diagram**



**7.2 Sequence Diagram**



**7.2 UML Diagram**

****

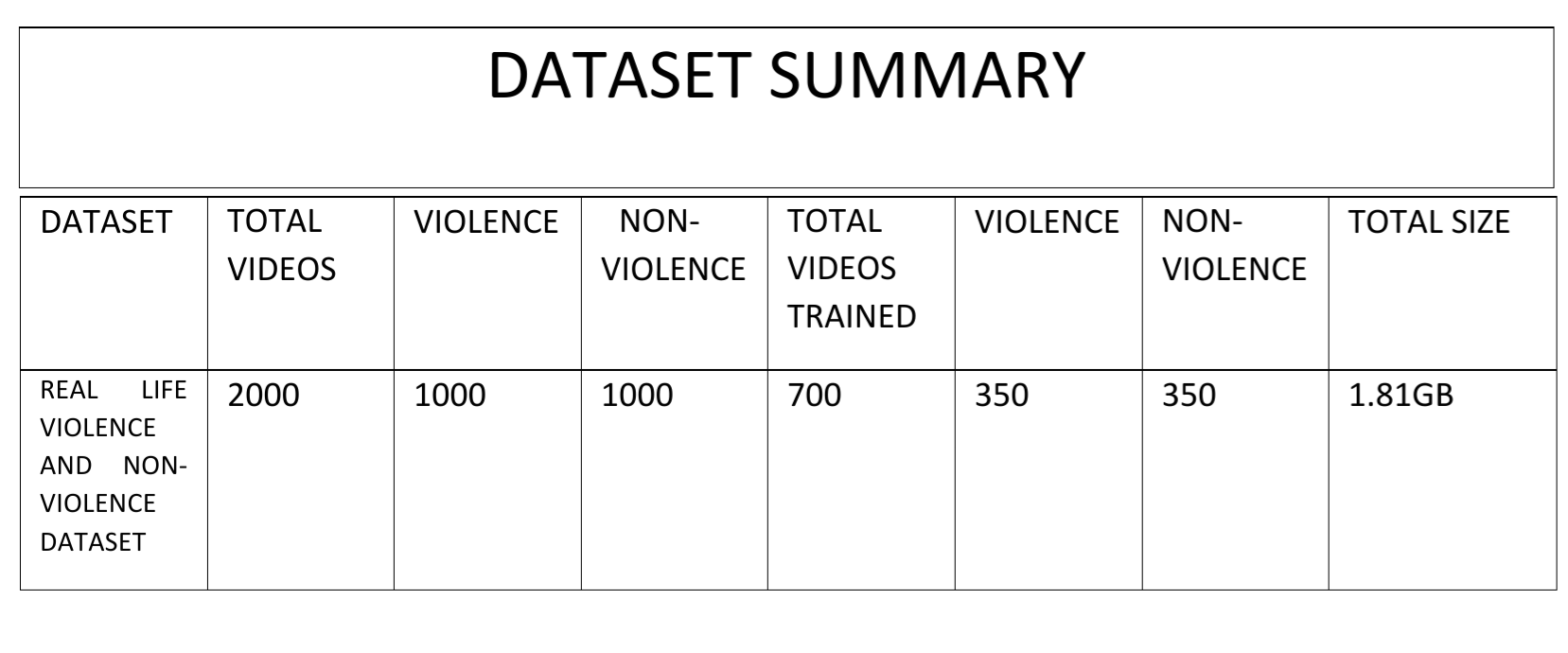
**8. Coding**

**8.1 Dataset Summary**

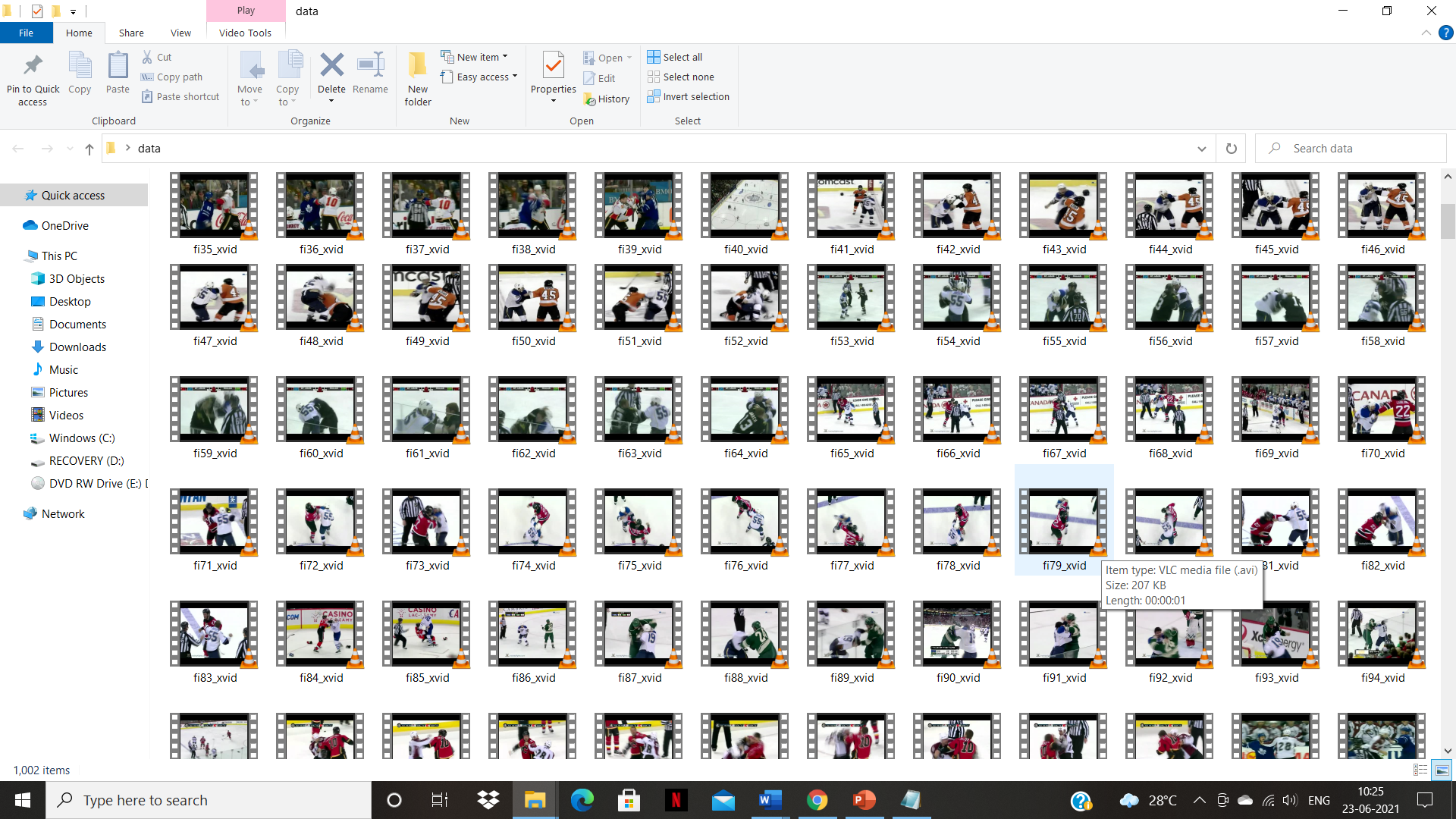
The Dataset is composed of 1000 Violence and 1000 non-violence videos collected from youtube videos, violence videos in our dataset contain many real street fights situations in several environments and conditions.

