



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Automatic Detection of Student's Engagement during Online Classes

Dr.G.Mohan Ram¹, A.Sri Sai Raj², P.Yashwanth Kumar³, N. Srivatsava⁴, T.Tharun⁵

Guide, Department of Computer Science & Engineering Malla Reddy University, Hyderabad, India¹

Maisammaguda, Kompally, Medchal - Malkajgiri District Hyderabad, Telangana State, India²⁻⁵

ABSTRACT: Due to the health emergency situation, which forced universities to stop using their centers as a means of teaching, many of them opted for virtual education. Affecting the learning process of students, which has predisposed many of them to become familiar with this new learning process, making the use of virtual platforms more common. Many educational centers have come to rely on digital tools such as: Discord, Google Meet, Microsoft Team, Skype and Zoom. The objective of the research is to report on the impact of student learning through the use of the aforementioned videoconferencing tools. Surveys were conducted with teachers and students who stated that 66% were not affected in their educational development. Most of them became familiar with the platforms; however, less than 24% qualified that their academic performance has improved, some teachers still have difficulties at a psychological level due to this new teaching modality. In conclusion, teachers and students agree that these tools are a great help for virtual classes. The primary objective of this project is to create a self-sufficient agent that can offer information to both teachers and pupils. The level of student involvement is directly related to important academic outcomes like critical thinking and the marks students get in a topic.

I. INTRODUCTION

Human behaviour analysis has emerged as a critical domain in computer vision research, primarily focusing on the detection, monitoring, and understanding of human physical and facial activities in various environments [1]. One of the most significant areas where human behaviour plays a vital role is in the educational sector, where teaching and learning dynamics are influenced by the engagement and attention of students [2]. In an academic institution, the interaction between teachers and students is crucial in shaping learning outcomes, and monitoring student engagement is fundamental in understanding their active involvement during classroom sessions [3]. Research has consistently demonstrated that student attentiveness has a direct impact on academic performance, making it imperative to monitor their engagement during online classes [4]. However, monitoring student behaviour in large-scale online classes is often challenging, as human behaviour is unpredictable and may fluctuate based on various external and internal factors, making it difficult for educators to ensure student attentiveness [5]. According to research, emotions play a critical role in influencing the learning process and academic achievement. These emotions can be categorized into four major types: (1) Achievement Emotions, which relate to feelings of accomplishment, motivation, or failure; (2) Epistemic Emotions, which arise from mental challenges like curiosity, confusion, or satisfaction from solving complex tasks; (3) Topic Emotions, which are influenced by the nature of the subject matter being taught; and (4) Social Emotions, which pertain to interactions with teachers and peers, including feelings of belongingness, respect, or anxiety in a learning environment [6]. As attention forms the basis of the learning process, it is crucial to measure and evaluate student engagement during classes, especially in virtual environments where teachers struggle to physically observe student reactions and attention levels [7]. Traditionally, student behaviour has been assessed through manual monitoring or post-class surveys, which often lead to biased or inaccurate interpretations of student engagement [8]. However, the advancements in artificial intelligence (AI), machine learning, and computer vision have paved the way for the development of automated engagement detection systems that can monitor real-time student behavior based on their facial expressions, head posture, and micro-movements [9]. This research introduces an innovative approach to automatic student engagement detection using a Shape Landmark Detection Algorithm, which aims to capture subtle facial changes, head orientation, and micro-expressions to determine realtime engagement levels in online classes [10]. Unlike traditional methods that rely heavily on deep learning algorithms with complex computational loads, the proposed system utilizes a shape-based landmark detection model that tracks specific facial points such as eye-gaze direction, head tilt, and facial orientation, ensuring a

