International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 01 | Jan 2025 www.irjet.net p-ISSN: 2395-0072

ENHANCING ONLINE EDUCATION THROUGH REAL TIME ATTENTION ASSESSMENT

Ketha Krishna Durga Devi[1], Gajulapally Aishwarya Reddy[2], Ms. P. Aswani[3]

[1]B-Tech 4th year, Dept. of CSE (DS), Institute of Aeronautical Engineering
[2] B-Tech 4th year, Dept. of CSE (DS), Institute of Aeronautical Engineering
[3] Assistant Professor, Dept. of CSE (DS), Institute of Aeronautical Engineering, Telangana, India

---

Abstract - *This project introduces a comprehensive system to monitor and assess student engagement during online learning sessions using computer vision and audio processing techniques. The framework captures live video streams via a webcam and leverages MediaPipe's Face Mesh to identify key facial landmarks for real-time head pose and estimation. The gaze tracking determines the direction of the student's attention, distinguishing between engaged ("Forward") and distracted ("Left," "Right," or "Down") states. Simultaneously, the system incorporates audio analysis through the PyAudio library to detect prolonged noise, marking periods of potential disengagement. By integrating these visual and auditory cues, the model provides a robust evaluation of student attention, enhancing the understanding of engagement in virtual learning environments. This system is designed to run efficiently on devices with limited computational power, making it suitable for broader adoption in online education platforms.*

Key Words: Student Engagement, Computer Vision, Gaze Estimation, Head Pose Detection, MediaPipe Face Mesh, Audio Processing, PyAudio, Real-Time Monitoring, Low Computational Power

1. INTRODUCTION

The shift towards online education has brought significant opportunities and challenges to the educational landscape.

While digital platforms enable greater accessibility and flexibility in learning, they also present unique hurdles in maintaining student engagement and ensuring effective instructional delivery. One of the critical challenges in online education is the ability to accurately assess and sustain student attention, a key determinant of learning success.

Traditional methods of attention assessment, such as quizzes and participation tracking, fall short in virtual environments where direct, in-person observation is not feasible. This gap necessitates the development of innovative solutions that can provide real-time insights into student engagement levels.

Recent advancements in artificial intelligence and computer vision offer promising avenues to address this challenge. This report explores the application of real-time attention assessment in online education using the MediaPipe framework, a cutting-edge computer vision solution designed for efficient and accurate tracking of facial landmarks and head pose estimation. MediaPipe's ability to perform real-time facial recognition and gaze tracking makes it an ideal choice for continuously monitoring student attention during online classes.

The goal of this project is to use the MediaPipe architecture to create a real-time attention assessment system in response to these difficulties. During online classes, students' facial cues, such as head position and gaze direction, can be continuously monitored thanks to MediaPipe, an innovative computer vision tool that provides effective facial landmark detection.

The suggested system can identify minute variations in facial expressions and movements that may be indicators of a student's concentration by utilizing MediaPipe's face mesh solution. Compared to more conventional techniques like tests or attendance records, these cues offer a more dynamic and real-time way to monitoring attentiveness and, when examined, reveal insightful information about student participation.

This study aims to do more than merely measure attentiveness. This method is intended to assess how well it detects attention lapses, or instances in which students are unfocused or disengaged, and gives teachers timely, useful feedback. Through the integration of this technology into online education platforms, educators can intervene promptly and modify their pedagogical approaches to re-engage learners.

Additionally, the project intends to show how this technology can be used more broadly to improve learning outcomes by promoting an interactive and adaptable learning environment. The ultimate objective is to improve student engagement by providing real-time feedback and providing a more individualized and adaptable virtual learning environment.

The suggested approach has the potential to enhance engagement and provide long-term learning analytics. Through the process of gathering data on student attentiveness throughout online sessions, the system is able

to produce comprehensive reports that illustrate trends in student involvement over time. Teachers and educational institutions alike can benefit greatly from these findings, which provide a deeper knowledge of the elements—like the time of day, lesson design, or teaching style—that affect students' ability to focus.

Furthermore, by identifying students who might require extra support or modifications to their learning environment, this data can be leveraged to create tailored learning experiences. With time, a more data-driven approach to online education may be possible thanks to this attention tracking technology provided by MediaPipe, which enables schools to make wise decisions to improve student performance and retention.