Also non-violence videos from our dataset are collected from many different human actions like sports, eating, walking.This dataset is readily available on Kaggle.

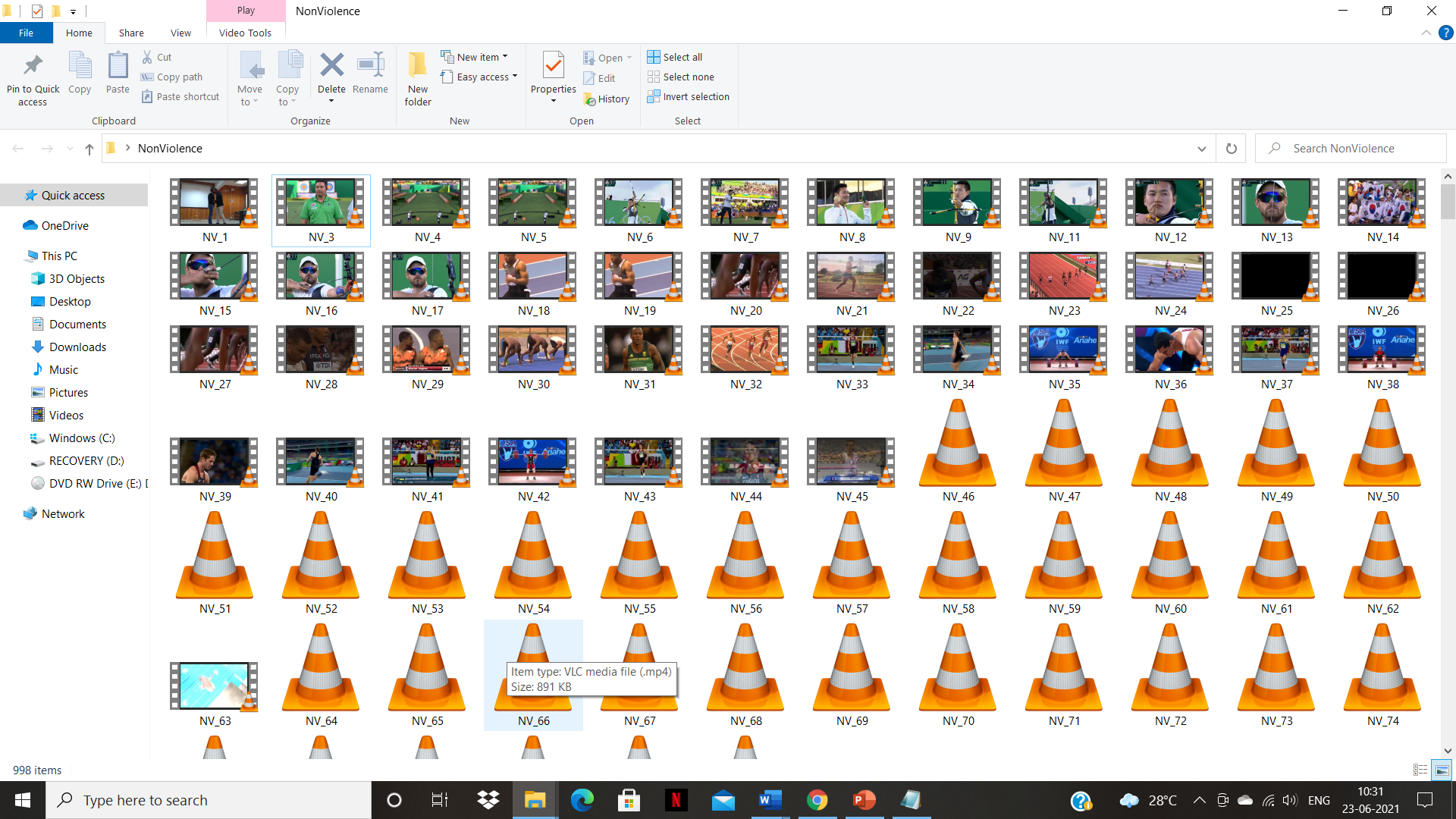
This dataset was uploaded in the drive and then was mounted in google colab.



