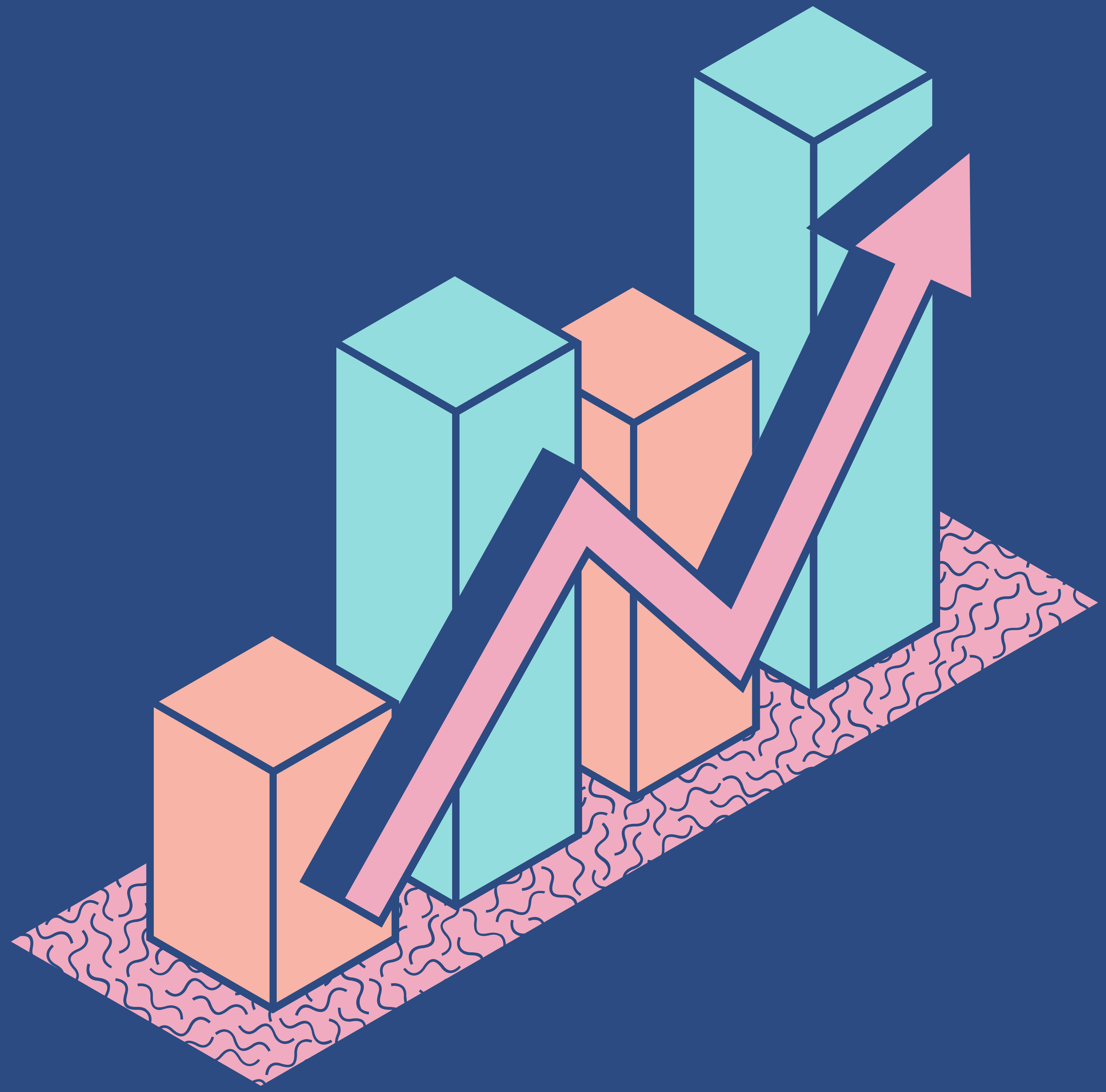
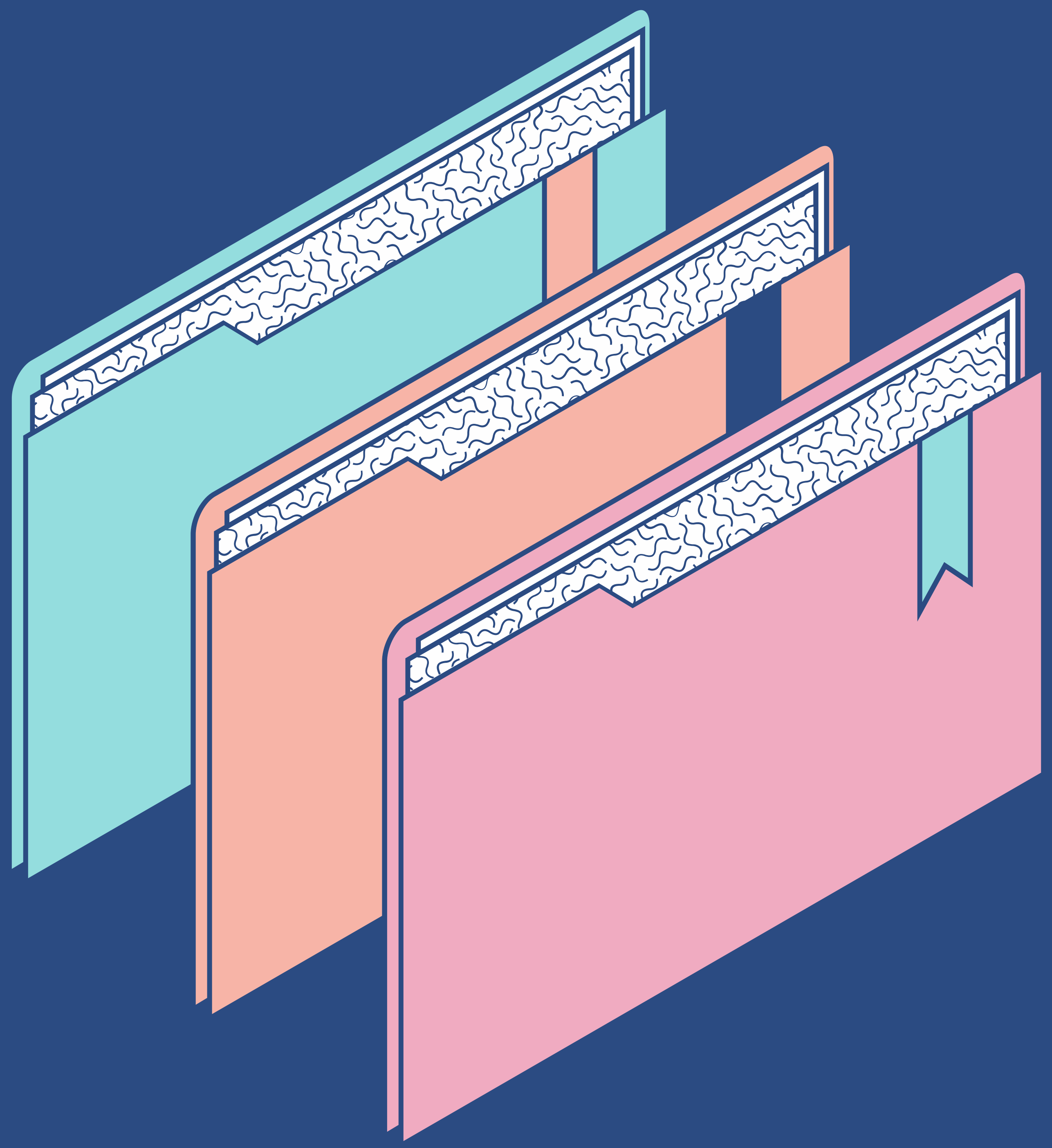
GROUP 8

Classroom Attention Monitoring

Nandan Mandal, Priyanshu Singhal, Vaishnavi Rathi

Problem Statement





Problem Statement

Assessing students' attention and classroom behavior is essential for enhancing the quality of teaching and learning. However, traditional methods, such as classroom observations and questionnaires, are subjective, inefficient, and incomplete for monitoring classroom concentration.

Solution and impact

To address this challenge, we are proposing a computer vision driven system designed for real-time monitoring of student attention within a classroom environment using a camera. This system not only provides continuous attention tracking but also offers an insightful visualization in the form of a post-lecture graph.