2. LITERATURE SURVEY

Thaman, Cao, and Caporusso In their paper "Face

Mask Detection using MediaPipe Facemesh", Thaman, Cao, and Caporusso introduce a system for detecting face masks using the MediaPipe framework, focusing on compatibility across platforms, particularly mobile devices. The system utilizes MediaPipe's facial landmark detection technology to determine whether individuals are wearing face masks. The model was evaluated in real-world environments, demonstrating its effectiveness in mask detection, which was especially critical during the COVID-19 pandemic. [1]

Wu, Zhang, and Tian In "Simultaneous Face Detection and Pose Estimation Using Convolutional Neural Network Cascade", Wu, Zhang, and Tian propose a multi-task convolutional neural network (CNN) cascade for simultaneous face detection and head pose estimation. The model benefits from multi-task learning and feature fusion, improving accuracy and real-time performance. The system was validated using datasets like FDDB and AFW. However, while it offers runtime efficiency, challenges such as scalability, reliance on high-performance GPUs, and difficulties in generalization across varied datasets remain. [2]

Chen, Zhang, Yin, and Wang The study "TRFH: Towards Real-Time Face Detection and Head Pose Estimation" by Chen, Zhang, Yin, and Wang aims to create a multi-task model combining face detection and head pose estimation into a unified system, eliminating the need for separate face detection processes. The efficiency of the model is tested with multi-stage face attribute analysis on the AFLW dataset. Despite achieving improvements in real-time estimation, especially for single-person scenarios, the model faces limitations in precision and computational complexity, particularly for multi-person settings. [3]

Yuan in "Face Detection and Recognition Based on

Visual Attention Mechanism Guidance Model in Unrestricted Posture", Yuan explores enhancing facial occlusion handling in face detection and recognition by integrating a visual attention mechanism. The model is designed to simplify detection into a semantic feature extraction task, aiming for high accuracy while maintaining speed, suitable for security surveillance applications. However, the reliance on simulation-based evaluations rather than real-world testing could limit the model's practical effectiveness, especially in complex environments. [4]

Wang, Lei, and Qian The research "Siamese PointNet: 3D Head Pose Estimation with Local Feature Descriptor" by Wang, Lei, and Qian focuses on creating a Siamese network for robust 3D head pose estimation, especially under challenging conditions such as occlusions and varying viewpoints. The model was evaluated using public datasets to minimize error margins while ensuring real-time performance. However, challenges such as generalization issues, overfitting, and the dependency on synthetic data, as well as the model's complexity for resource-constrained devices, remain significant hurdles. [5]

Zhu, Chen, and Gao In "Improvement of Face Detection Algorithm Based on Lightweight Convolutional Neural Network" , Zhu, Chen, and Gao present a face detection algorithm optimized for mobile platforms, leveraging a lightweight CNN architecture built on MobileNet and deformable convolution layers. Knowledge distillation is used to enhance the model's performance by training a smaller network with the guidance of a larger, more complex model. While this approach reduces model size and enhances efficiency, the use of deformable convolution layers increases computational complexity. Additionally, fine-tuning the distillation process to achieve optimal performance across various mobile devices presents challenges.[6]

3. METHODOLOGY

3.1 Proposed Work

Accurate head pose and gaze estimation plays a critical role in real-time student attention tracking, especially in online education settings. By leveraging MediaPipe, the system utilizes Face Mesh for precise face detection and head pose estimation to continuously monitor student engagement during virtual classes. This approach is both lightweight and efficient, making it ideal for deployment on low-power devices such as mobile phones, ensuring accessibility for a wide range of users. The system can be seamlessly integrated into existing online learning platforms or used independently as a standalone tool to provide valuable insights into student focus levels. Pilot testing across diverse environments has demonstrated the system's effectiveness in assessing and enhancing student attention, contributing to improved learning outcomes. Additionally, continuous updates to the technology ensure the system remains adaptable to new advancements, further optimizing student engagement and providing real-time feedback to educators.

This dynamic, real-time monitoring enhances the learning experience, fosters better classroom management, and empowers educators with actionable data to improve student participation and retention.

