FPT UNIVERSITY QUY NHON

ARTIFICIAL INTELLIGENCE DEPARTMENT

ABNORMAL BEHAVIOR DETECTION IN ONLINE EXAM

SUBJECT: DEEP LEARNING

Instructor: Nguyen Vo Thanh Khang

Implemented: Alphabet team – AI17B

Nguyen Thanh Dat – QE170110

Dang Thi Le Chi – QE170156

Le Huy Hoan – QE170006

Ngo Dinh An – QE170066

Ho Ton Bao – QE170017

QUY NHON 2023

Table of Contents

I. Introduction

1. Problem statement:

According to information from the law library, each exam room is assigned 2-3 invigilators. And more than 70% of students have cheated on an exam according to a survey by the American Educational Testing Service (ETS). In contrast, more than half a century ago, this number was only about 20%. Truly an interesting number. Grasping that problem, our team created an Abnormal Behavior Detection In Online Examination project. The project aims to develop a system that can detect and recognize abnormal behavior during university online exams. The project utilizes computer vision and deep learning techniques to monitor and analyze student behavior to identify any actions that may indicate academic dishonesty or irregularities.

2. Solution:

The system works by receiving image frames as input during online testing and then classifying them as "normal" or "abnormal" behavior. So, the key point of this project is to determine what is normal behavior and what is abnormal behavior. You will have to follow exam regulations such as the student's laptop having a camera and being operable. When taking the exam, do not cover the camera. If you cover up, you are cheating. Classroom conditions: good lighting conditions (bright lights). As for normal behavior, when taking the online test, you will sit close to the computer, you can sit solemnly or lean back. As for abnormal behavior, there are more cases. For example, you bend down to look at your phone. Turn left, turn right to copy other people's work. Or lean forward. Some cases such as wearing masks and dark glasses will also be considered abnormal behavior. The system serves as a support mechanism for proctors during online exams. If a student violates prescribed behavioral norms and triggers an abnormal behavior warning that exceeds a set threshold, the test will be automatically locked. However, in special cases, students can continue taking the exam by entering the token sent by the system to the proctor, thereby allowing the student to continue and complete the exam under supervision.

3. Value:

The Abnormal Behavior Detection In Online Examination project holds significant value by offering a robust system that supports proctors in upholding exam integrity and curbing irregular behavior during online assessments. By accurately identifying abnormal behaviors, including instances like unauthorized device use or suspicious movements, this solution contributes to a substantial reduction in cheating occurrences during exams.

The system serves as a support mechanism for proctors during online exams. If a student violates the stipulated behavioral norms and triggers an abnormal behavior warning surpassing a set threshold, the test will be automatically locked. However, in exceptional cases, students can proceed with the exam by inputting a token sent to the proctor by the system, thus allowing them to resume and complete the exam under supervision.

This solution aims to provide a robust support system for proctors, ensuring adherence to exam guidelines and preventing potential cheating or irregular behavior during online assessments.