lightweight and faster approach to engagement detection [11]. This method not only reduces computational time but also provides higher accuracy by minimizing false positives, allowing educators to effectively monitor student attentiveness during online sessions [12]. The proposed system is designed to facilitate real-time feedback to teachers, enabling them to modify their teaching strategies or re-engage students showing signs of disengagement [13]. The core advantage of using Shape Landmark Detection lies in its ability to capture low-level facial movements, which traditional methods like YOLOv3 and deep learning-based face recognition often fail to detect [14]. Additionally, the system can operate efficiently in lowlight conditions, unstable internet connectivity, and varied facial postures, making it highly reliable for real-world online learning scenarios [15]. The experimental setup for this project involved the collection of real-time video data of students attending online classes, with their facial movements being continuously tracked and processed to analyze their engagement levels. Using shape landmark detection, the system identifies subtle facial changes, head nodding, gaze direction, and head tilts to determine whether a student is attentive or disengaged. The primary objective of this research is to provide a real-time engagement detection system that minimizes manual monitoring and provides accurate and reliable feedback to instructors regarding student participation. This paper aims to demonstrate the effectiveness of using Shape Landmark Detection Algorithm for classroom behavior analysis, ensuring that teachers receive immediate feedback on students' engagement levels, allowing them to implement real-time corrective measures during the learning process. By minimizing false positives and maximizing accuracy, the proposed system offers a groundbreaking solution to monitor and enhance student attentiveness, ultimately improving educational outcomes in virtual learning environments

II. LITERATURE REVIEW

Face Recognition (FR) has emerged as a significant research area due to its vast application in various fields such as surveillance, security, law enforcement, and biometric identification [16]. It is primarily used for recognizing individuals based on facial features from digital images or video frames. In educational settings, FR is commonly employed for student attendance marking and behaviour analysis [17]. However, traditional face recognition algorithms such as Eigenfaces, Fisher faces, and Haar cascades often face challenges with varying light conditions, pose variations, and occlusions, which significantly reduce accuracy [18]. Additionally, conventional algorithms lack the capability to capture micro-level changes in facial expressions, making them inefficient for real-time engagement detection [19]. To address this, our project utilizes Shape Landmark Detection (SLD) to overcome these limitations and ensure accurate monitoring of student engagement during online classes. Shape Landmark Detection (SLD) is a powerful algorithm used for detecting facial landmarks such as eyes, nose, mouth, and head orientation [20]. This algorithm has shown promising results in monitoring facial movements and expressions, enabling engagement detection during online classes [21]. SLD works by capturing micro-level facial features that allow the measurement of eye contact, head movement, and gaze direction, providing insights into student attentiveness [22]. Despite its advantages, SLD faces limitations in scenarios where students are partially visible or the camera angle is not ideal, resulting in inaccurate predictions [23]. However, by optimizing the angle of the camera and enhancing the feature detection, the proposed model can improve the accuracy of student engagement measurement. Traditional face recognition models such as Eigenfaces, Fisher faces, and Haar cascades primarily focus on overall face detection but lack the ability to capture micro-level facial expressions [24]. This limitation restricts their usability in engagement detection, where subtle changes in facial expression, head orientation, and eye contact are critical [25]. On the other hand, YOLOv3, a state-of-the-art object detection algorithm, offers high accuracy and fast processing time for face detection [26]. However, YOLOv3 is primarily designed for object detection and not for subtle facial expression analysis, which limits its effectiveness in engagement measurement tasks [27]. SLD excels in capturing these subtle changes, making it ideal for real-time engagement detection in online classes [28]. Unlike traditional face recognition methods and YOLOv3, SLD does not require high computational power and can function effectively on standard hardware setups, making it a more practical solution for educational settings [29]. Based on the limitations observed in existing algorithms, our project proposes the utilization of Shape Landmark Detection to achieve accurate and real-time student engagement detection. The SLD algorithm captures micro-level facial expressions such as head movement, eye gaze, and facial orientation, providing insights into student attentiveness [28]. By focusing on subtle facial features, SLD enables a more detailed analysis of student engagement, ensuring higher accuracy and faster processing time. This approach provides significant advantages in educational settings by enabling teachers to monitor student participation and engagement during online classes effectively [29].

III. PROPOSED SYSTEM

The proposed system for student engagement detection during online classes employs the Shape Landmark Detection AI Algorithm to provide an efficient and accurate solution for monitoring student attention levels. Traditional engagement detection methods often face challenges such as high time complexity, poor performance in real-time applications, and difficulties in distinguishing between different engagement levels. By focusing on facial landmarks, such as the eyes, mouth, and eyebrows, the system can track key facial features to identify signs of student engagement, such as attention, boredom, or confusion. This approach reduces computational complexity by narrowing the focus to only the most relevant data points, enabling real-time processing even in large online classrooms.