Educational Impact

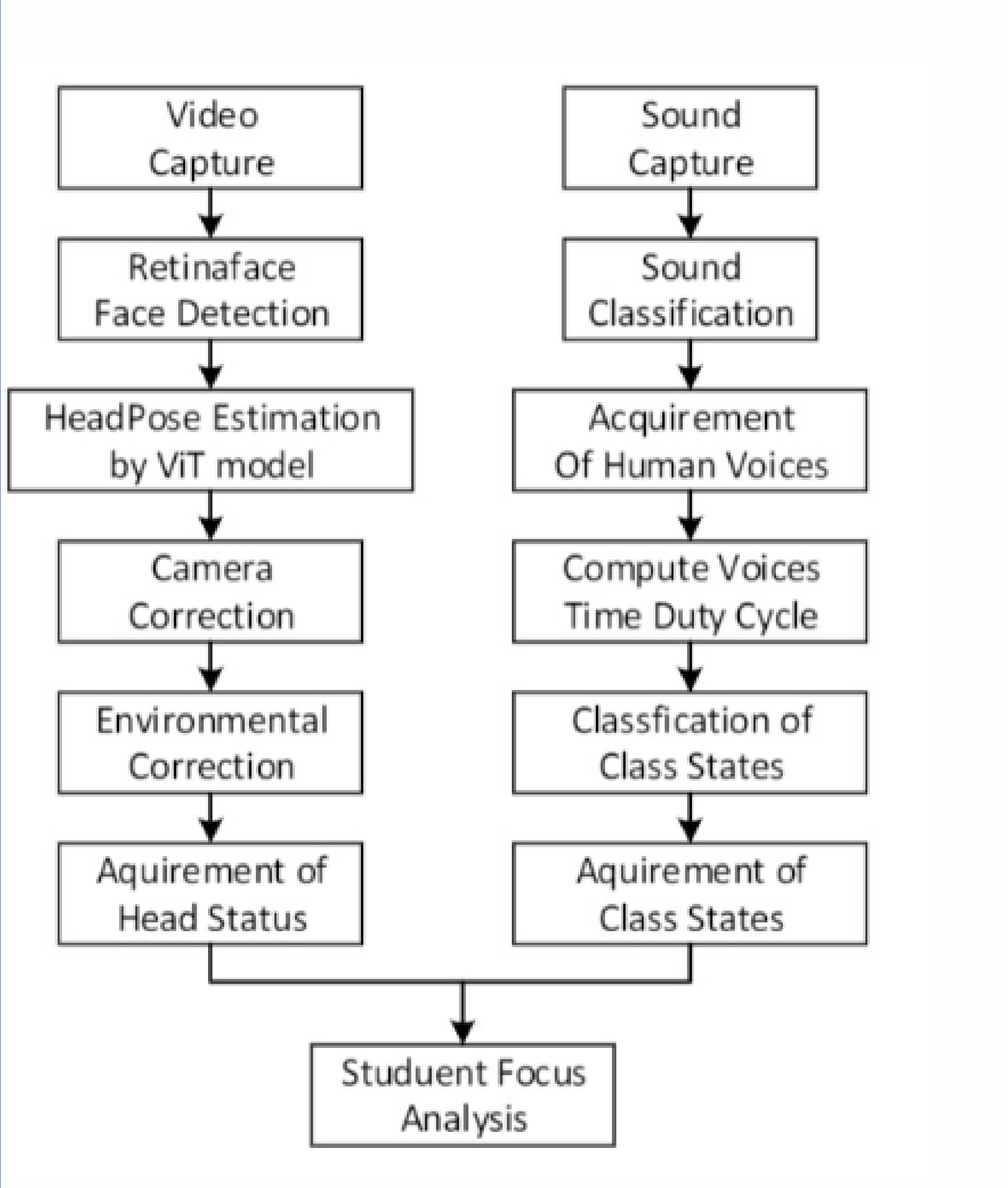
This will help educators assess classroom engagement at different time stamps, enabling them to optimize teaching content and enhance instructional quality.

Enhanced Learning Experience

Students will benefit from a more engaging and tailored learning experience, since instructors will be able to identify what parts of the lecture had low attention scores.

Literature Survey

Paper 1: A Deep-Learning Based Method for Analysis of Students' Attention in Offline Class

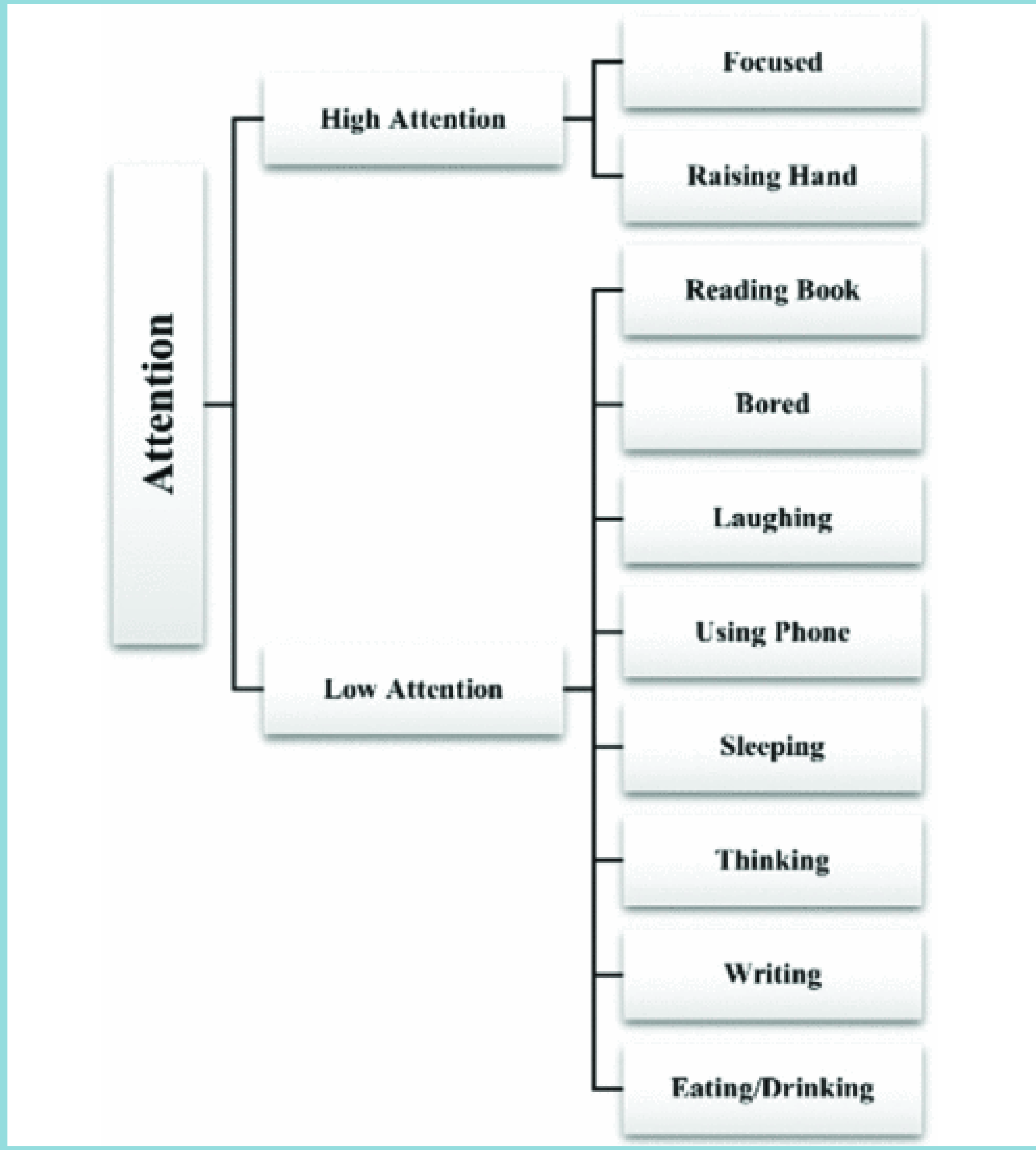


- **The paper proposes a method to measure students' attention in offline classes using deep-learning models**
- **The class-state sequence and each student's head-pose parameters are used to estimate if the student's attention**

	Focused Attention	Unfocused Attention
Lecture	Forward-looking state	Head-down state Head-side state
Interaction	Forward-looking state	Head-down state Head-side state

- Retinaface , ViT (Vision Transformer) , and ASR for face detection and location, head pose estimation, and speech recognition to accurately extract the learning attention of each student
- The paper also corrects the head pose angele for the position of the camera and the movement of the professor.

Paper 3: Smart Classroom: A Deep Learning Approach towards Attention Assessment through Class Behavior Detection



Classification:

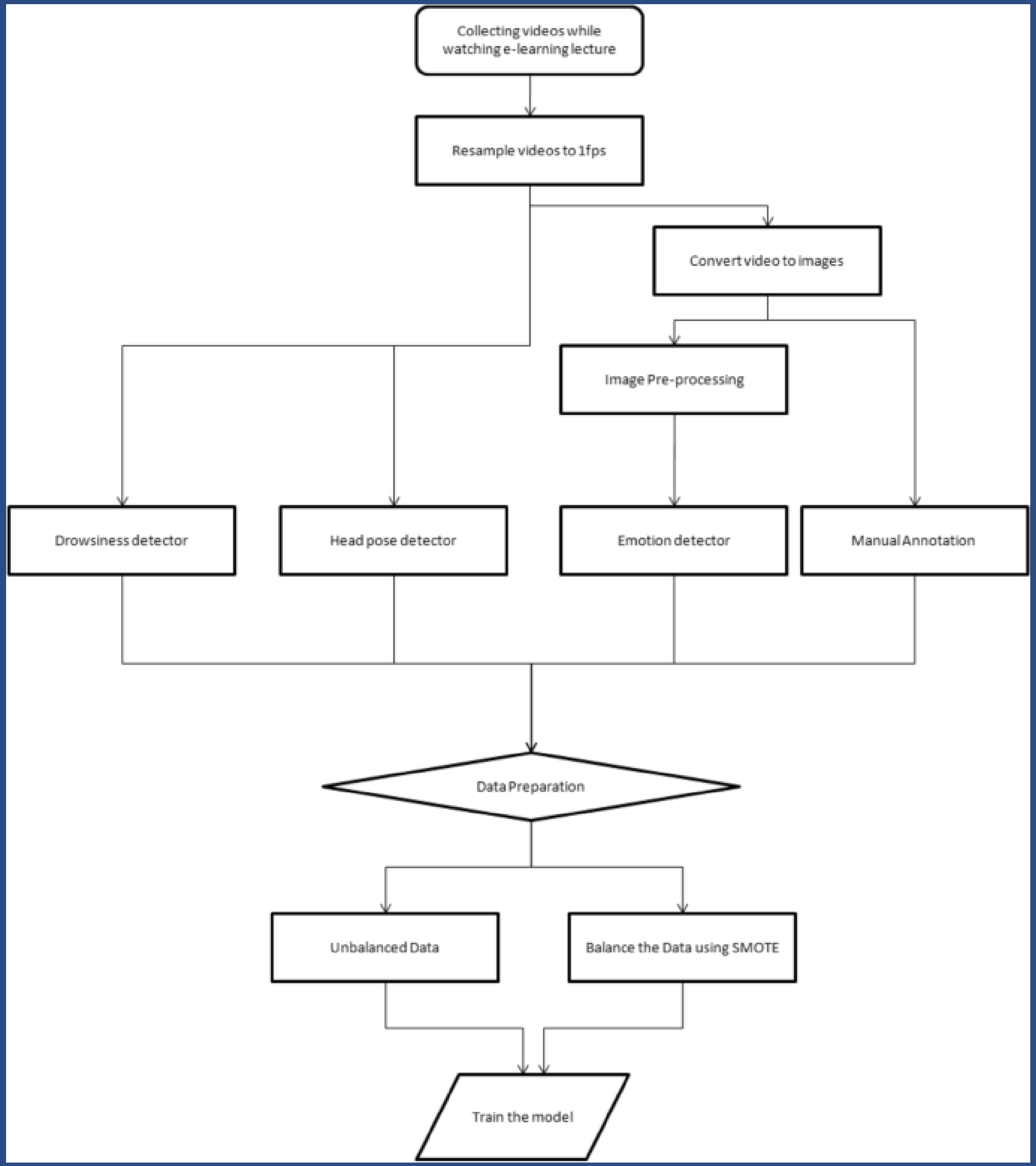
- The proposed system is vision-based and uses a camera to monitor students' behavior in real time.
- It classifies student actions/behaviors into high attention and low attention categories based on raised hands, boredom, eating/drinking, etc
- Emotion Recognition:The system also recognizes students' emotions, including happy, sad, angry, neutral, and surprise, .