4. System diagram workflow:

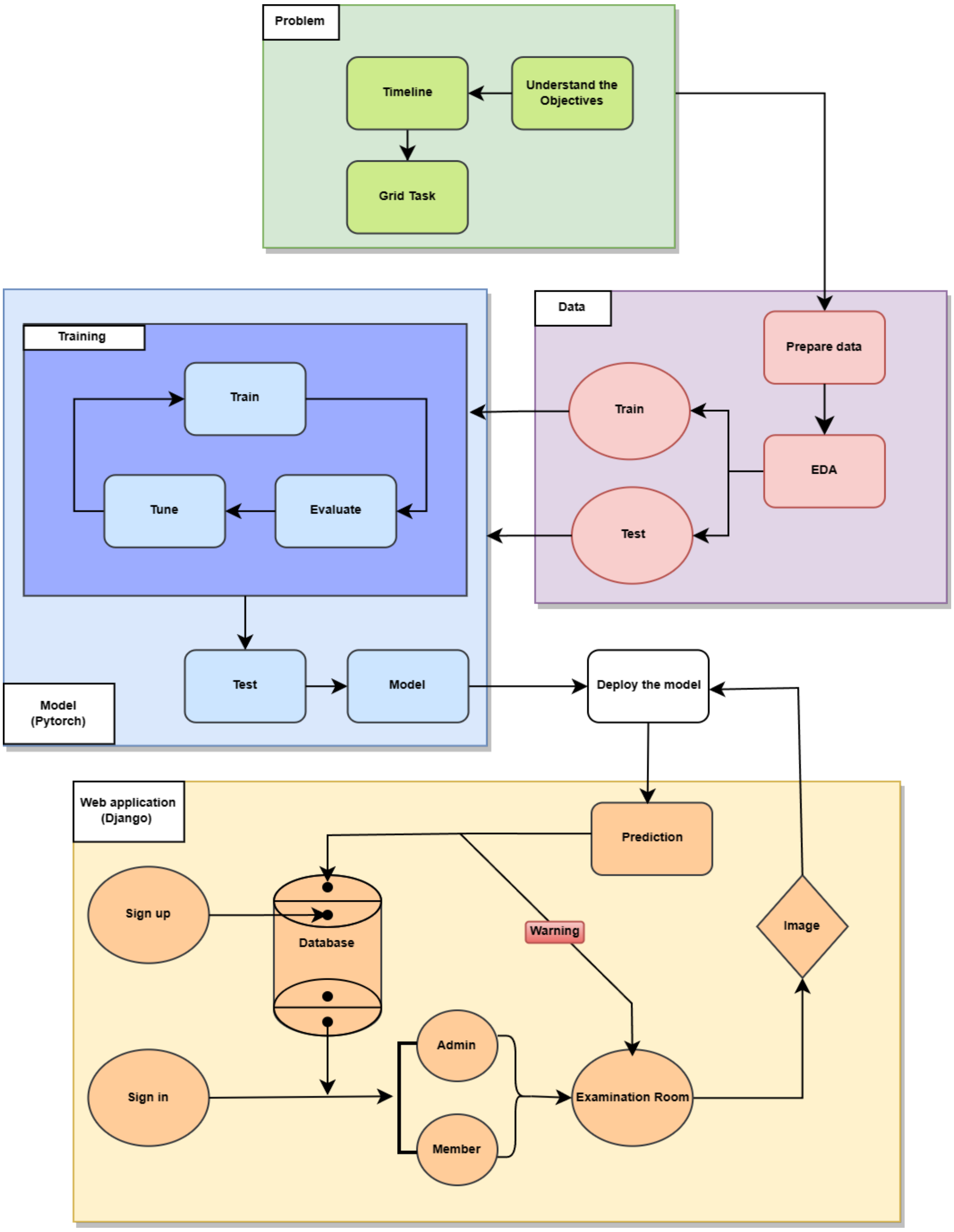

Image 1. System diagram workflow

II. Methodology

1. Data:

This part discusses the methodology and the process while we prepare data. My colleagues and I had searched for valid datasets, but in this case, because we want datasets that contain images that were taken by a laptop webcam and can contain our abnormal criteria, it is not available. Therefore, we decided to collect data by ourselves. Firstly, we collected 400 images from 2 people in my team, in detail 200 images/class.

After that, we discovered that our dataset did not cover all abnormal cases, so we collected more than 400 images from the 2 teammates which contained many abnormal cases. In conclusion, we collected a total of 800 images from 4 people in my team and included relatively complete abnormal cases, such as students looking right/left or wearing sunglasses while taking exams.

After data collection, we split images into a train set and a test set with a ratio of 8-2. As a result, the train set has 638 images and the test set has 162 images with an image size is 640x480. Why do we choose that size? Because that size is a common webcam resolution in most modern laptops and we don't want to resize the images, this will make the image distorted. One more reason, in most abnormal cases, students tend to look right/left and we want an observable area as high as possible in the width of the image. So we decided to use the original size of the webcam. Next step, we proceed to transform images into landmark points using Mediapipe.