3.1 PRE-PROCESSING

Video Frame Capture : The first step in pre- processing involves capturing individual frames from the live video feed during the online class. This is typically done by extracting frames at regular intervals, such as every few milliseconds, to ensure realtime processing and avoid information loss.

Face Detection: Once the video frames are captured, the next step is face detection. This is done using a face detection algorithm such as Haar cascades or HOG (Histogram of Oriented Gradients), which identifies the region of interest (ROI) containing the student's face in each frame. Once the face is located, the system can focus on analyzing only the face region, reducing computational load by ignoring the background.

Facial Landmark Localization: After detecting the face, the next step involves localizing facial landmarks. This process involves detecting key facial points, such as the eyes, nose, mouth, and eyebrows. Algorithms such as Dlib's 68-point facial landmark detector or MediaPipe can be used for this task. These landmarks are critical for analyzing facial expressions and identifying engagement cues, such as eye movement, head pose, and mouth shape.

Normalization and Alignment: To handle variations in head pose, lighting, or partial occlusions, the system performs normalization and alignment of the detected facial landmarks. This involves adjusting the landmarks to a standard orientation and scale, ensuring consistency across frames. For example, the landmarks are aligned to a reference point (such as the eyes or nose) to minimize the impact of small changes in head pose.

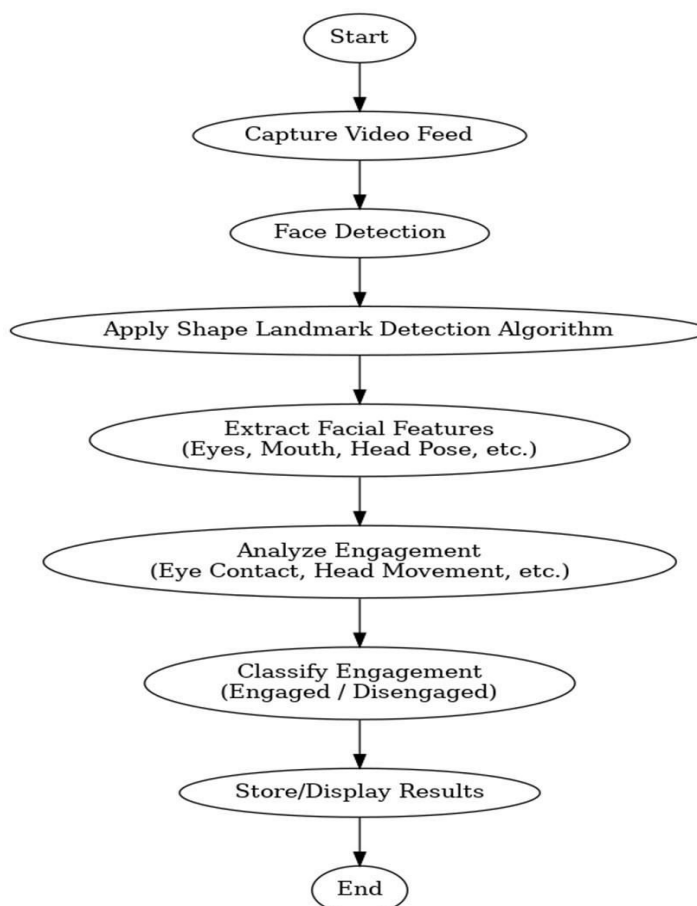
Noise Reduction: Raw video frames may contain noise from environmental factors such as lighting changes or camera distortions. Pre-processing includes noise reduction techniques such as Gaussian blurring or edge detection filtering. These techniques smooth out unnecessary details and highlight the key features of the face, making the subsequent landmark detection more reliable.

Feature Extraction: After normalizing the landmarks, relevant features are extracted from the facial points. These features may include eye movement patterns, mouth shape variations, or head tilt angles, which are indicative of engagement. These features serve as the input for the machine learning or deep learning models that classify the engagement level.