ML Algorithms

- The system uses YOLOv5, a deep learning-based object detection model, to monitor students' behavior.
- DeepSORT algorithm is used for tracking students.
- Facial recognition technology, specifically the HaarCascade algorithm, is employed for student identification and attendance

M. M. A. Parambil, L. Ali, F. Alnajjar and M. Gochoo, "Smart Classroom: A Deep Learning Approach towards Attention Assessment through Class Behavior Detection," 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 2022, pp. 1-6, doi: 10.1109/ASET53988.2022.9735018.

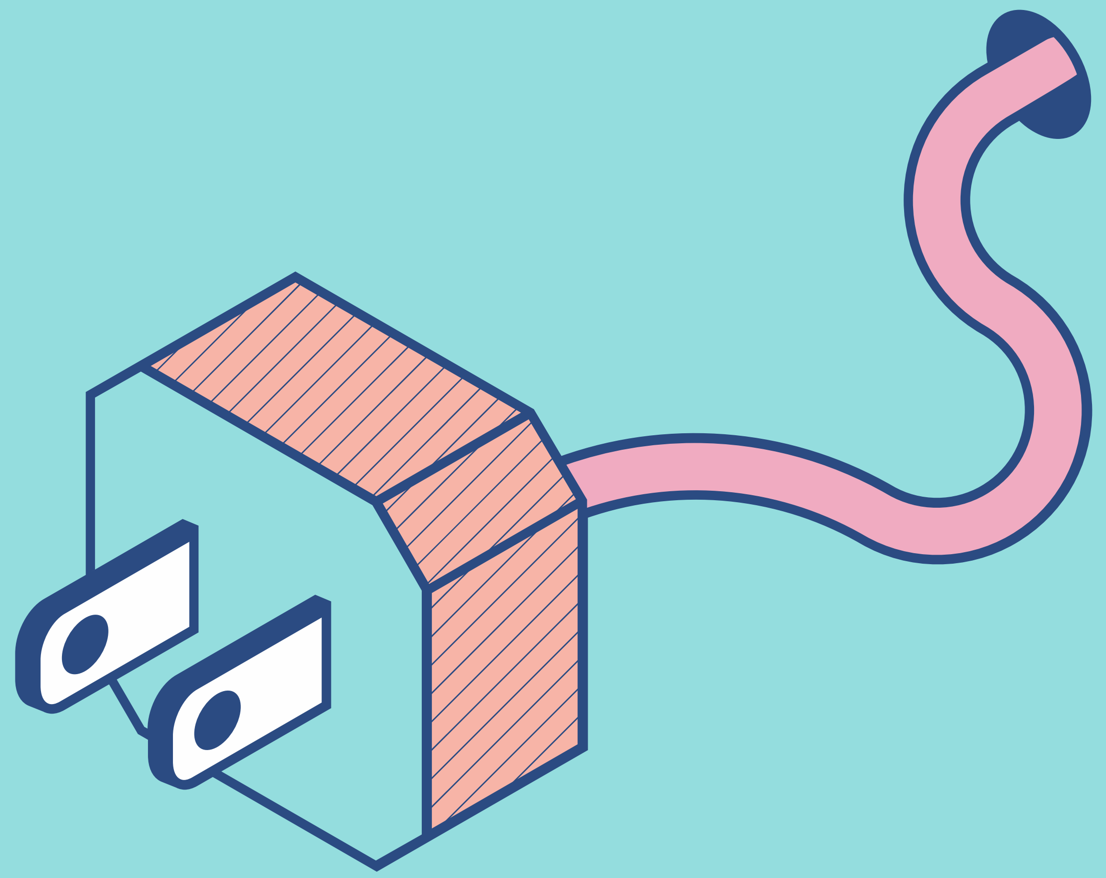
Paper 4: Machine Learning applied to student attentiveness detection: Using emotional and non-emotional measures

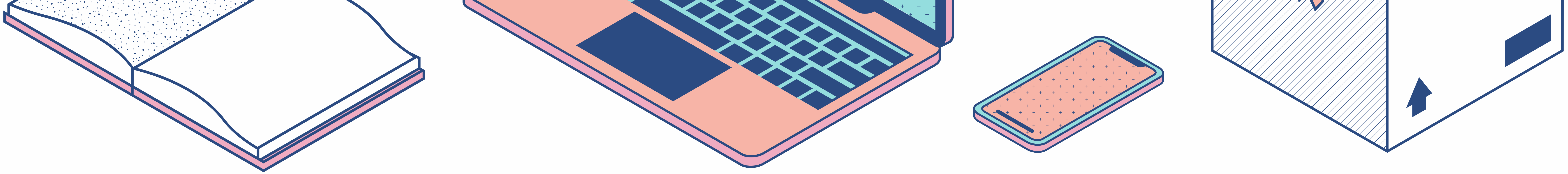


1. Drowsiness Detection:
    - EAR and YAR calculated from facial landmarks.
2. Head Pose Analysis:
    - Determine head rotation angles from facial keypoints.
3. Emotion Detection:
    - Use FER2013 dataset and VGGNet variant model.
    - Detect sseven emotions.
4. Machine Learning Models:
    - Compare decision trees, random forest, SVM, and XGBoost.
5. Metrics:
    - Focus on accuracy and AUROC.
    - Best Model: XGBoost with AUROC OVR: 92.12%, Accuracy: 80.52%.

Elbawab, M., Henriques, R. Machine Learning applied to student attentiveness detection: Using emotional and non-emotional measures. Educ Inf Technol (2023). https://doi.org/10.1007/s10639-023-11814-5

Data collection and Feature Preprocessing





Data Collection

1. Informed Consent: An official consent form was sent to the entire batch of participants. The form clearly explained the project's purpose and the intended use of the collected data. It informed participants that their classroom activities would be recorded and that only the project team members would have access to this data.

Consent form for MLPR project

Our project aims to improve classroom engagement and learning experiences by developing a machine learning model capable of gauging attention levels in a classroom. To achieve this, we need your consent to capture photos during class sessions. All the data will be used only for training purpose and will be kept private and within the team comprising of me (Nandan), Vaishnavi Rathi and Priyanshu Singhal.