With each image, Mediapipe returns 468 landmark points on a human face, at first we just select randomly 23 points and return the x, y, and z coordinates. Below is our development process:

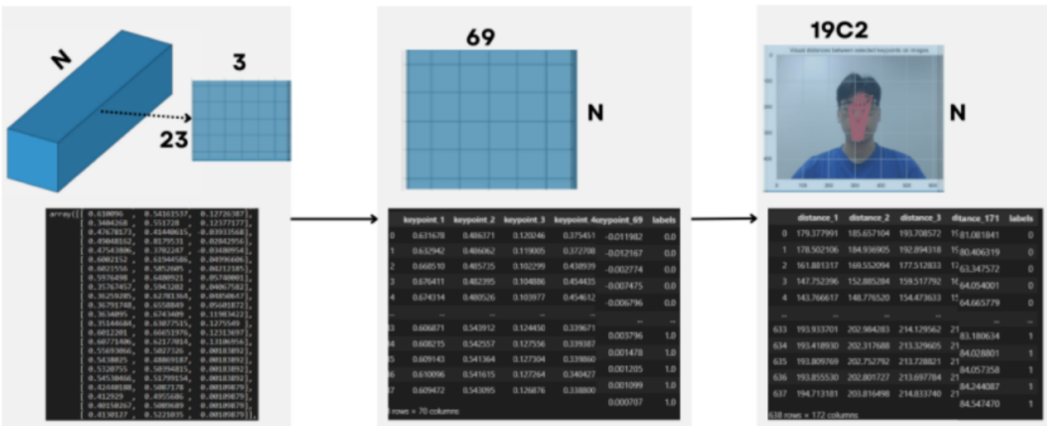

Image 2. Development process

In the first stage, we return the x, y, and z coordinates values and return it in a line, 23 points are 23 lines. With each image, we have a 23x3 matrix representing one image.

In the second stage, we flatten the 23x3 matrix to one line, which means one line contains all points in one image. So we have 69 features.

In the third stage, we discovered that the relation between each point is better, so we decided to calculate pairwise distances with just 19 points (we will explain why we choose 19 points). So we have 171 features.

How to choose landmark points?

Mediapipe API returns 468 landmark points on the human face (more than 10 points for iris eyes). My team will choose the points that have the biggest change between Normal and Abnormal classes. We will find that in FindAllLandmarks.ipynb as below:

- Step 1: Take all landmarks points in one image and calculate distances pairwise.
- Step 2: Calculate mean distance value of all images in one class (Normal/Abnormal), return distance_matrices_normal and distance_matrices_abnormal variables.
- Step 3: Subtract two variables above to find the difference between two classes.
- Step 4: Find the biggest sum on landmark distances (represent for biggest difference), and take the first 20 biggest distances. As a result, we have 17 landmarks and 2 iris center points.

Finally, we discovered some problems in datasets and preprocessed data (clear all zero rows) and saved it to the ".npz" file, the remaining total data is 702 (563 for train set and 139 for test set) and proceeded with the training process.

2. **Model:**

For traditional binary classification problems in computer vision, we require CNN layers to extract features from images and utilize them for classifying into 0 or 1.