**Dataset sample : Violence dataset**

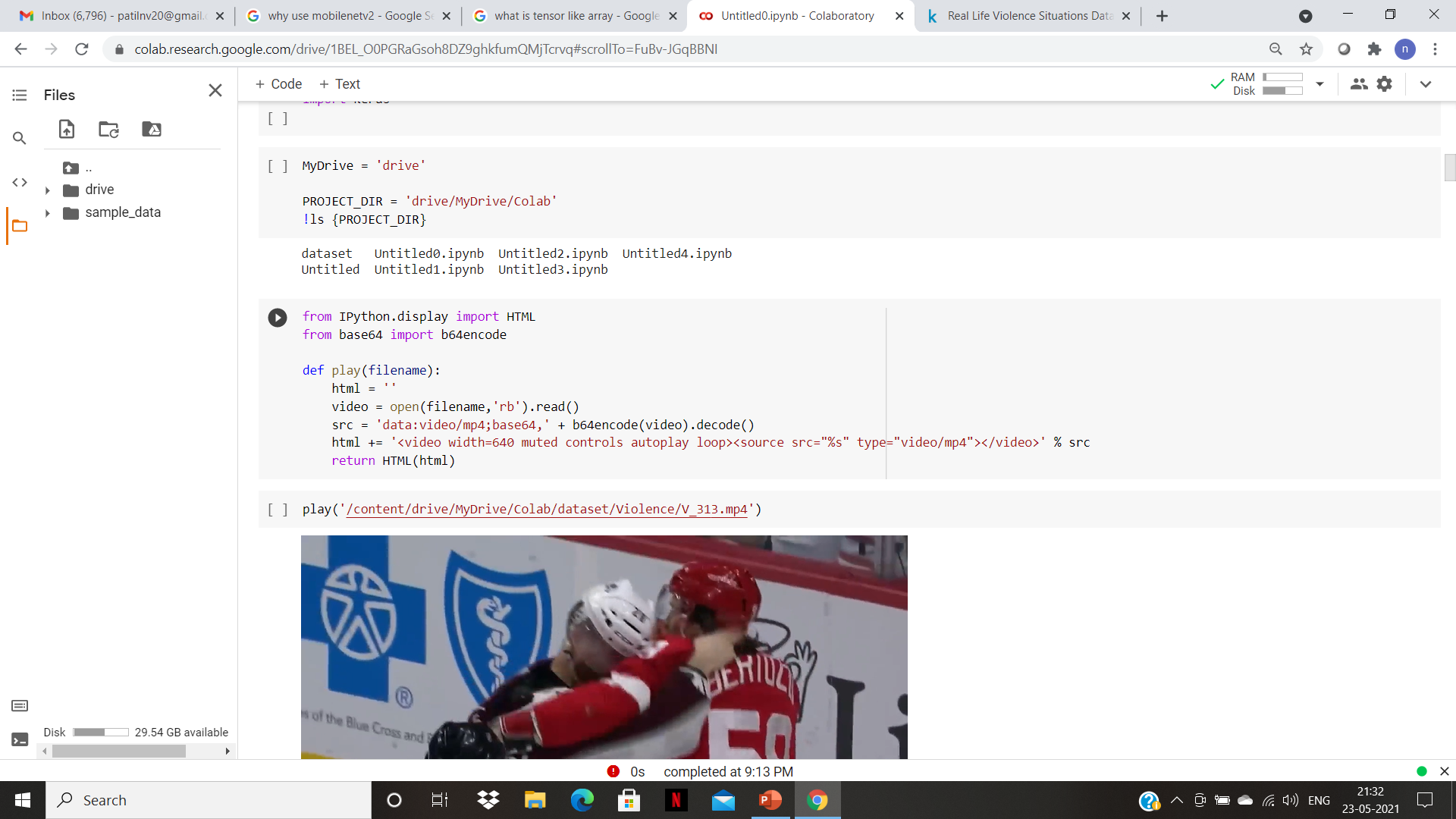


**Dataset sample : Non-Violence**