3.2 System Architecture

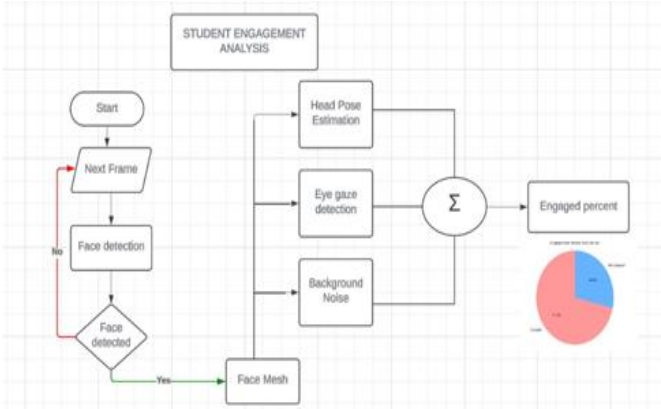

Fig -1: Proposed Architecture

The Student Engagement Analysis System Architecture processes a real-time video stream to assess student engagement during online learning. The system captures each frame and performs face detection; if no face is detected, it loops to the next frame. When a face is detected, the Face Mesh module extracts facial landmarks, which serve as input for three parallel subsystems: Head Pose Estimation, Eye Gaze Detection, and Background Noise Detection. These components analyze the student's head orientation, gaze direction, and surrounding auditory environment, respectively. The outputs of these subsystems are aggregated in a summation module to compute the engagement percentage, which is then visualized (e.g., a pie chart) to classify the student as either "Engaged" or "Not Engaged." This architecture ensures efficient and real-time analysis by leveraging modular components to provide insights into student behavior and attention during online sessions.

3.3 Dataset

In this area, we offer a thorough rundown of the datasets that were used for the model's training and testing. Effective learning of the model is ensured by a well-chosen training dataset, and trustworthy evaluation under a variety of real world scenarios is guaranteed by a strong testing dataset.

The pre-trained models of MediaPipe are able to generalize well across a range of individuals and contexts since they have been trained on extensive, diversified datasets that include a variety of facial photos and videos. Comprehensive facial annotations are included in these datasets, which are frequently derived from open datasets like 300W, AFLW, and LFW. The training set of data includes a wide range of head positions, facial expressions, and lighting scenarios, which improves the model's real- time recognition of important face traits.

Because of the diversity of the data, the model can function well in real-world settings with widely varying facial angles, lighting conditions, and user demographics. The pre-trained model's ability to identify important facial landmarks, like the eyes, nose, and mouth, is essential for gaze tracking and attentiveness assessment. These landmarks play a crucial role in identifying a person's gaze direction, facial alignment, and minute shifts in expression that may indicate distraction or attentiveness. The diversity of the training data is essential for enabling real-time processing in a variety of contexts, such as online meetings and live lectures.

We used the 300W-LP dataset to test, as well as to validate the model's accuracy in identifying facial landmarks and head positions. In the field of computer vision, this dataset is well known for its comprehensive annotations of head positions. The 300W-LP dataset offers a solid standard for assessing the effectiveness of head pose estimation algorithms since it contains thousands of photos with thorough annotations of facial landmarks taken with various head positions. The dataset can be used to assess how well the MediaPipe model holds up in practical settings because it includes a variety of facial expressions, occlusions, and lighting conditions.

The images in 300W-LP are captured from a variety of perspectives, including awkward positions, to replicate difficult real-world scenarios. This guarantees that the model can accurately identify facial landmarks in challenging scenarios, as when the face is slanted or partially obscured, in addition to accurately identifying frontal faces. Additionally, the dataset has a variety of backgrounds, which introduces another level of complexity to evaluate the model's adaptability to various settings. We can assess the model's generalizability and capacity to sustain high performance in a variety of scenarios by using this dataset for testing. This makes the model suitable for usage in online learning environments where students may interact with the system in a range of scenarios.

3.4 Data Processing

Preparing datasets for model training depends on data processing leveraging Pandas and Keras DataFrames. Pandas effectively manages missing values by imputing or deleting them, therefore guaranteeing dataset integrity and lowering bias. Usually between 0 and 1, normalizing numerical features to a common scale helps to avoid bigger features controlling the model training process. To reduce dimensionality and improve computing speed, unnecessary columns are deleted. Keras DataFrames provide effective data preparation for neural network designs by elegantly interacting with deep learning frameworks. They also preserve data integrity by handling missing values just like Pandas does. Keras DataFrames maximize resource use and

minimize overfitting by normalizing numerical characteristics and eliminating useless columns, therefore guaranteeing improved model convergence and performance.