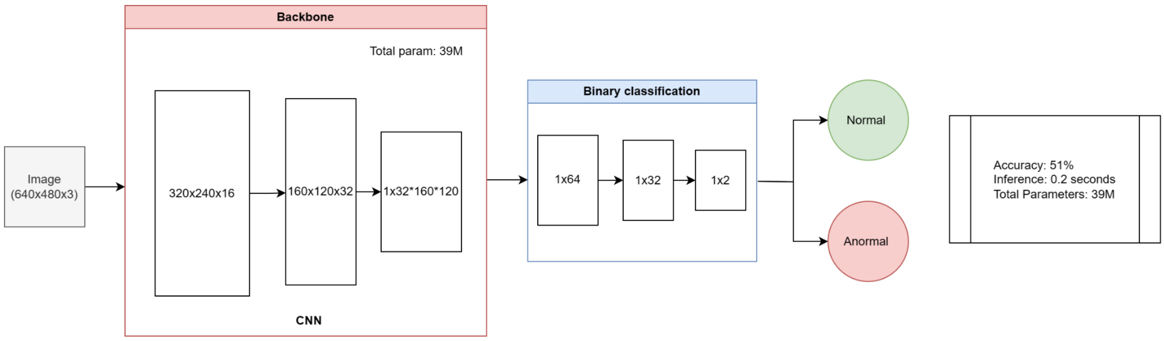

Image 3. Model with traditional CNN

In the traditional solution, we lack visibility into the specific features being extracted, making it challenging to control the classification process. Therefore, we have opted to utilize Mediapipe to extract key points on faces, employing them for model training.

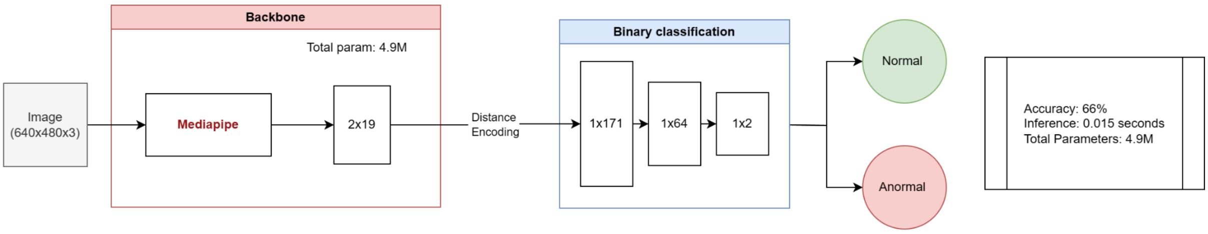

Image 4. Model with mediapipe

We've significantly reduced the parameters of the feature-extracting backbone from 39 million to 4.9 million, leading to a performance boost from 51% to 66%. This enhancement is credited to our precise understanding of classification features using Mediapipe. Focusing on the essential 19 facial points, we generate 171 pairwise distances for effective classification.