1. Do you agree with us taking your pictures in class? *

◯ Yes

◯ No

+ Add new

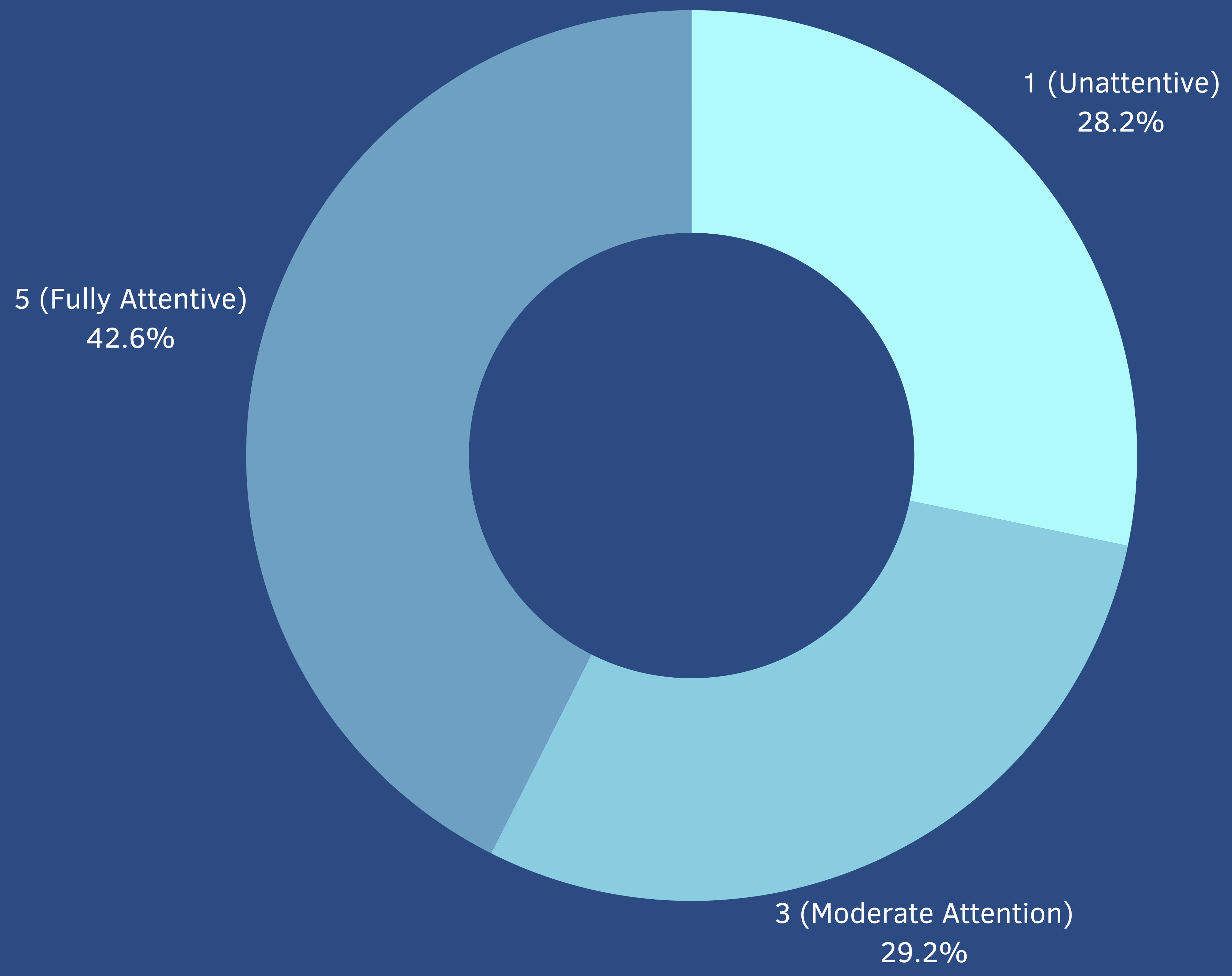
ID	Start time	Completion time	[illegible]	[illegible]	Last modified time	Do you agree with us
1	9-18-23 10:30:15	9-18-23 10:30:18	[illegible]	[illegible]		Yes
2	9-18-23 10:31:38	9-18-23 10:31:41	[illegible]	[illegible]		Yes
3	9-18-23 10:32:02	9-18-23 10:32:08	[illegible]	[illegible]		Yes
4	9-18-23 10:36:51	9-18-23 10:36:54	[illegible]	[illegible]		Yes
5	9-18-23 10:37:24	9-18-23 10:37:26	[illegible]	[illegible]		Yes
6	9-18-23 10:37:44	9-18-23 10:37:49	[illegible]	[illegible]		Yes
7	9-18-23 10:39:19	9-18-23 10:39:22	[illegible]	[illegible]		No
8	9-18-23 10:39:19	9-18-23 10:39:23	[illegible]	[illegible]		Yes
9	9-18-23 10:43:51	9-18-23 10:43:53	[illegible]	[illegible]		Yes
10	9-18-23 10:47:09	9-18-23 10:47:12	[illegible]	[illegible]		No
11	9-18-23 10:48:28	9-18-23 10:48:35	[illegible]	[illegible]		Yes
12	9-18-23 11:04:29	9-18-23 11:04:30	[illegible]	[illegible]		No
13	9-18-23 11:06:27	9-18-23 11:06:29	[illegible]	[illegible]		Yes
14	9-18-23 20:45:15	9-18-23 20:45:21	[illegible]	[illegible]		Yes
15	9-18-23 22:56:30	9-18-23 22:56:36	[illegible]	[illegible]		Yes
16	9-18-23 22:57:24	9-18-23 22:57:32	[illegible]	[illegible]		Yes
17	9-18-23 23:12:20	9-18-23 23:12:23	[illegible]	[illegible]		Yes
18	9-19-23 0:27:23	9-19-23 0:27:26	[illegible]	[illegible]		Yes
19	9-21-23 14:01:06	9-21-23 14:01:38	[illegible]	[illegible]		Yes
20	9-21-23 14:08:30	9-21-23 14:08:41	[illegible]	[illegible]		Yes
21	9-21-23 14:14:51	9-21-23 14:15:00	[illegible]	[illegible]		Yes

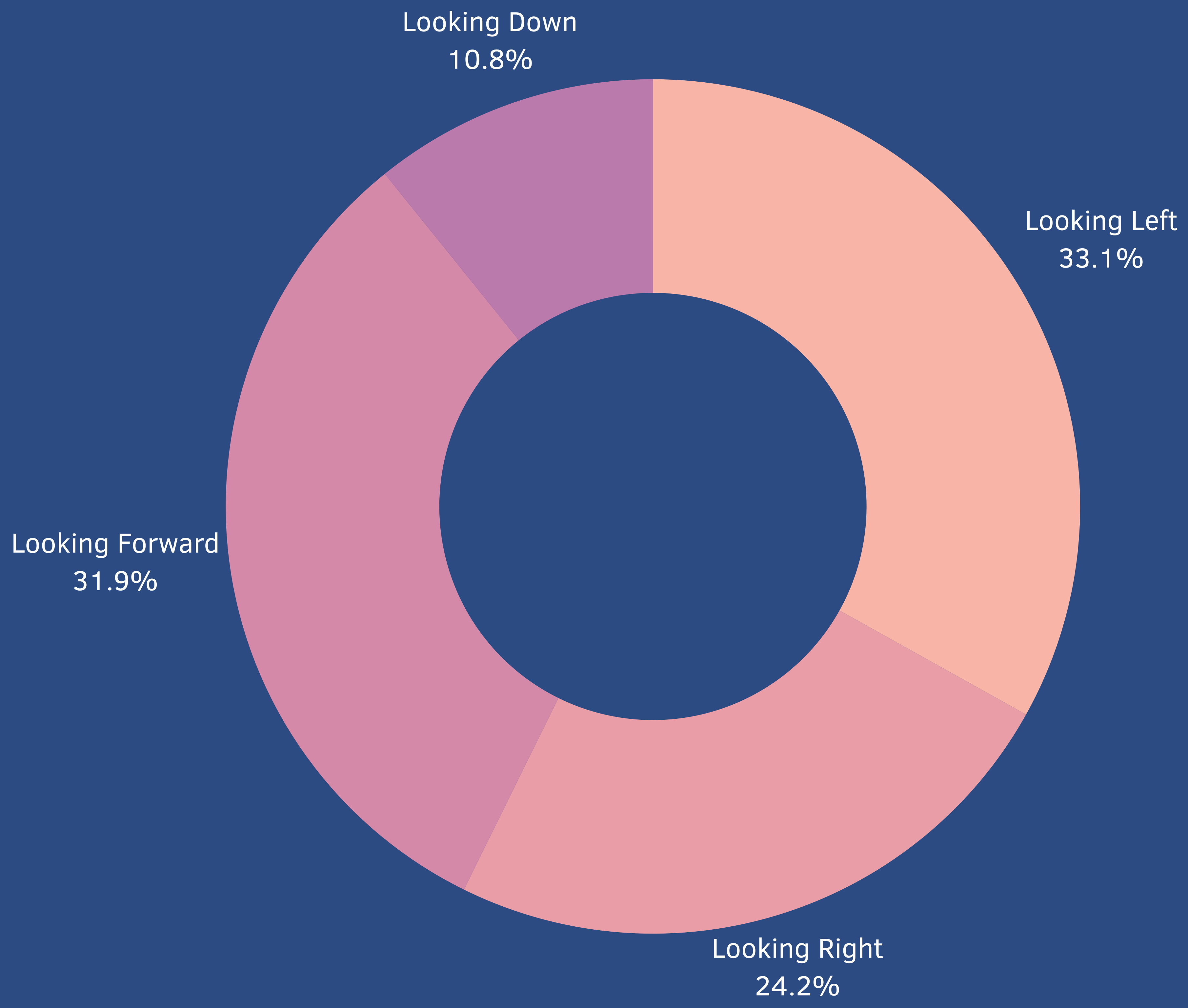
Data Collection

2. Data Collection Process: The data was collected from a classroom environment with approximately 10 participants on average. Data was collected in every MLPR lecture over a span of 8 weeks and we have 12,515 labelled images among 14,320 images.

3. Privacy and Ethics: Ethical concerns regarding privacy were addressed by ensuring that only the team members had access to the recorded data. Additionally, steps were taken to anonymize the data, such as not capturing individuals who did not provide consent or were not part of the study.