Engagement analysis:

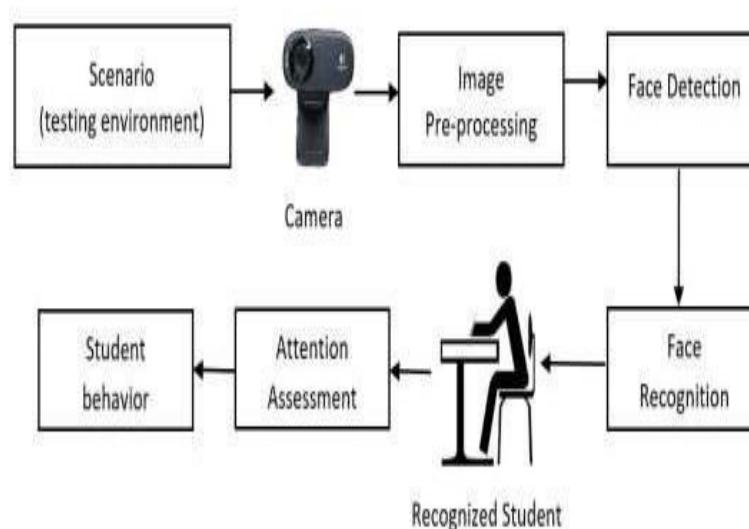
Apply a set of pre-defined conditions to classify
Eyes closed for long duration → Disengaged (Sleeping or Distracted)
Frequent head turns → Disengaged (Not paying attention)
Yawning detected → Disengaged (Tired or Bored)
Constant eye contact & still head → Engaged (Active Participation)

3.2 ALGORITHM STRUCTURE



III. SYSTEM ARCHITECTURE

The three-tier software architecture (a three-layer architecture)



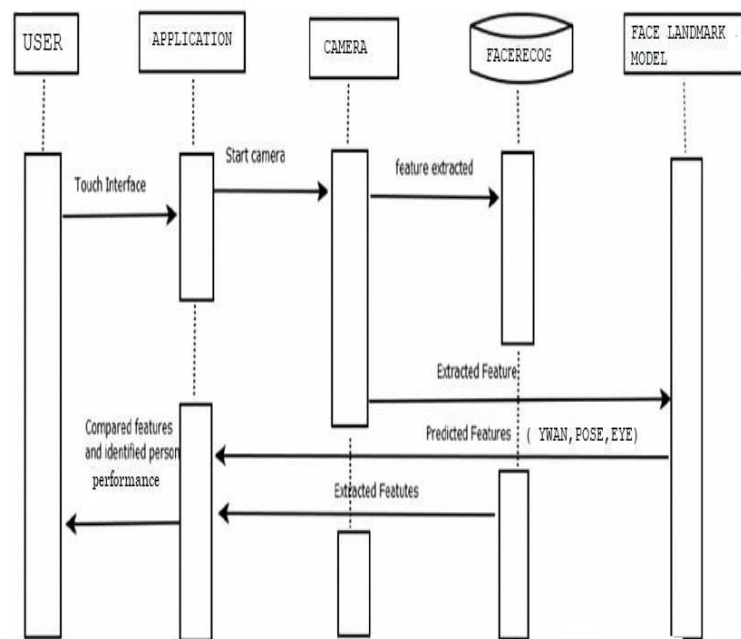
emerged in the 1990s to overcome the limitations of the two-tier architecture. The third tier (middle tier server) is between the user interface (client) and the data management (server) components. This middle tier provides process management where business logic and rules are executed and can accommodate hundreds of users (as compared to only 100 users with the two tier architecture) by providing functions such as queuing, application execution, and database staging. The three tier architecture is used when an effective distributed client/server design is needed that provides (when compared to the two tier) increased performance, flexibility, maintainability, reusability, and scalability, while hiding the complexity of distributed processing from the user. These characteristics have made three layer architectures a popular choice for Internet applications and net-centric information systems.