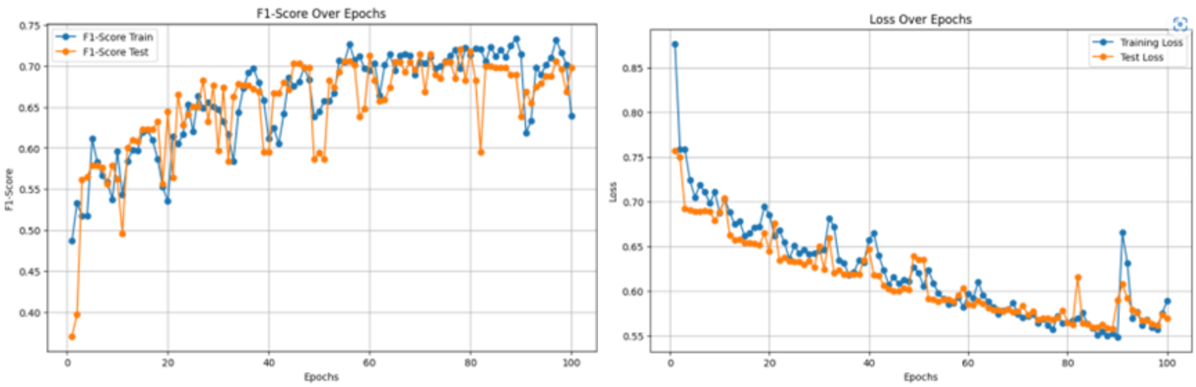

Image 5. Performance

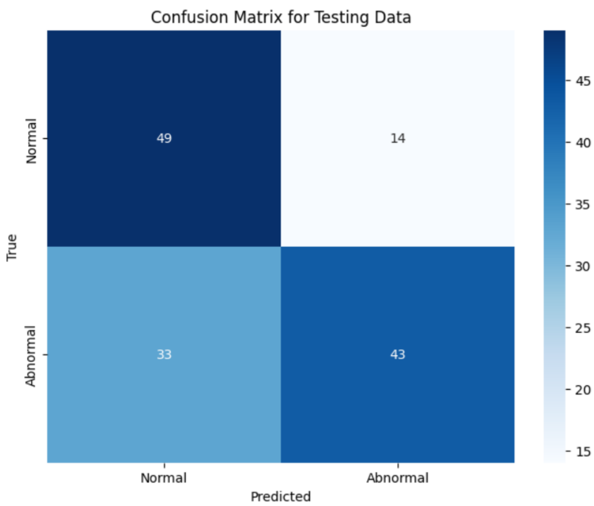

Image 6. Confusion matrix

3. Web:

The implementation methodology of an abnormal behavior detection system is integrated into a web platform. This system focuses on proctoring online exams and involves the utilization of HTML, CSS, and JavaScript for the frontend, while employing Django and FastAPI for the backend, including model API.

The web application comprises three primary sections:

- Home Page: Provides an overview of the project and introduces the services offered.
- Members Page: Provides an overview of the project and introduces the services offered.
- Service Pages: Focuses on the online exam service, involving a multi-step process

Web operation:

- User initiation: Opening the camera to capture the user's feed, which is displayed on the left side of the screen.
- Exam interface: The right side of the screen serves as the exam area.
- Frames from the camera are forwarded to the backend and subsequently to the model for abnormal behavior detection. Parameters vary for different devices (e.g., 30fps/s or 60fps/s). After extensive testing, approximately 80% of 160 frames are identified as abnormal behavior (roughly 5 seconds), prompting a warning to the user, locking their exam, generating a random exam code, and storing it in the database.
- An invigilator review is initiated to determine whether the user can resume the exam. If approved, the user receives the exam code and enters it on the exam page to unlock and continue.

This report provides insights into the technical process, including frontend development for user interaction, backend handling using Django and FastAPI to manage data flow, and the implementation of a model API for abnormal behavior detection. The focus remains on ensuring a secure and monitored environment for online examinations through a robust and comprehensive system.

III. Results

1. EDA Data:

Let’s proceed to data analysis, below are some things we have explored:

Check class distribution: we can see that the number of normal and abnormal classes is quite balanced.

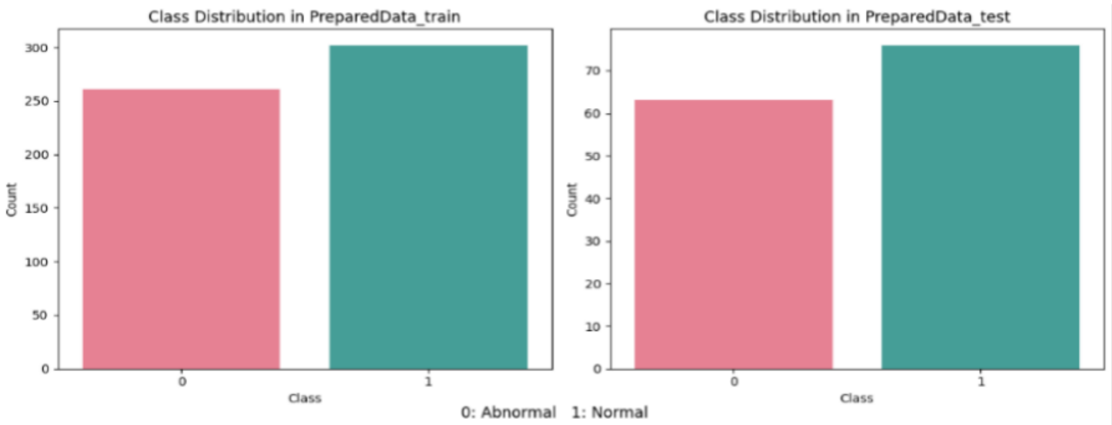

Image 7. Class distribution in train/test set

Data distribution in each class: The distribution in class Abnormal (Labels 0) is different with Normal (Labels 1)