Nature of the dataset



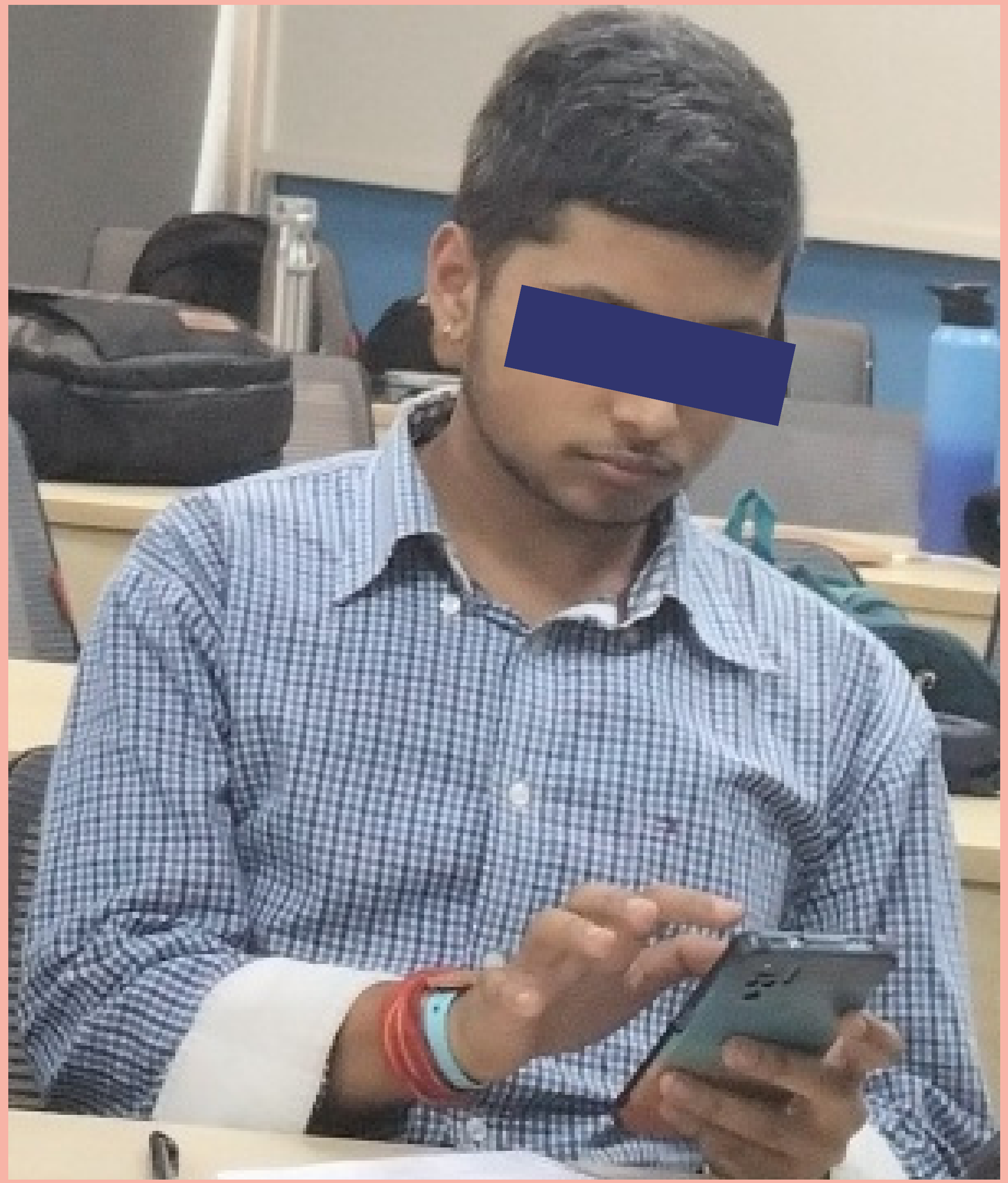


Simplified Guidelines for annotators

1	3	5
Looking Down	Head on hands, face on table	Hand Raise
Extreme Left, Right (Talking)	Looking slightly left	Looking Forward
Sleeping	Looking forward, but slouched or stretching	Completely upright posture
Looking at Laptop, Phone	Looking slightly right	Neutral Head Position

Sample Images

1 - Unattentive



3 - Moderately Attentive

5 - Fully Attentive

Feature
Pre Processing

Image Sampling

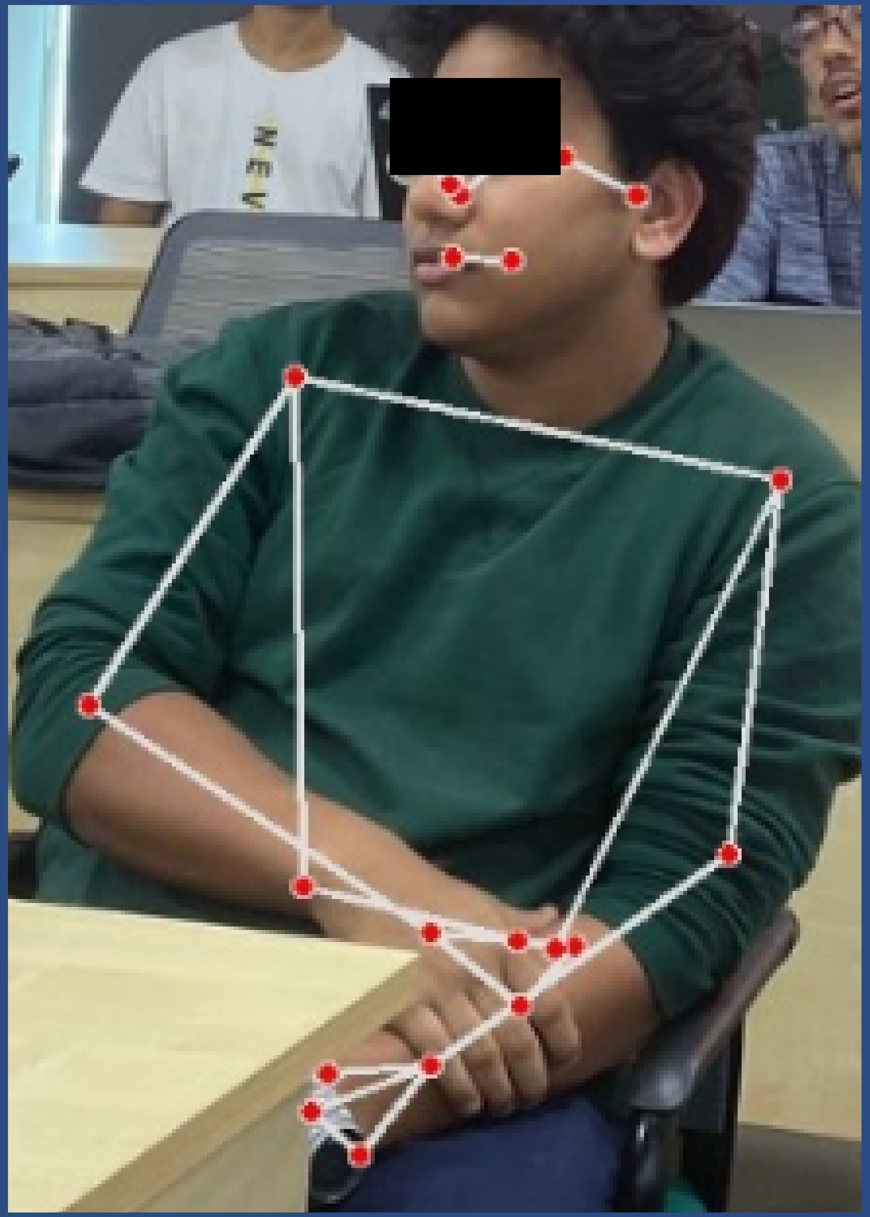
We realised that the attention level of a student does not change very frequently so we extracted images out of the recordings at every 5th second. This helps us to reduce the unnecessary processing and also capture the variation.

Data Cleaning and resizing the images

We filtered out images containing noise, which included instances of individuals passing in front of the camera, poor lighting conditions, and out-of-focus camera shots.

Landmark extraction

We are using Mediapipe to detect and track key landmarks on the human body in real-time from images or videos. Mediapipe Pose can return 33 key landmarks on a human body. MediaPipe Face Mesh iscan estimate 468 3D face landmarks in real-time.