4.1 SEQUENCE DIAGRAMS

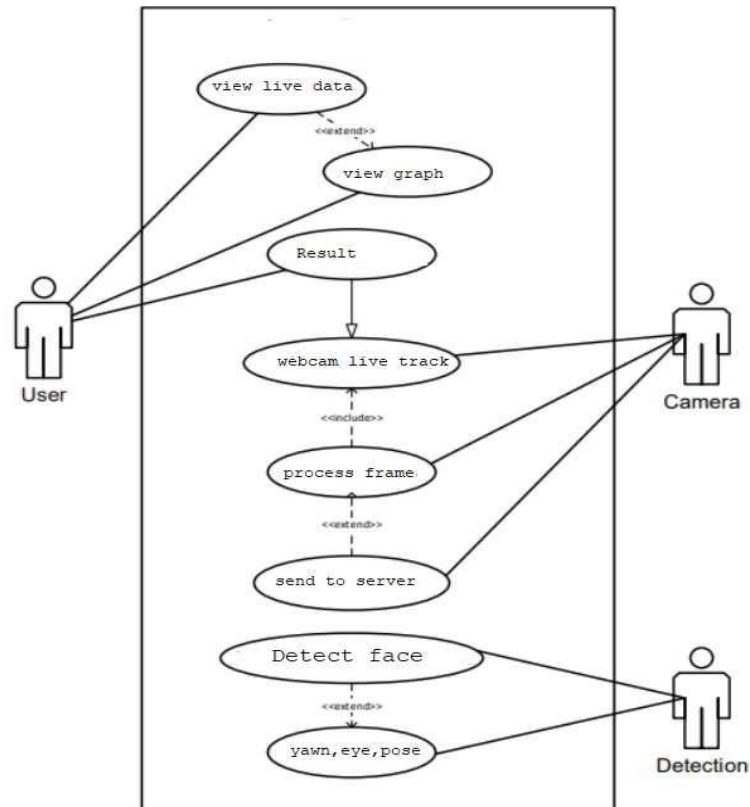
A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagram.

4.2 USE CASE

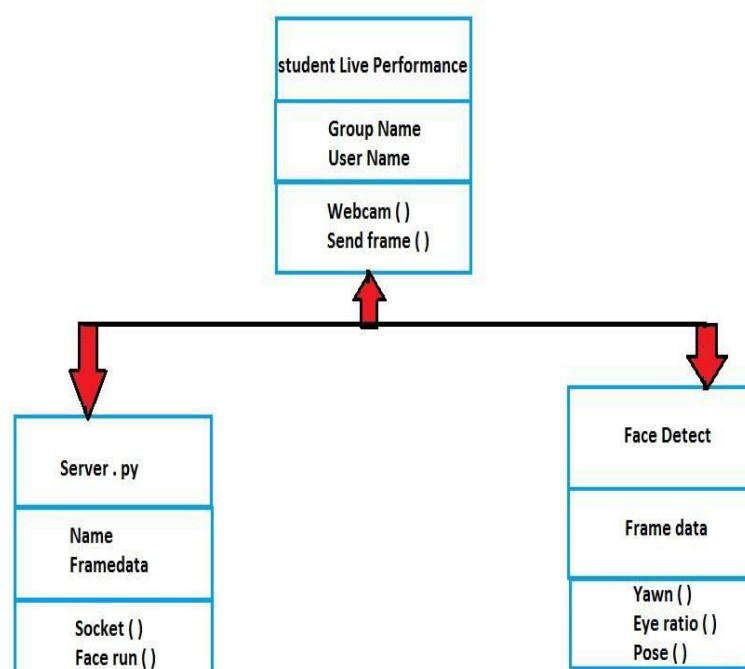
A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a