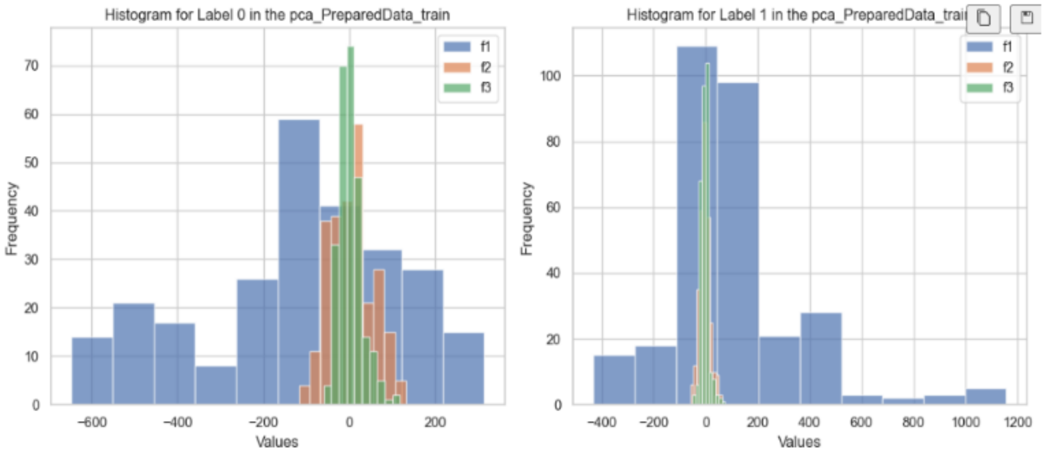

Image 8. Data distribution on each class in train set

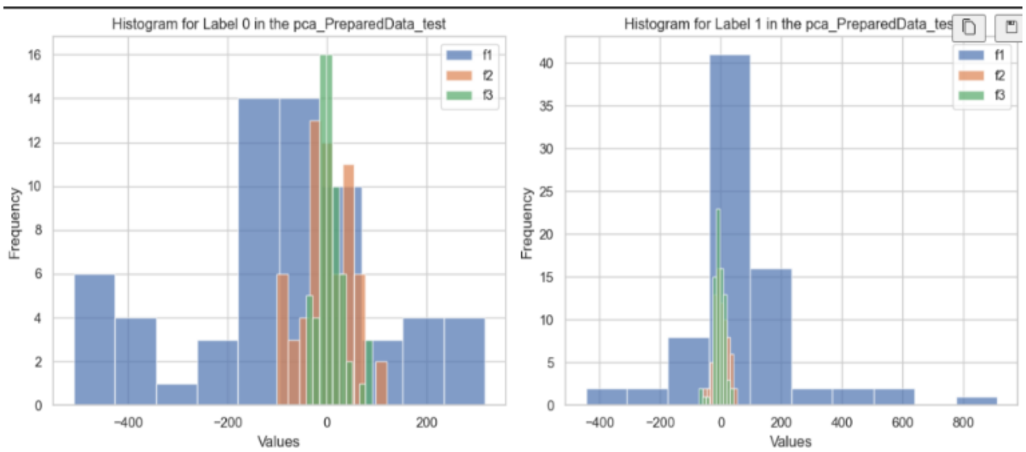

Image 9. Data distribution on each class in test set

PCA Data and see cluster: We can see that the Normal (yellow)/Abnormal (blue) is clearly clustered