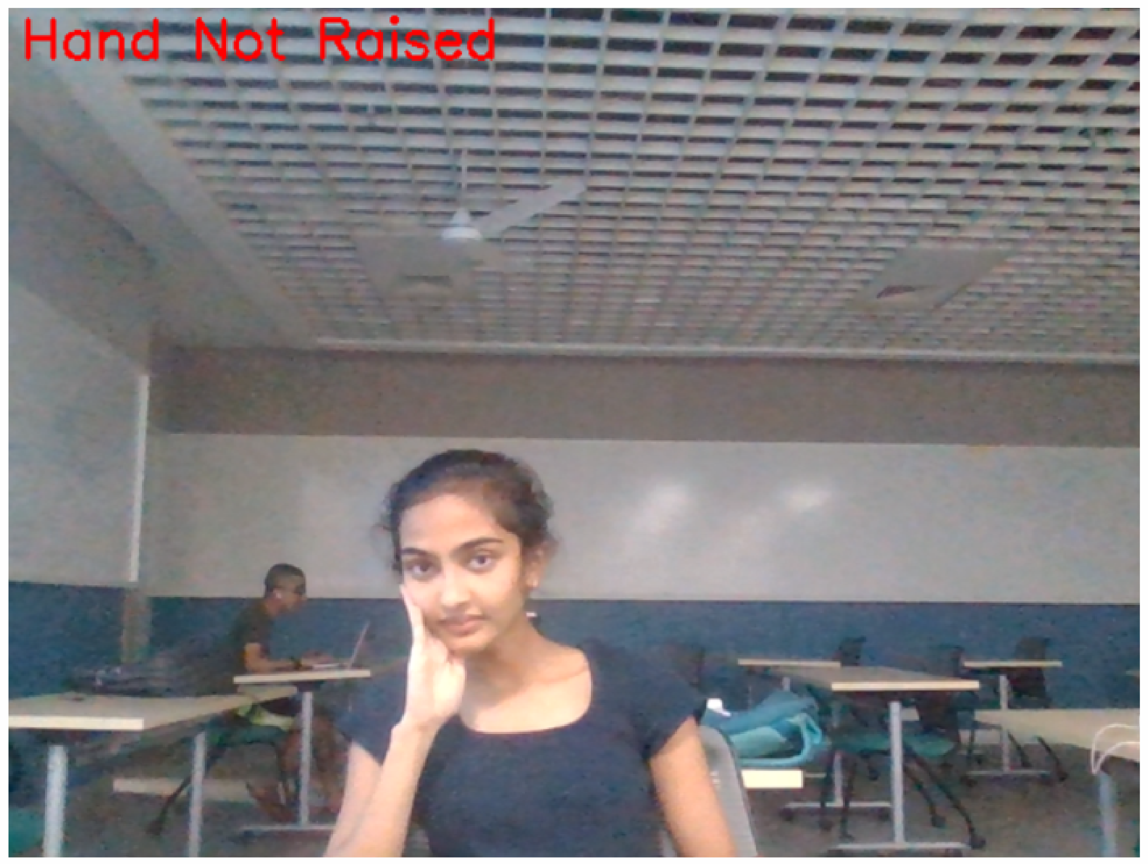


Person Detection

MediaPipe processes one person at a time. To work with multiple individuals, we'll incorporate an object detection model like Yolov8 to identify and crop a bounding box around each person for subsequent analysis using MediaPipe.

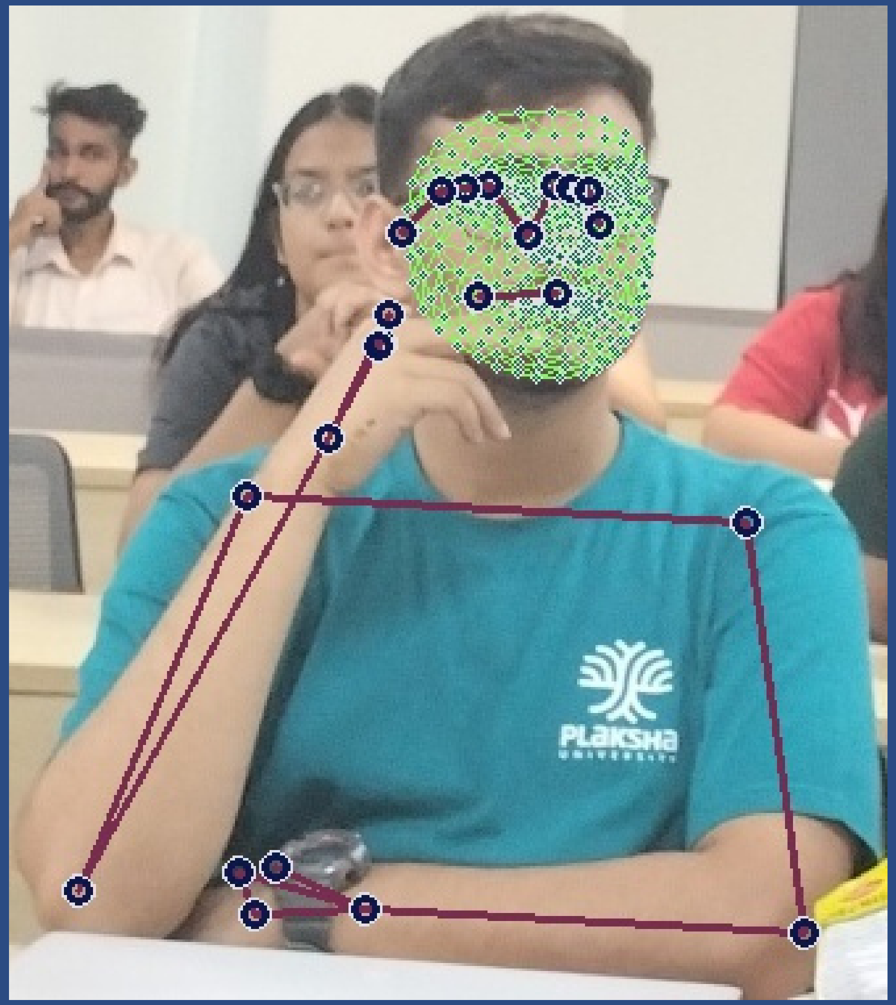
Feature Extraction - Preliminary approach used

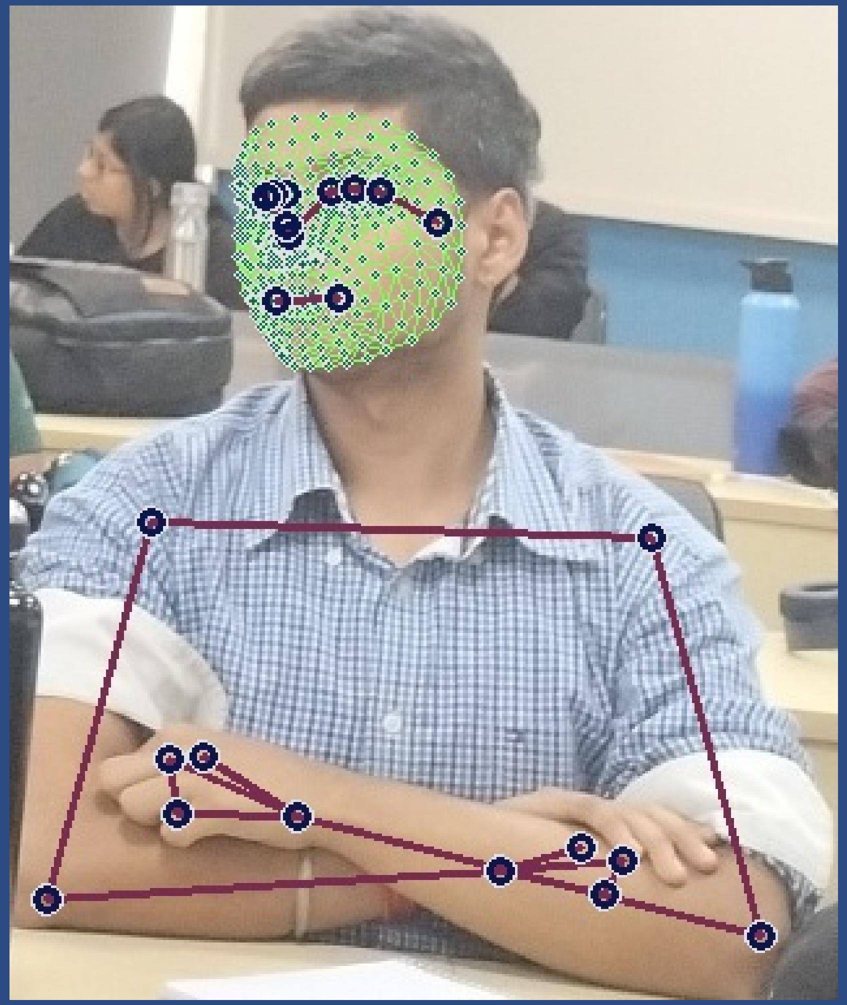
From these landmarks, we can make our own features like Hand raise, Looking down-up, Head on table, Looking towards the sides, Hand on head etc based on angles and coordinates.



Feature Extraction - Final Approach

- Extracts pose landmarks and calculates angles between combinations of three landmarks.
- Extracts facial landmarks to solve a Perspective-n-Point (PnP) problem, yielding head orientation in terms of rotation angles.