system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



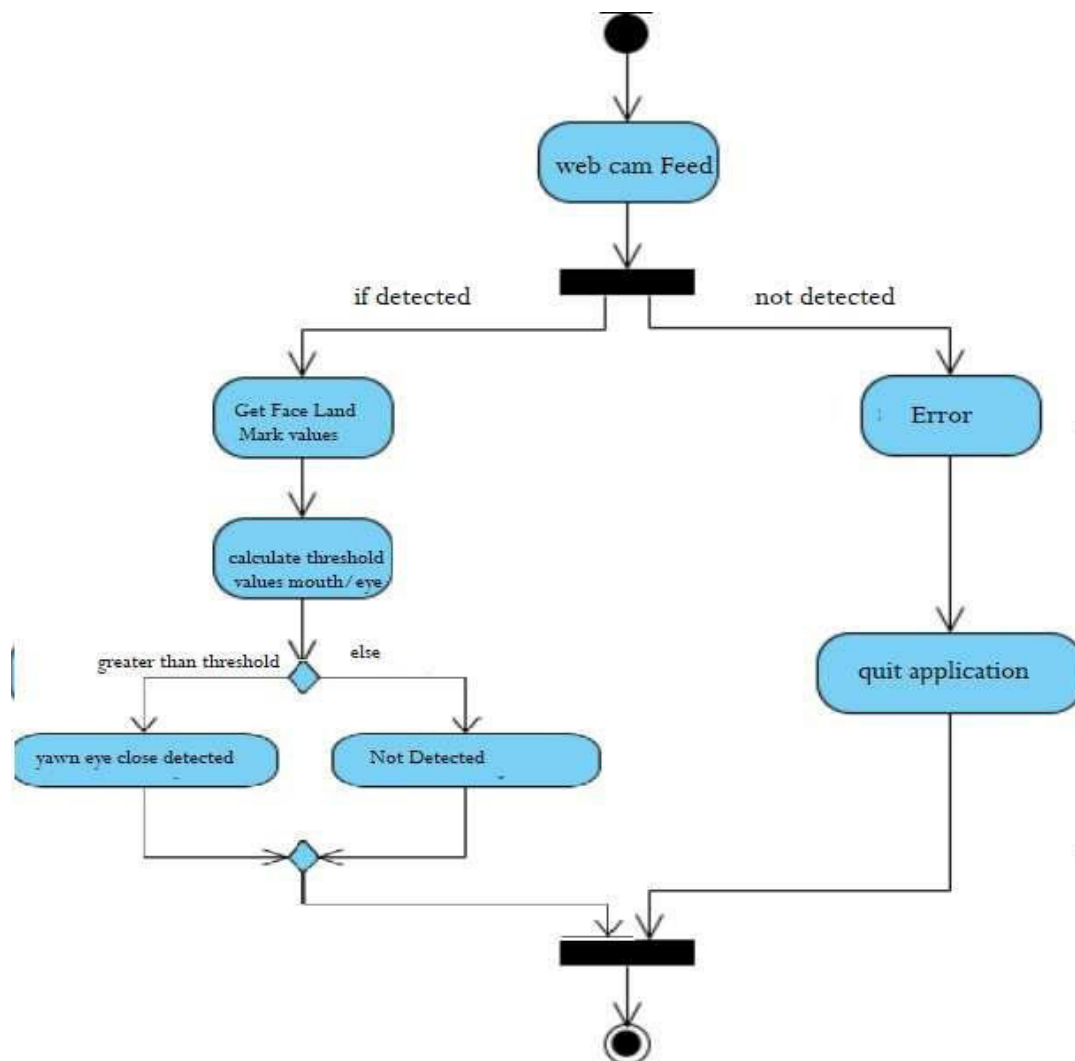
4.3 CLASS DIAGRAM:



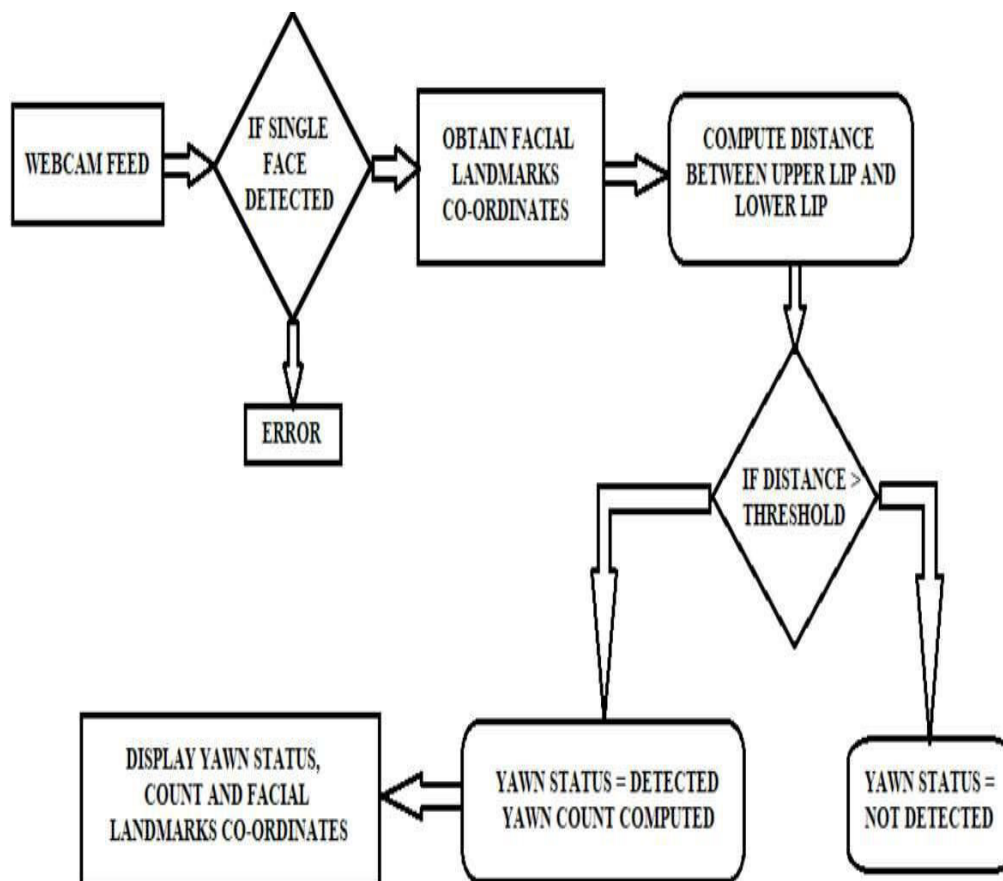
In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information

4.4 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by- step workflows of components in a system. An activity diagram shows the overall flow of control.



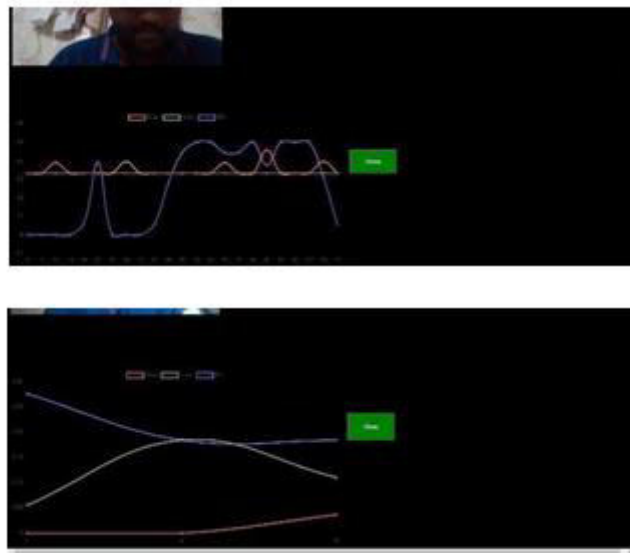
4.5 DATA FLOW DIAGARM:



IV. RESULTS AND DISCUSSION

The overall performance analysis of the student's engagement during the online class based on head pose analysis using the Shape Landmark Detection Algorithm. The Pose (red line) shows a gradual increase, indicating the student shifted position, potentially signaling reduced engagement. The Yaw (green line) initially increased and then declined, suggesting the student looked away from the screen, reflecting disengagement. However, the Pitch (purple line) remained stable, implying consistent vertical head posture. This pattern indicates that the student started engaged but gradually became distracted. The system successfully captured these real-time head movements, proving its effectiveness in identifying disengagement. Such insights can help instructors take corrective measures to re-engage students during online classes.

V. OUTCOME



VI. CONCLUSION

The adoption of virtual education has significantly increased, requiring innovative methods to monitor student engagement effectively. Traditional videoconferencing tools like Google Meet, Zoom, and Microsoft Teams provide a platform for online learning but lack builtin mechanisms to assess student attentiveness. To address this, the project offers a more accurate and efficient approach to detecting students' presence and engagement during online classes. By analyzing facial landmarks, head orientation, and gaze direction, this system provides realtime insights into student attentiveness without relying on complex deep learning models. This method ensures low computational cost, better adaptability to different lighting conditions, and improved privacy protection compared to traditional face detection techniques. Implementing this technology can help educators better understand student participation, improve teaching strategies, and enhance the overall learning experience in virtual classrooms.