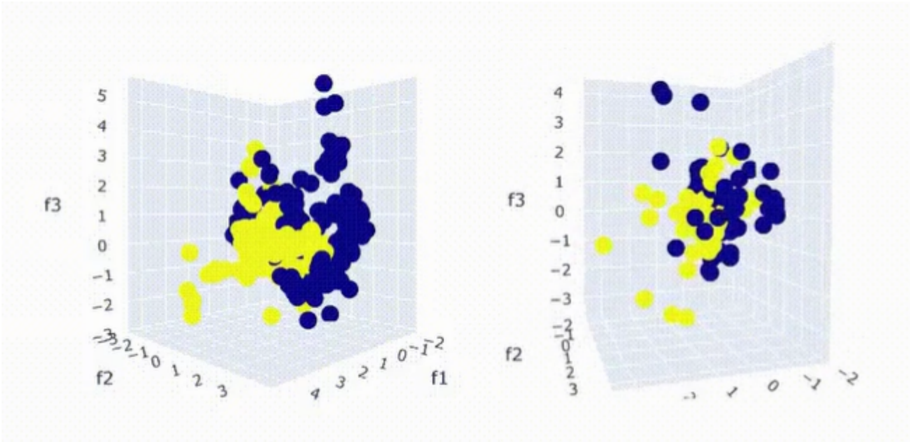

Image 10. Visualize PCA Data in train (left) and test set (right)

Visualize landmark points:

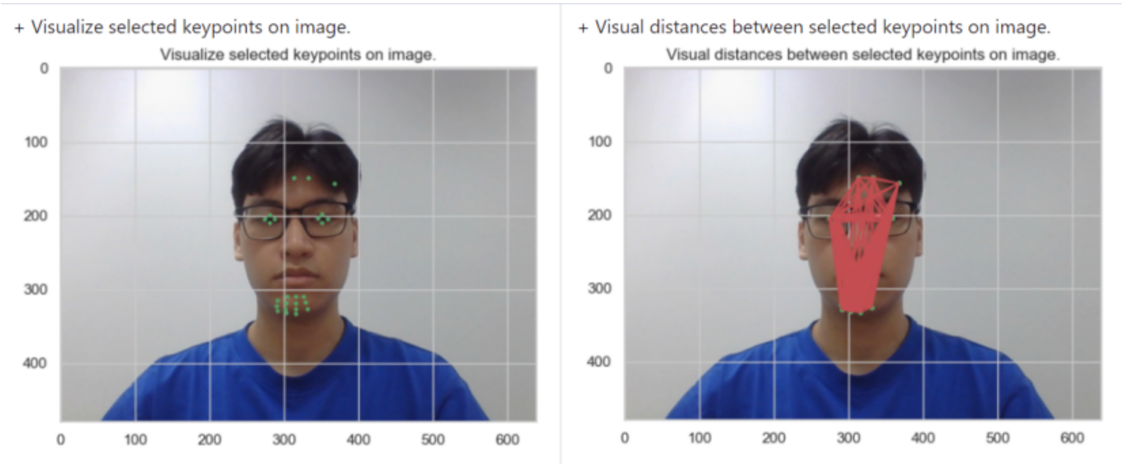

Image 11. Visualize landmark points (left) and distances between them (right)