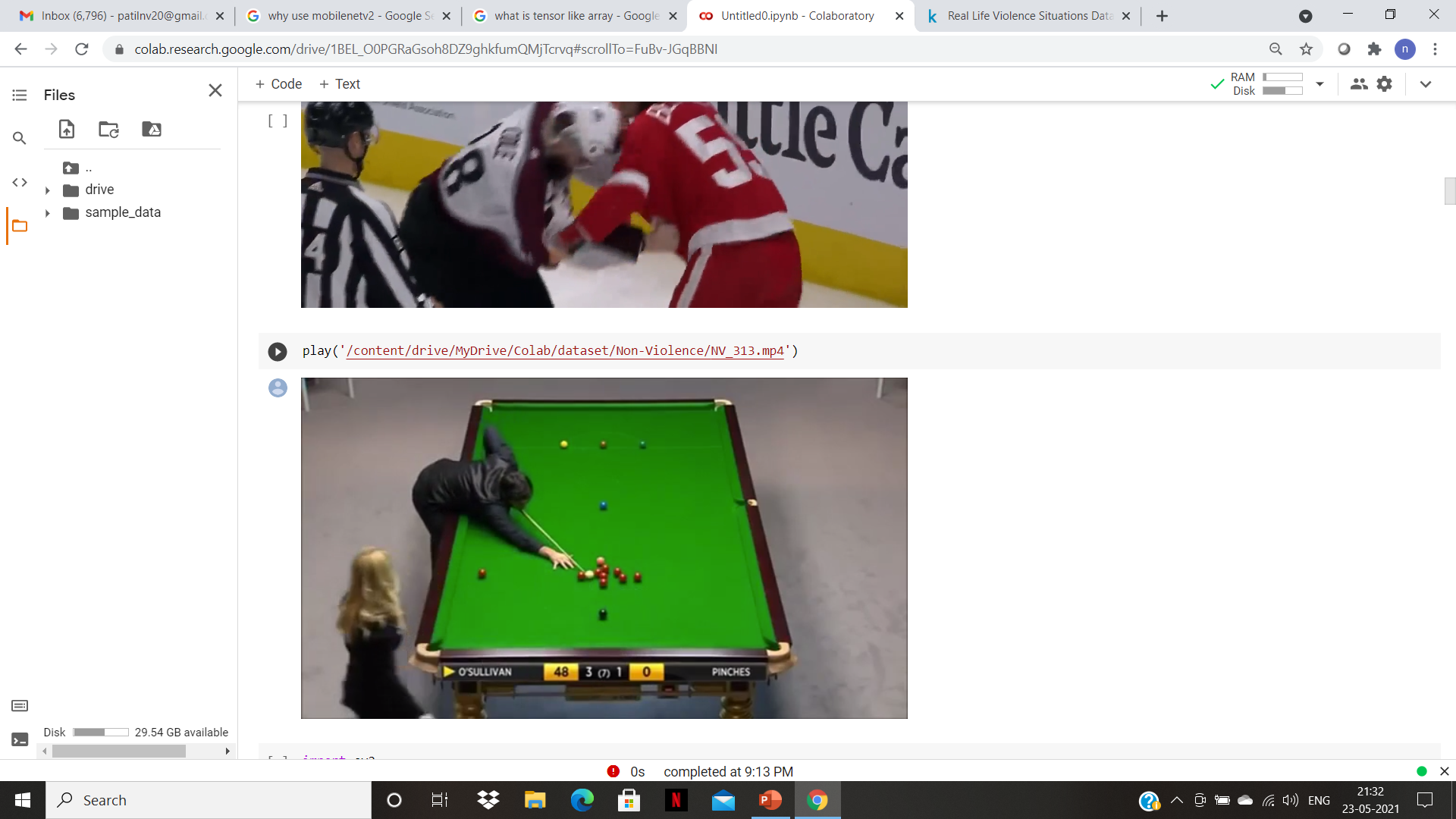


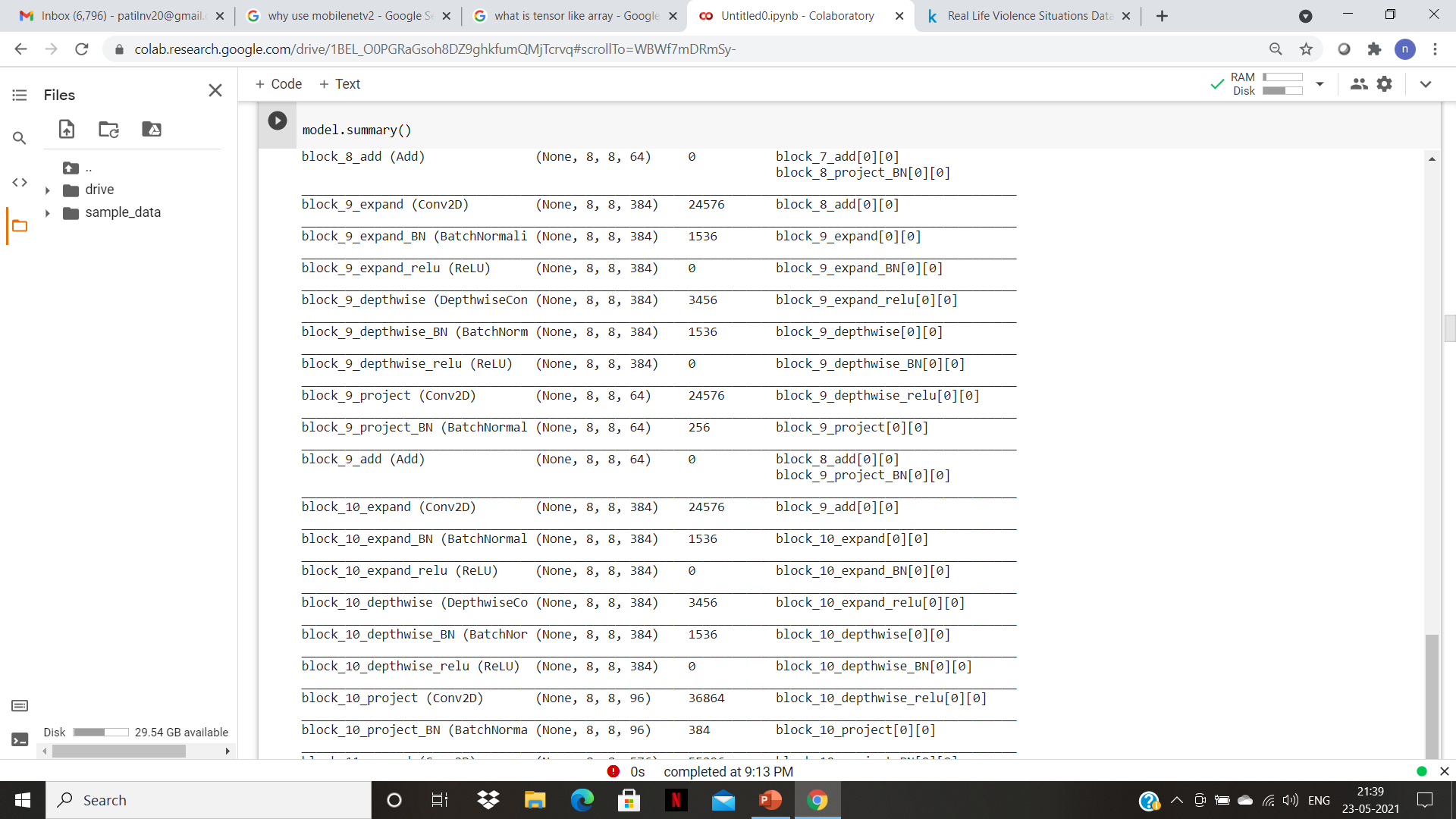
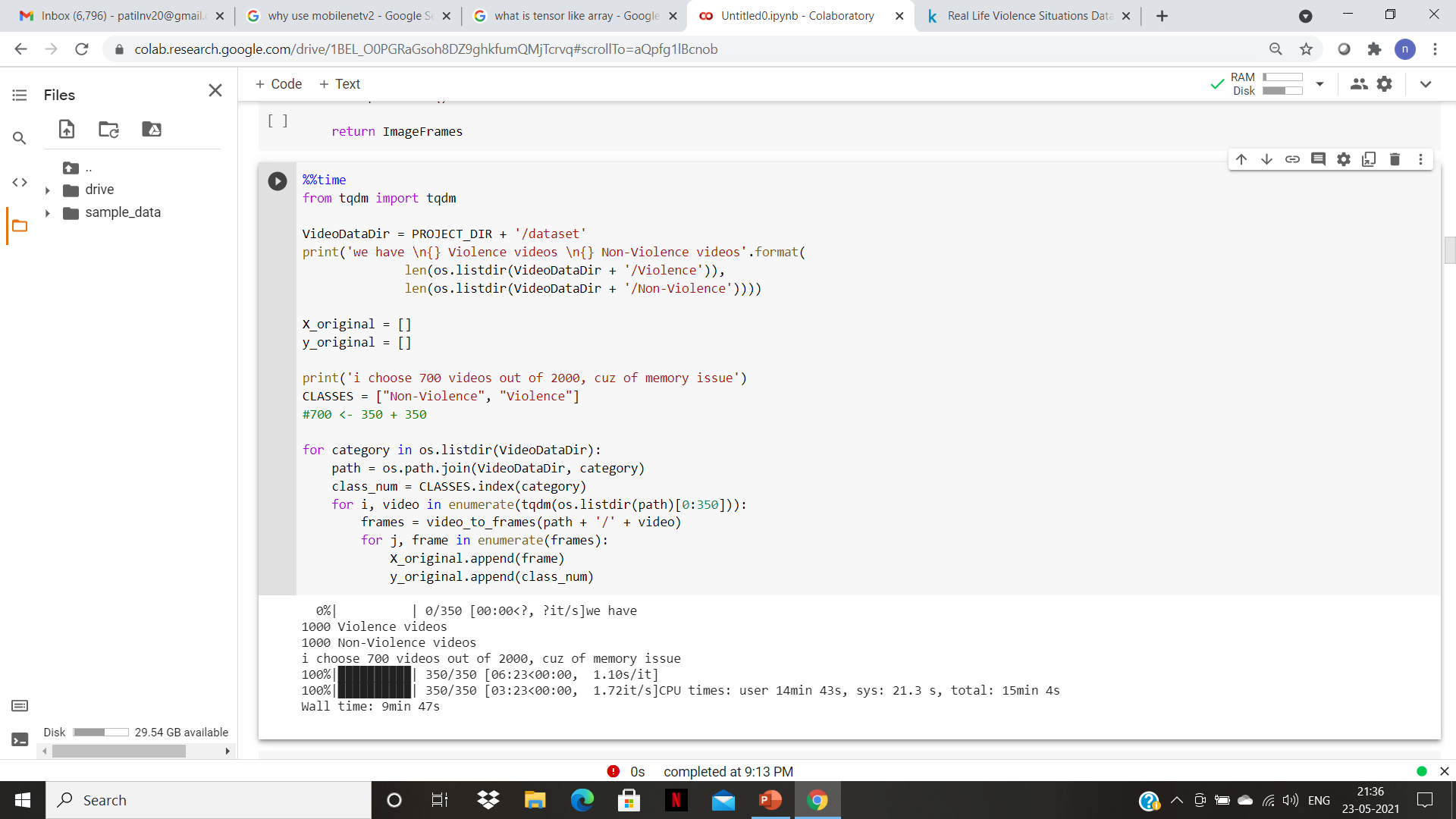
**8.2 Code Snippets**