3.5 Feature Selection

This code initializes MediaPipe's Face Mesh solution's settings for head pose estimation. It creates variables (X AXIS CHEAT and Y AXIS CHEAT) to track the head postures on the X and Y axes, as well as placeholder values for extra modifications. A maximum of two faces with defined detection and tracking confidence levels can be detected in real-time from a video feed using the 'mp face mesh module. A 3D array, 'face 3d', defines essential facial locations, such as the nose tip and corners of the eyes and mouth, which are crucial for determining head pose orientation. In order to facilitate real-time input for the face mesh recognition and analysis, the code additionally sets up a video capture from the default camera. The left and right eye corners are designated as the origin for gaze-tracking computations by modifying the 3D coordinates of facial landmarks.

To ensure proper alignment for further analysis, the coordinates of the left eye are adjusted by adding and removing specified values, and the right eye undergoes a similar change. In order to facilitate seamless transitions and the tracking of gaze movements over time, variables are further set to store the gaze scores from the previous frame. In order to enable real-time audio input for background noise detection, this code initializes the PyAudio library. It uses a single channel and a buffer size of 1024 frames to set up an audio stream with a sample rate of 44.1 kHz. The program can continually monitor audio levels since variables are defined to track the start time of noise detection and whether noise has been detected. This configuration makes it possible to monitor ambient noise levels effectively, which makes it easier to assess classroom circumstances while analyzing student involvement.

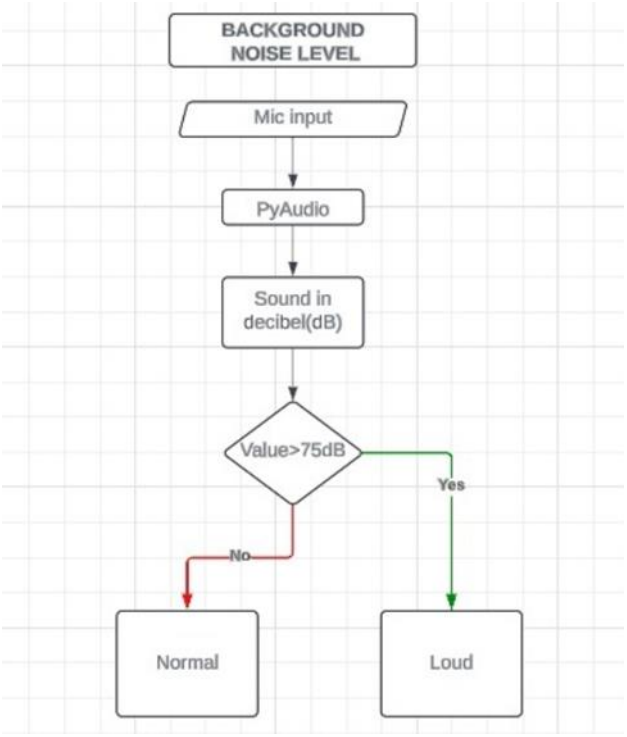

Fig -2: Background Noise Detection Process

4. EXPERIMENT RESULTS

Promising outcomes have been observed in the monitoring learning sessions since the real-time attention evaluation system was implemented to improve online education. The system proficiently records and examines a range of indications, including facial expressions, gaze patterns, and interaction behaviors, by means of smooth integration with widely-used online learning platforms and the utilization of sophisticated data processing and machine learning techniques.

```
Looking Straight    91.571429
Looking Left         5.571429
Looking Right        2.857143
Name: statement, dtype: float64


Engaged         78.0
Not Engaged     22.0
Name: Engagement, dtype: float64
```

Fig -3: Analysis Report

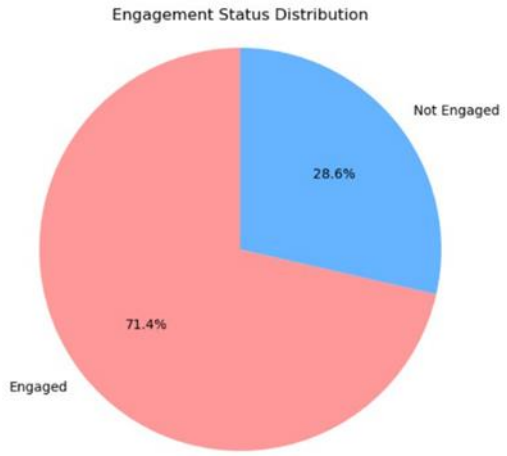

Fig -4: Student Engagement Status Distribution

The output data shows that pupils focused primarily on the screen, gazing straight ahead and enhancement of student involvement during virtual astounding 91.57% of the time. Furthermore, the examination of student engagement indicates that 78% of the students were classified as engaged during the sessions, whilst just 22% were classified as not involved. This highlights the effectiveness of the system in promoting a more attentive and dynamic learning environment.