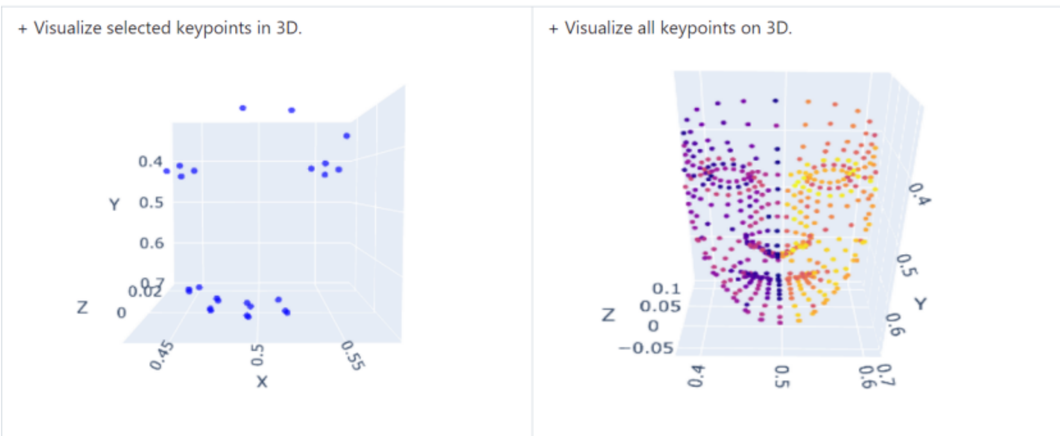

Image 12. Visualize selected landmark points (left) and all landmark points (right) in 3D

2. **Model accuracy:**

Total Frames Processed	**467**
True Prediction	**398**
Precision	**0.9688**
Recall	**0.7778**
Accuracy	**75.87 %**

Table 1. Real Time Testing

IV. Discussion

1. **With the F1-Score 75% evaluation method, does it guarantee accuracy in the real world?**

   Although the model's F1-score is only 75%, we actually tested it by directly labeling each frame and comparing it with the model's prediction, and the F1-score is up to ~88%.

2. **In case there are 2 or more people in the frame, what will the result be?**

   The model still only detects 1 person, and the remaining people will be abnormal on their computers if they look at other people's computers.

3. **What happens if students turn away or close their eyes but they don't cheat?**

   They must sit upright, and focus on the screen because exam room regulations require them to be serious, except when submitting the test.

4. **For subjects related to calculation, students can calculate on paper before circling the answer, the model will predict abnormality, how to solve this case?**

   In the future, the team will develop an abnormal multi-label classification model in the case of subjects related to calculation, and set the case where looking down will be normal.

5. **Why not train one label abnormal, but must train two labels abnormal and normal?**

   Because inference has a set threshold index, train on 2 labels to be able to adjust the threshold to predict abnormal action more flexibly.

V. Conclusion

The Abnormal Behavior Detection In Online Examination project stands as a pivotal step towards enhancing exam integrity and mitigating irregularities during online assessments. Through the implementation of computer vision and deep learning techniques, our project aimed to scrutinize student behavior during exams to identify potential instances of academic dishonesty.

Future Directions:

1. **Development of Database System:**
   Implementing a comprehensive database system will fortify data management, ensuring efficient storage and retrieval of exam-related information. This database will serve as a reliable repository for all exam-related data, enhancing system reliability and accessibility.

2. **System Authorization Enhancements:**
   Strengthening system authorization protocols is imperative. Future iterations will include refined authentication mechanisms to ensure secure access and user verification, elevating the system's overall security posture.

3. **Statistical Insights Integration:**
   Incorporating statistical analysis modules will enable the extraction of valuable insights from exam data. This will empower educators and administrators with meaningful statistics, aiding in informed decision-making and performance evaluation.

4. **Model Retraining Protocols:**
   Continuous improvement is pivotal. Our roadmap includes regular model retraining sessions, ensuring our system evolves with changing behaviors and stays adept at detecting newer forms of irregularities.

5. **Multi-Classification and One-Shot Learning:**
   Advancing towards multi-classification capabilities and integrating one-shot learning methodologies are crucial objectives. This evolution will enable our system to accurately categorize diverse behaviors with greater granularity and efficiency, enhancing its adaptability and accuracy.

These planned enhancements signify our commitment to not only enhancing user interaction but also reinforcing the system's fundamental capabilities, ensuring it remains at the forefront of proctoring online exams with utmost reliability and sophistication.