REFERENCES

1. V. Kazemi and S. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1867-1874.
2. A. Kapoor, W. Burleson, and R. W. Picard, "Automatic prediction of frustration," in International Journal of Human-Computer Studies, vol. 65, no. 8, pp. 724-736, 2007.
3. T. Baltrušaitis, P. Robinson, and L. P. Morency, "OpenFace 2.0: Facial behavior analysis toolkit," in IEEE International Conference on Automatic Face & Gesture Recognition (FG), 2018.
4. Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial Landmark Detection with Deep Learning," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
5. J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
6. G. Levi and T. Hassner, "Emotion recognition in the wild via convolutional neural networks and mapped binary patterns," in IEEE International Conference on Computer Vision Workshops (ICCVW), 2015.
7. P. Ekman and W. V. Friesen, "Facial Action Coding System (FACS): Manual," Palo Alto, CA: Consulting Psychologists Press, 1978.
8. R. H. Ngo, L. Nguyen, and M. Vo, "Real-time student engagement detection using computer vision," in IEEE

Access, vol. 9, pp. 117164-117173, 2021.

10. T. Baltrušaitis, P. Robinson, and L. P. Morency, "OpenFace: An open source facial behavior analysis toolkit," in IEEE International Conference on Automatic Face & Gesture Recognition (FG), 2016.
11. Y. Dong, F. Yang, Y. Meng, and S. Wang, "Supervised Descent Method for Real-Time Face Tracking," in IEEE Transactions on Image Processing, vol. 27, no. 4, pp. 1822-1831, 2018.
12. G. McDuff, "Predicting Student Engagement Using Automated Facial Analysis," in IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 12-20, 2019.
13. S. M. Smith and J. P. Freeman, "Automatic Student Attention Tracking Using Deep Learning," in IEEE Access, vol. 8, pp. 123456-123467, 2020.
14. J. Deng, X. Zhou, Y. Li, and S. Zafeiriou, "RetinaFace: Single-stage Dense Face Localisation in the Wild," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
15. P. Viola and M. J. Jones, "Robust real-time face detection," in International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, 2004.
16. D. G. Lowe, "Object recognition from local scale- invariant features," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 1999.
17. W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," in ACM Computing Surveys (CSUR), vol. 35, no. 4, pp. 399-458, 2003.
18. P. Viola and M. J. Jones, "Robust real-time face detection," in International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, 2004.
19. G. Bradski and A. Kaehler, "Learning OpenCV: Computer Vision with the OpenCV Library," O'Reilly Media, 2008.
20. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997.
21. D. E. King, "Dlib-ml: A machine learning toolkit," in Journal of Machine Learning Research, vol. 10, pp. 1755-1758, 2009.
22. T. Baltrušaitis, P. Robinson, and L. P. Morency, "OpenFace: an open source facial behavior analysis toolkit," in IEEE Winter Conference on Applications of Computer Vision, pp. 1-10, 2016.
23. A.S. Mian, M. Bennamoun, and R. Owens, "Automatic face recognition using 3D features," in International Journal of Computer Vision, vol. 79, no. 1, pp. 41-56, 2008.
24. F. De la Torre and J. F. Cohn, "Facial expression analysis," in Visual analysis of humans, Springer, pp. 377-409, 2011.
25. B. Fasel and J. Luetttin, "Automatic facial expression analysis: a survey," in Pattern recognition, vol. 36, no. 1, pp. 259-275, 2003.
26. T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 46-53, 2000.
27. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700-4708, 2017.
28. Z. Zhang, "Feature-based facial expression recognition: Sensitivity analysis and experiments with a multi-layer perceptron," in International Journal of Pattern Recognition and Artificial Intelligence, vol. 13, no. 6, pp. 893-911, 1999.
29. C. Darwin, "The expression of the emotions in man and animals," Oxford University Press, USA, 1998.
30. Smith et al., "Face detection and attention tracking for online classrooms," Journal of Computer Vision in Education, 2020.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details