5. CONCLUSION

In conclusion, the real-time attention evaluation system's integration of MediaPipe has proven essential in revolutionizing the online learning environment. Celebrated for its effectiveness and real-time processing skills in computer vision tasks, MediaPipe has made it easier to capture facial landmarks and gaze direction with accuracy and ease. These are two essential components for evaluating student involvement. Teachers can gain valuable insights on their students' attention levels by using MediaPipe's pre-trained models to examine a variety of engagement markers. This ability enables prompt interventions that can improve learning results in addition to helping to detect when kids could be losing concentration. The system's functionality is further enhanced by the employment of sophisticated gaze estimation algorithms in conjunction with noise detection, enabling educators to modify their pedagogical approaches to suit the changing needs of their students in virtual settings.

As the system develops via ongoing enhancements and feedback-driven modifications, MediaPipe continues to be a vital component of its success. Real-time alerting systems enable teachers to make well-informed decisions based on up to-date information, creating a virtual learning environment that is more participatory and interesting. Because it promotes active participation, this proactive strategy to measuring student engagement can enhance academic performance and satisfaction among students.

6. FUTURE SCOPE

To further improve the system's capabilities, future developments might involve adding more complex machine learning algorithms and broadening the spectrum of observable behavioral and emotional signs. The real- time attention assessment system promises to make a big contribution to the future of online education by utilizing MediaPipe's capabilities and keeping students interested and supported throughout their learning journeys.

REFERENCES

[1] Hao Wu; Ke Zhang; Guohui Tian. Simultaneous Face Detection and Pose Estimation Using Convolutional Neural Network Cascade, IEEE

[2] Li Pengyu, Zhou Zhenkun, Li Haiyan, Zhu Yajing, "Improving CNN Model for Residential Building Image Classification: Enhancing Parameter Estimation Accuracy Through Transfer Learning and Reducing Model Complexity with MobileNet", 2023 3rd International Signal Processing, Communications and Engineering Management Conference. (ISPCEM), pp.50-54, 2023.L. Bai, J.

[3] Rowley HA, Baluja S, Kanade T (1998) Neural network-based face detection. IEEE Trans Pattern Anal Mach Intell 20(1):23–38.

[4] Wang, Q.; Qian, W.Z.; Lei, H.; Chen, L. Siamese Neural Pointnet: 3D Face Verification under Pose Interference and Partial Occlusion. Electronics 2023, 12, 620.

[5] B. Thaman, T. Cao, N. Caporusso, "Face Mask Detection using MediaPipe Facemesh".

[6] H. Ran, W. Xiang, S. Zhenan, and T. Tieniu, "Wasserstein CNN: learning invariant features for NIR-VIS face recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 7, pp.1761–1773, 2018.

[7] Li Pengyu, Zhou Zhenkun, Li Haiyan, Zhu Yajing, "Improving CNN Model for Residential Building Image Classification: Enhancing Parameter Estimation Accuracy through Transfer Learning and Reducing Model Complexity with MobileNet", 2023 3rd International Signal Processing, Communications and Engineering Management Conference (ISPCEM), pp.50-54, 2023.

[8] C. Huang, J. Shi, "Real-time Head Pose Estimation Using Deep Neural Networks and Face Landmarks," IEEE Access, 2021.

[9] S. Kumar, A. Sharma, “Efficient Head Pose and Gaze Estimation Using Lightweight CNNs,” arXiv preprint, 2022.

[10] J. Nguyen, T. Pham, “Gaze and Face Detection for Attention Estimation in Online Learning Environments,” ACM, 2021.

[11] M. Zhou, L. Zheng, ”Real-time Student Engagement Detection with Gaze Estimation,” IEEE Transactions on Education, 2022

[12] F. Zhang, Z. Yu,”Integrating Facial Landmark Detection and Head Pose Estimation for Real-Time Gaze Tracking,” IET Computer Vision, 2022.

[13] E. E. Ye, J. E. Ye, J. Ye, J. Ye, R. Ye,”Low cost Geometry-based Eye Gaze Detection using Facial Landmarks Generated through Deep Learning”.

[14] D. Patel, A. Chattopadhyay, “Real-Time Eye Gaze Estimation for Engagement Tracking in Virtual Classrooms”.

[15] L. Wang, Z. Zhang, Y. Yang, ”Lightweight Face and Pose Estimation Models for Real-Time Applications”.