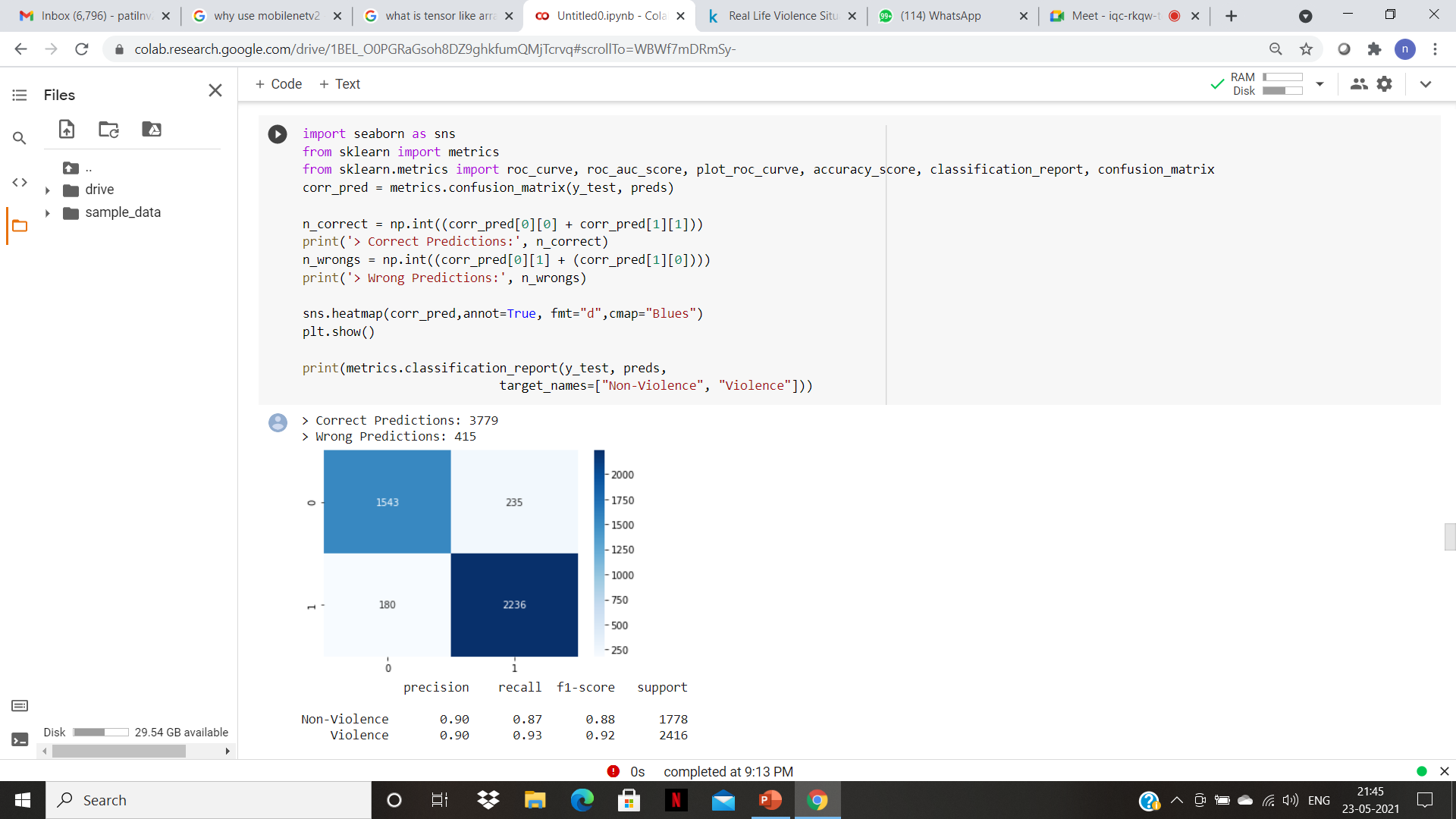
Connection to drive

****

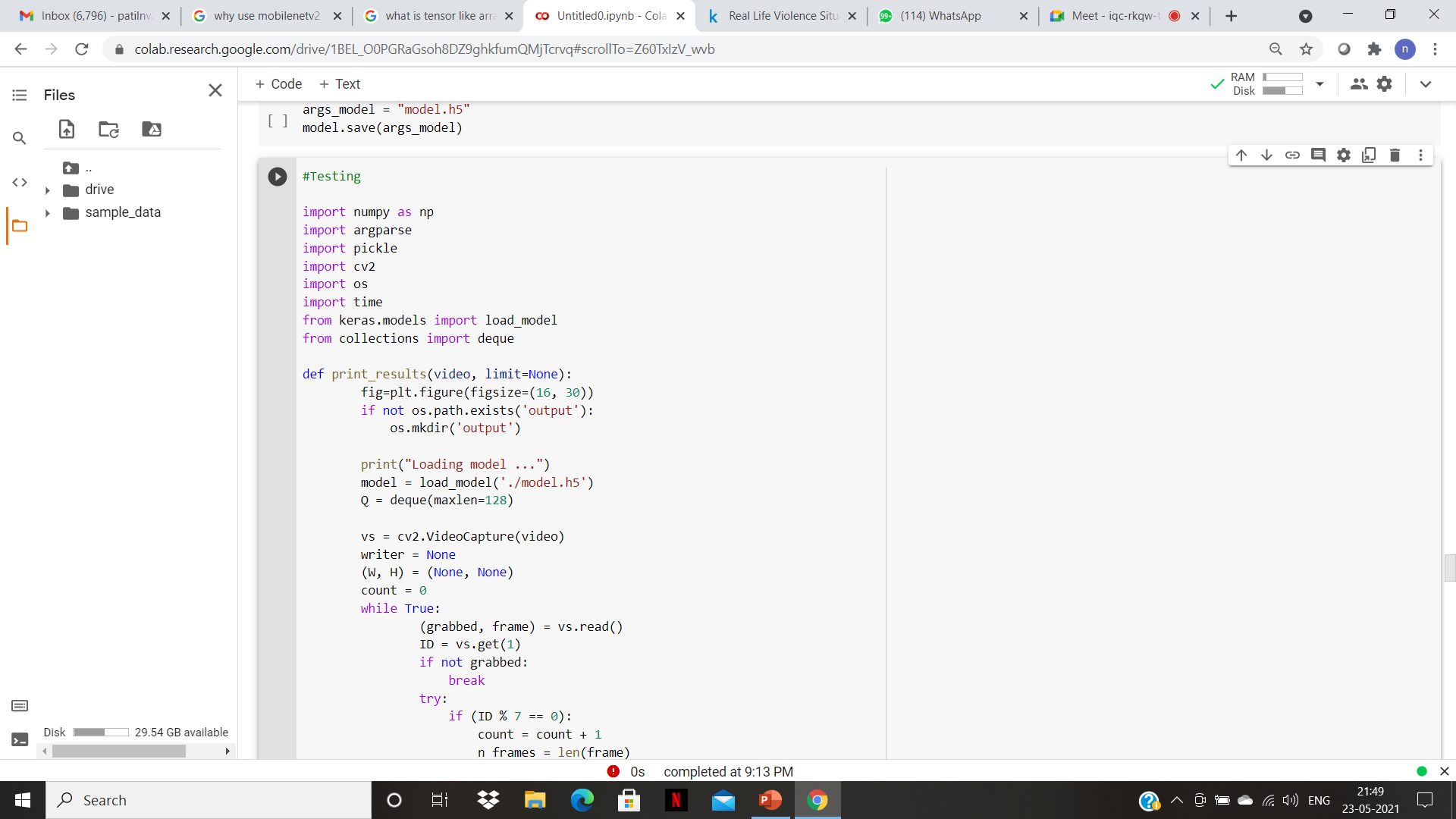
Sample Video Data

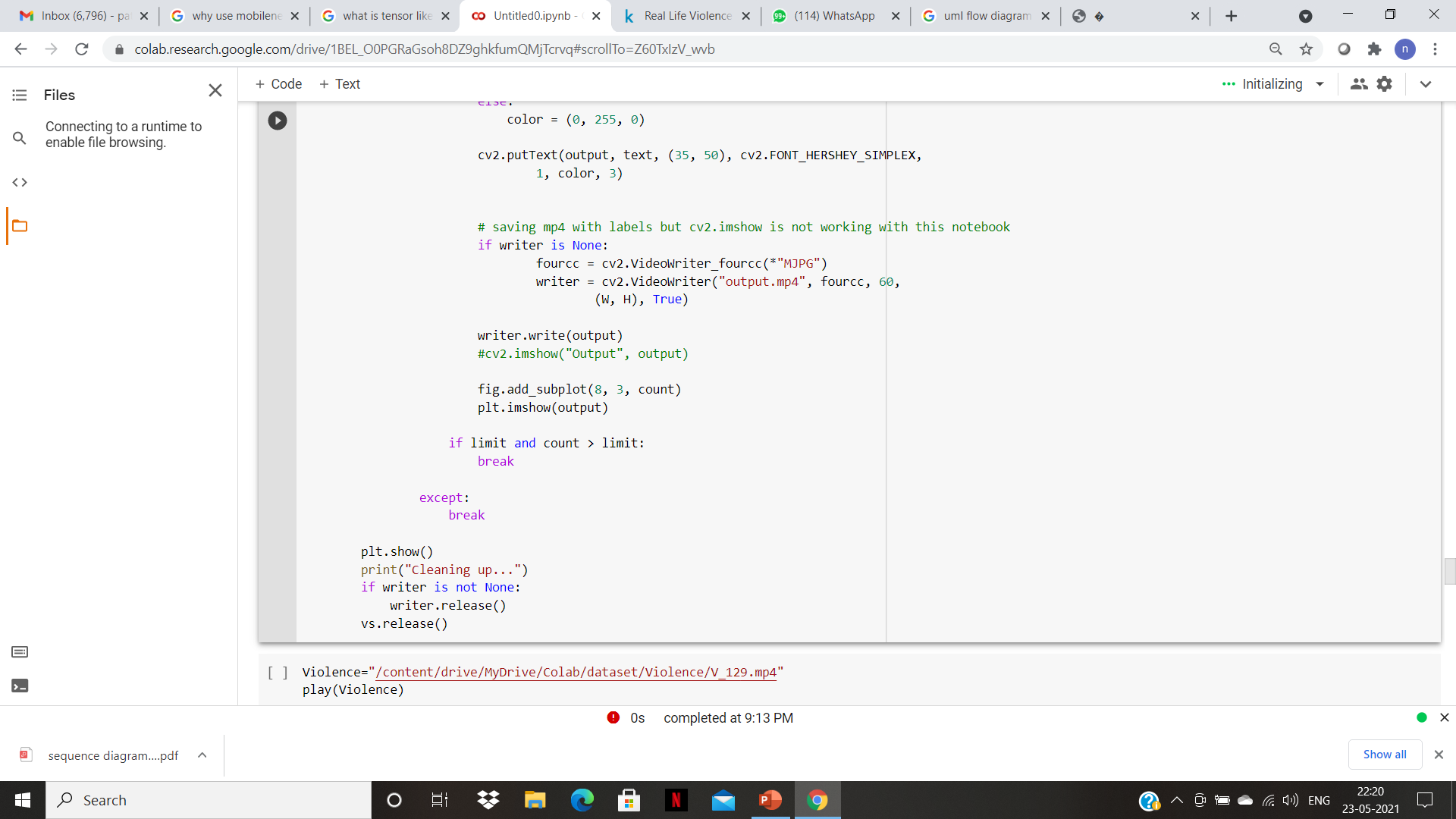




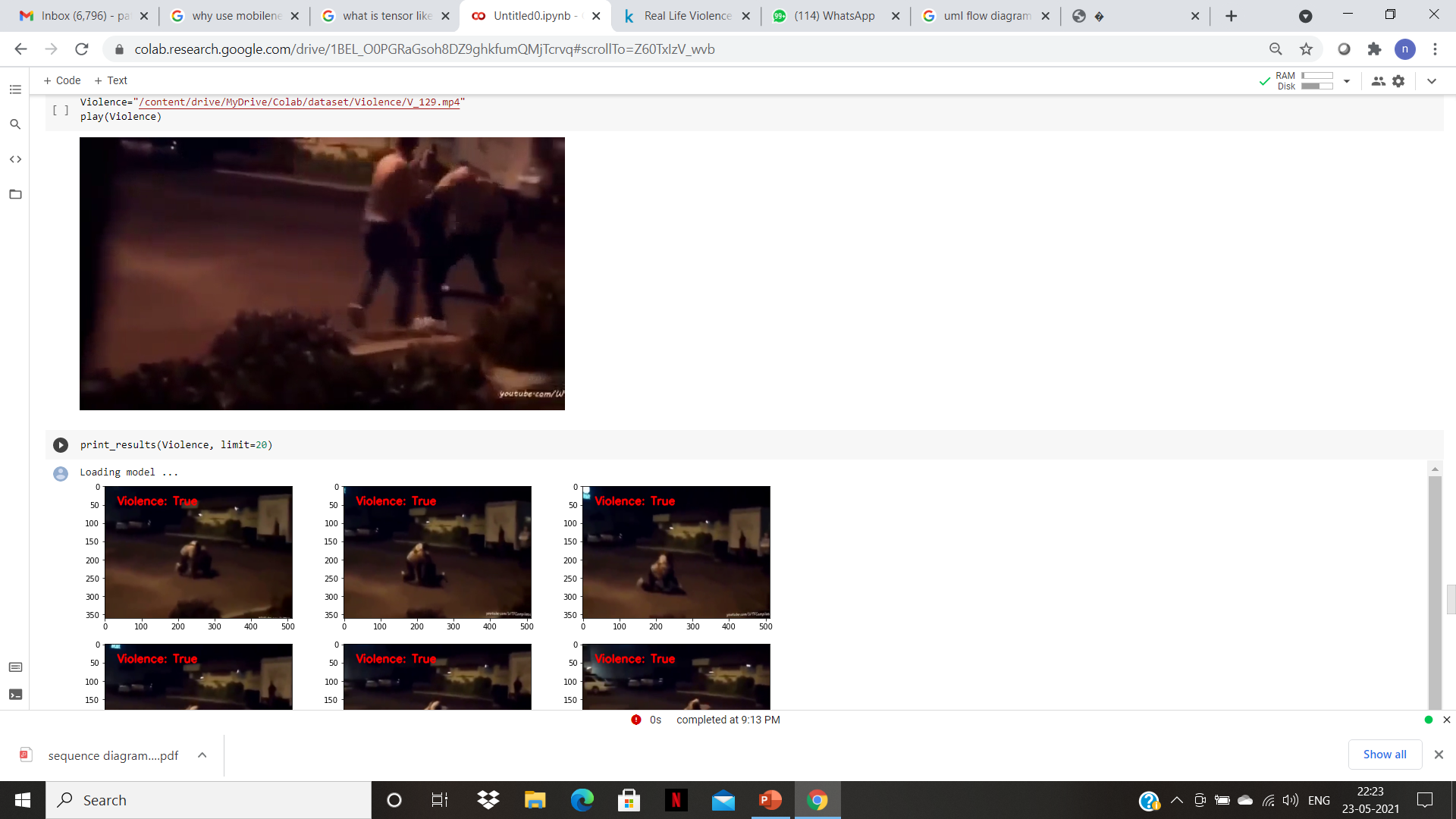


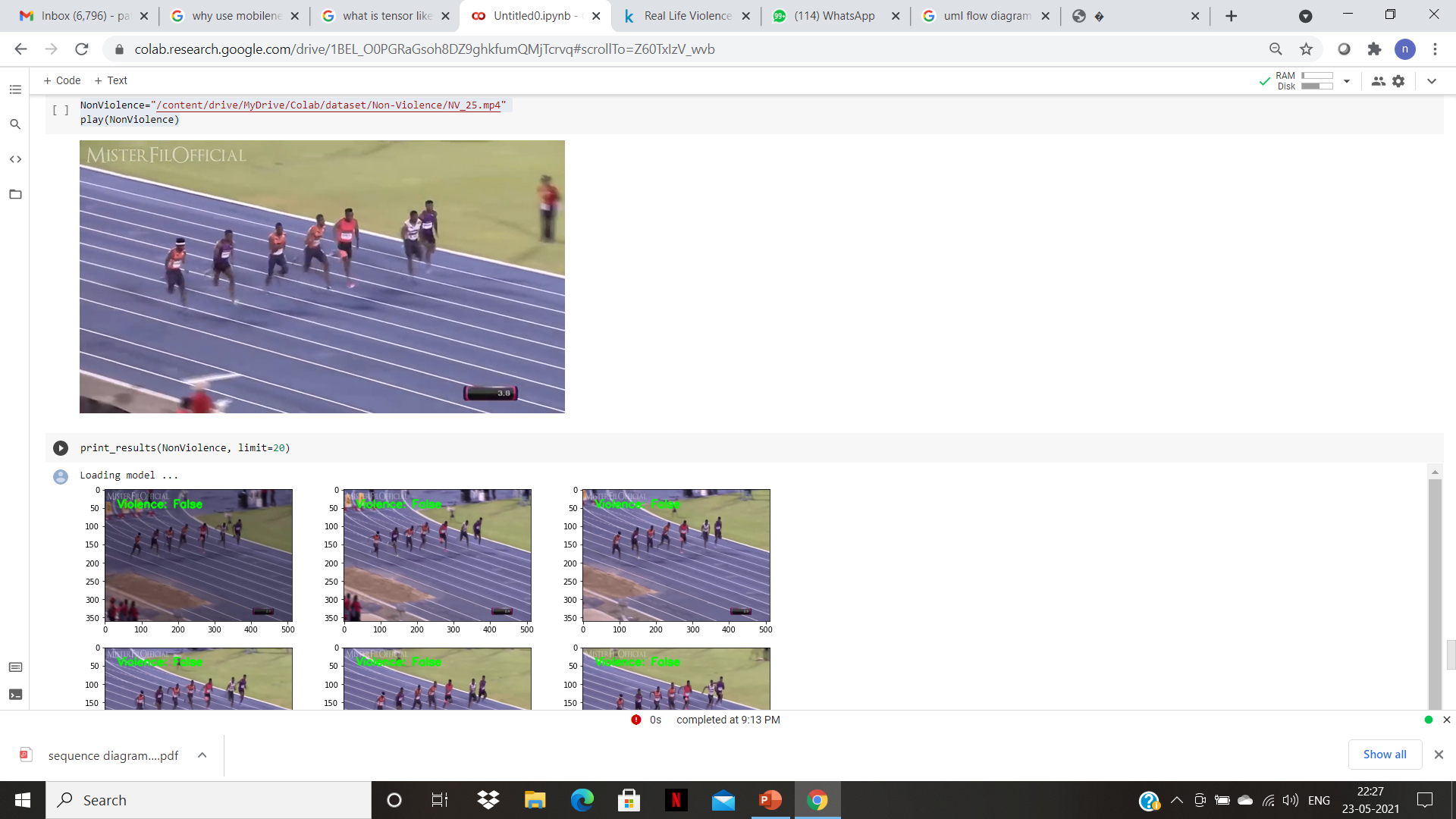
**8.3 Testing Snippets**

****

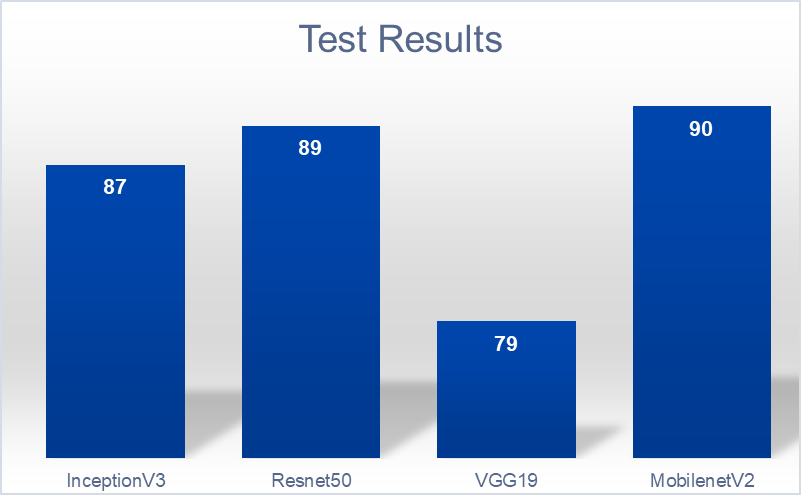


**8.4 Final Result**

****

****

**9. Result Analysis**



Here we have referenced the results and compared these architectural models which have used similar dataset .The most preferred architectures used for CNN are inceptionV3 , Resnet50 , VGG19 due to their heavy hidden layers.As we can see , the classifer which we have used is at par in terms of accuracy with other models.Also MobileNet is faster and can be reliable over smaller dataset.Hence MobileNets achieve this performance by reducing dramatically the number of learnable parameters, which also makes them faster and easier to train compared to more traditional networks.

**10. Conclusion**

Thus ,we have proposed an efficient framework which can detect violence in sensitive areas and detect weapons by analyzing the video streams collected from the surveillance cameras.The proposed model consists of a convolutional neural network (CNN) for frame level feature extraction.