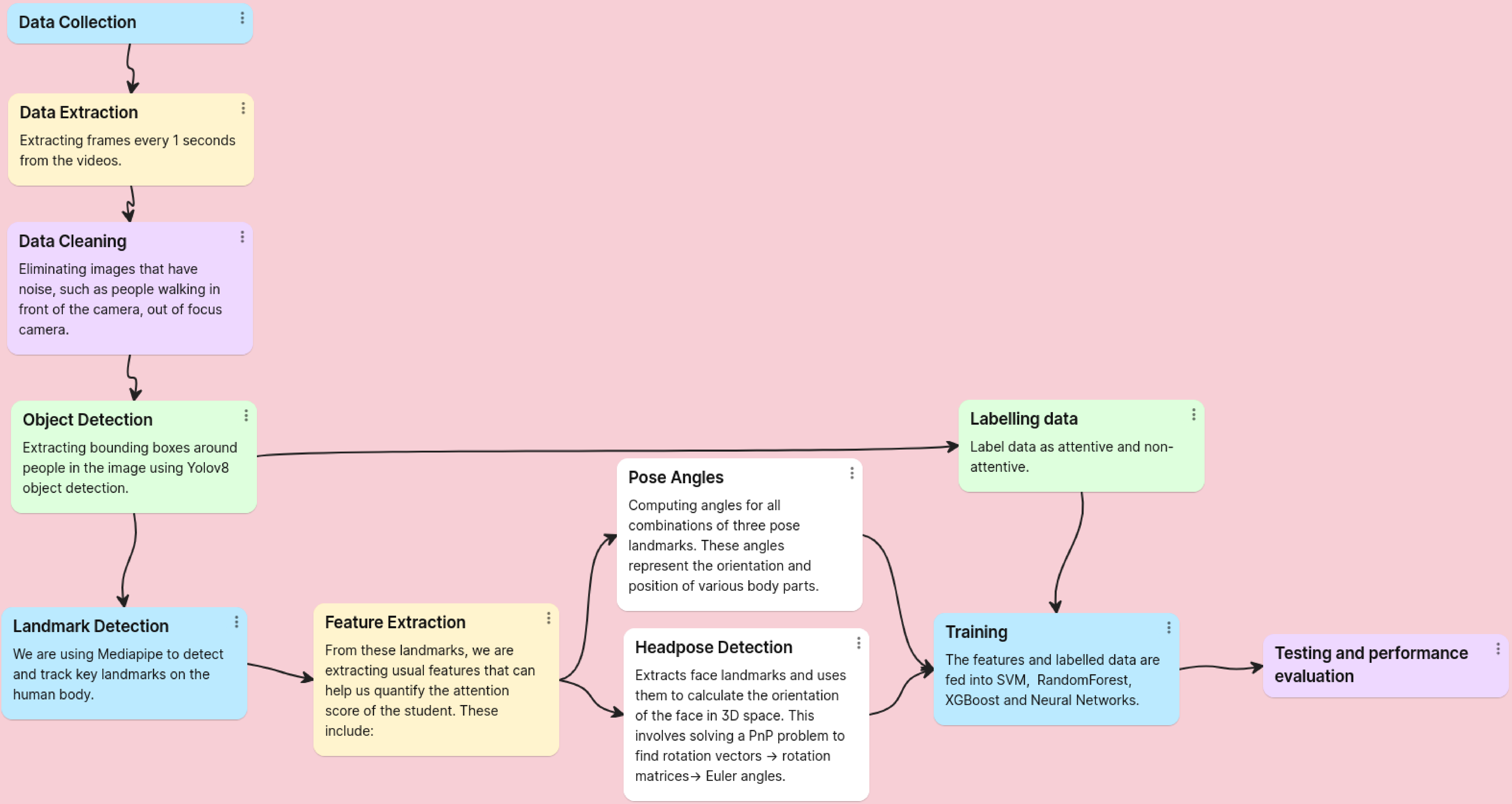


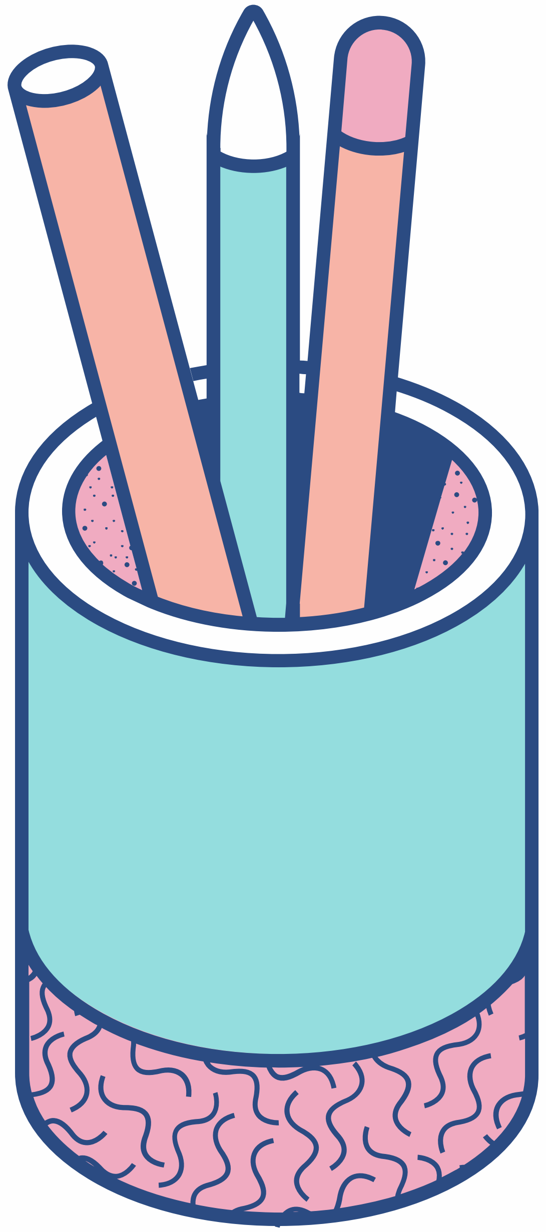
ML
Methodology

- YOLOv8
- MediaPipe

- Random Forest
- XGBoost
- SVM
- NN

Methodology

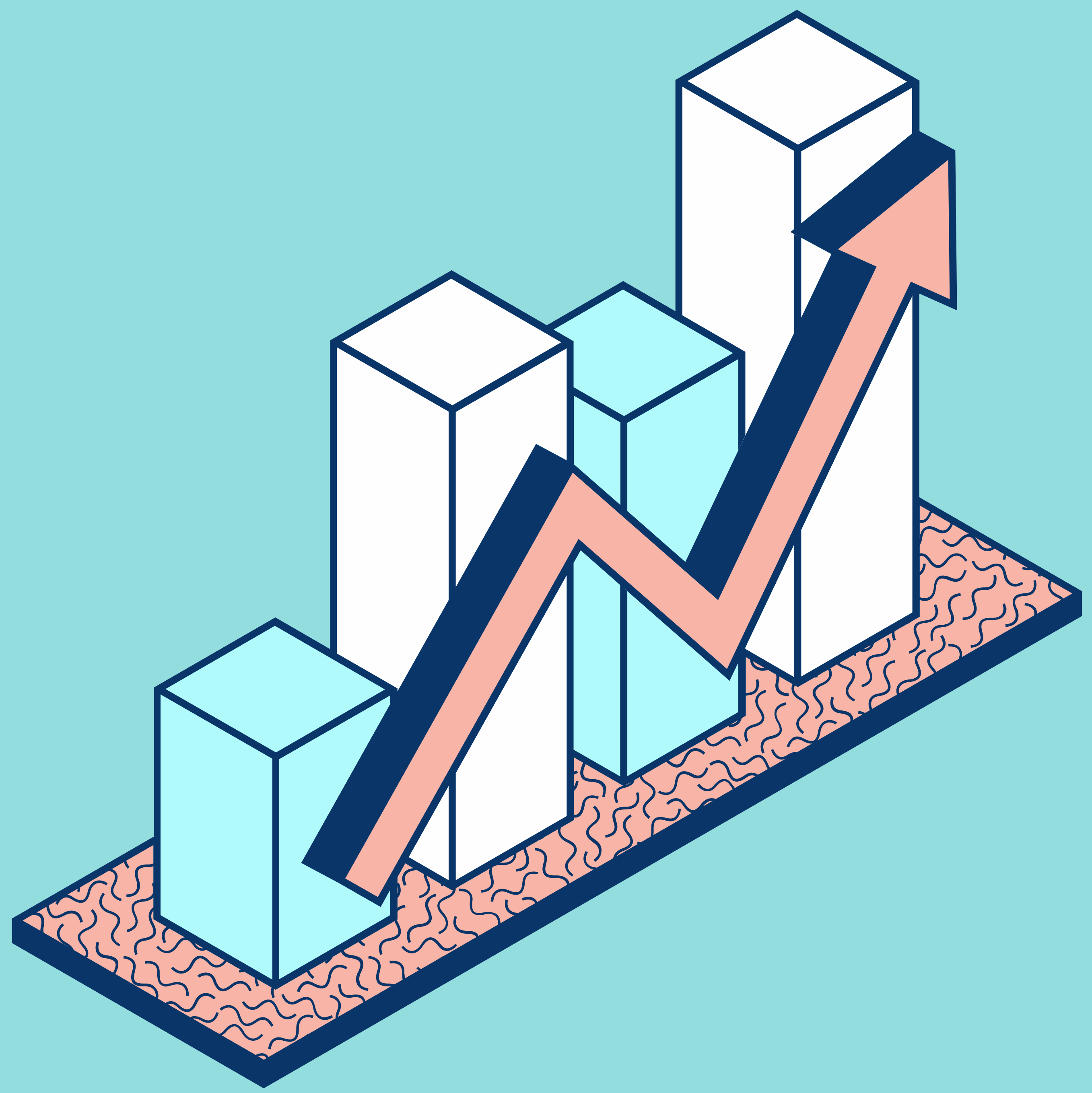




Challenges

- Data Labeling: Manually labeling data as "attentive" or "non-attentive" can be subjective and time-consuming.
- Variability in data: Extreme variability between student poses. Especially when the professor walks around class.
- Generalizing to different classrooms: Classroom settings can vary significantly. Models trained in one classroom may not generalize well to others.
- Gaze detection, EAR, YAR will be difficult to detect in individuals who are located at a considerable distance.