Therefore we implemented deep learning model to predict violence in video data, We found that our implementation dealt well with this task even though our GPU power was relatively low.The data augmentation process helped the model to deal with the small amount of labeled data, the augmentation increased the amount of samples and helped the model to find meaningful patterns in the frames. Also, we found that retraining the CNN networks improves the performance of the network dramatically.

Thus we have proposed an efficient framework which can detect violence in sensitive areas and detect event of violence by analyzing the video streams collected from the various sources. Therefore with the help of video surveillance we can analyze the presence of violence and a particular event of violence using Deep learning techniques.

**11. Future Scope**

The planned system solely detects suspicious human behavior.As present and future work, we are evaluating reducing the number of false positives, by preprocessing the videos, i.e. increasing their contrast and luminosity, and also by enriching the training set with weapons in motion.We can also use an alarm system to report violence to the nearest police station in near future.Detection of fire and different weapons can be enforced in future.Real-time autonomous drone surveillance system can be developed to identify violent individuals in public areas.

**12. References**

**12.1 Journal References**

1. Swathikiran Sudhakaran, Oswald Lanz,” Learning to Detect Violent Videos using Convolutional Long ShortTerm Memory,” 978-1-5386-2939-0/1720 IEEE 2017
2. Prakhar Singh, Vinod Pankajakshan, “A Deep Learning Based Technique for Anomaly Detection in Surveillance Videos, “Twenty Fourth National Conference on Communications 2018.
3. S.M. Rojin Ammar Md. Tanvir Rounak Anjum Md. Touhidul Islam,“Using Deep Learning Algorithms to Detect Violent Activities”.
4. Lyu, Y., Yang, Y., “Violence detection algorithm based on local spatio-temporal features and optical flow,”. 2015 International Conference on Industrial InformaticsComputing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), pp. 307– 311. IEEE, December 2015.
5. Wei Liu et al., “SSD: Single Shot MultiBox Detector”, European Conference on Conputer Vision, Volume 169, pp 20-31 Sep. 2017.
6. Jain, H., Vikram, A., Mohana, Kashyap, A., & Jain, A. (2020). Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications. 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). doi:10.1109/icesc48915.2020.9155832.
7. S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong and W.-c. Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting 2015.
8. J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Dar-rell. Long-term recurrent convolutional networks for visual recognition and description2015.
9. Lyu, Y., Yang, Y., “Violence detection algorithm based on local spatio-temporal features and optical flow,”. 2015 International Conference on Industrial Informatics Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), pp. 307– 311. IEEE, December 2015
10. Ruben J Franklin et.al., “Anomaly Detection in Videos for Video Surveillance Applications Using Neural Networks,” International Conference on Inventive Systems and Control,2020.
    1. **Web References**
11. <https://ieeexplore.ieee.org/document/9014714>
12. https://arxiv.org/pdf/1702.05147.pdf
13. https://models.roboflow.com/classification/mobilenetv2-classification
14. [www.tensorflow.org/tutorials/images/transfer\_learning.](http://www.tensorflow.org/tutorials/images/transfer_learning.)