Performance Metrics



SVM

```
from sklearn.metrics import accuracy_score

svm_accuracy = accuracy_score(y_test, y_pred_svm)
print(f"SVM Test Accuracy: {svm_accuracy}")
```

```
SVM Test Accuracy: 0.661134163208852
```

NN

Model Evaluation

In [7]:
```
loss, accuracy = model.evaluate(X_test, y_test)
print(f"Test Accuracy: {accuracy}")
```

```
22/22 [==============================] - 0s 1ms/step - loss: 0.8399 - accuracy: 0.6638
Test Accuracy: 0.6637930870056152
```

RandomForest

```
rf_accuracy = accuracy_score(y_test, y_pred_rf)
print(f"Random Forest Test Accuracy: {rf_accuracy}")
```

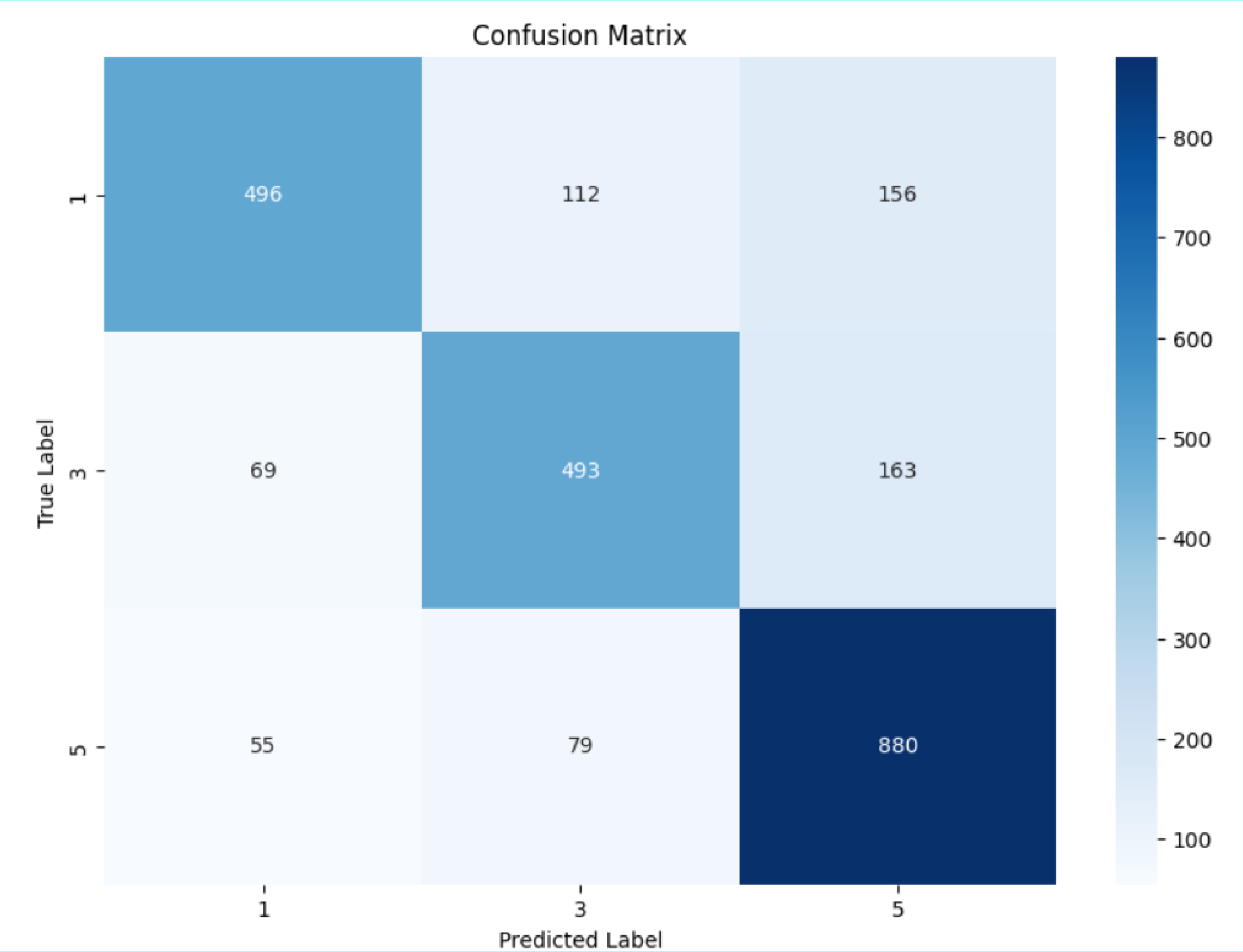
```
Random Forest Test Accuracy: 0.6874135546334716
```

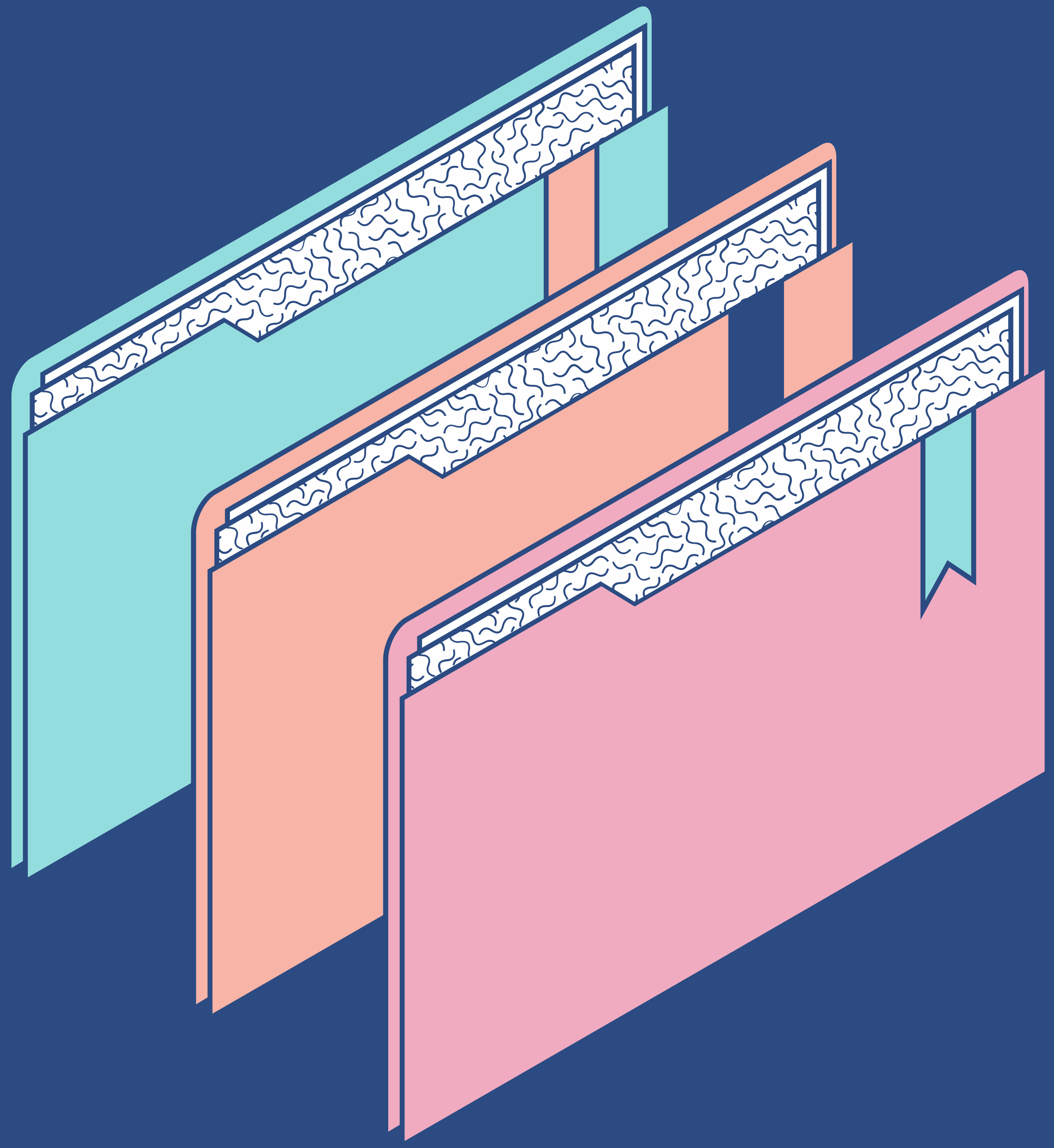
XGBOOST

Accuracy: 74.59%

Balanced Accuracy: 70.07%

	precision	recall	f1-score
0	0.82	0.62	0.71
1	0.69	0.58	0.63
2	0.74	0.89	0.81
accuracy			0.75
macro avg	0.75	0.70	0.72
weighted avg	0.75	0.75	0.74





Deployability of the ML solution

Deploying the Model at Plaksha

1 — **2** — **3** — **4**

STEP 1
Pilot Program

Initiate with a single course to assess the model's effectiveness.

STEP 2
Integration with Existing Infrastructure

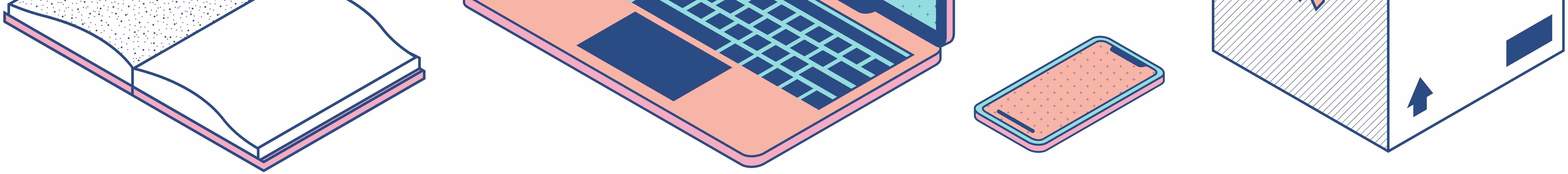
Utilize existing classroom cameras to minimize new hardware needs.

STEP 3
Generalizability

Broaden the training dataset to include various courses for better generalization.

STEP 4
Instructor Dashboard

Create a dashboard for teachers to view engagement analytics.



Challenges

- Privacy and Ethics: Navigate data privacy concerns and secure student consent.
- Behavioral Variability: Account for the diverse ways students express attention.
- Classroom Diversity: Adapt the model to different classroom environments.
- Camera Constraints: Overcome limitations due to varying camera quality and placements.
- Adaptability: Continuously update the model to maintain accuracy across different subjects
- System Integration: Seamlessly integrate with the existing IT framework of the university.

Do you have any questions?