IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96120

REAL TIME ATTENTION SPAN TRACKING IN ONLINE EDUCATION

Chinthu H A[1], Nirmala Hiremani[2]

[1]Student, Department Of MCA, Visvesvaraya Institution of Advance Technology, muddenahalli, Chickballapur, India.

[2]Professor, Department Of MCA, Visvesvaraya Institution of Advance Technology, muddenahalli, Chickballapur, India.

Abstract: Over the last 10 years, e-learning has changed the way students study by offering them access to high-quality instruction whenever and wherever they need it. However, understudies regularly become preoccupied for a variety of reasons, all of which have a negative impact on their learning ability. Many experts have attempted to improve the nature of online education, but we need a comprehensive solution to this problem. This research aims to present a method that uses a camera feed and a mouthpiece contribution to monitor student's continuing attention levels during online classes. Throughout this review, we look at several photo handling strategies and AI algorithms. We suggest a framework that utilises five specific non-verbal features to calculate the understudy's consideration score throughout computer-based tasks and generate continual input for both the understudies and the association. We can use the generated data as a heuristic value to investigate the general exhibition of understudy as well as the speaker's exhibiting guidelines.

Keywords: Artificial Intelligence, Attention, Blink rate, Drowsiness, Eye look following, Emotion arrangement, Face acknowledgment, Body Posture assessment, Noise recognition.

1. INTRODUCTION

Over the last decade, e-learning has revolutionized how students learn by providing them access to quality education whenever and wherever they need. However, students often get distracted due to various reasons, which affect the learning capacity to an excellent extent. Many researchers are trying to enhance the quality of online education, but we'd like a holistic approach to deal with this issue. This paper intends to provide a mechanism that uses the camera feed and microphone input to watch the real-time attention level of students during online classes. We explore various image processing techniques and machine learning algorithms throughout this study. We propose a system that uses five distinct non-verbal features to calculate the eye score of the student during computer based tasks and generate real-time feedback for both students and the organization. we will use the generated feedback as a heuristic value to research the overall performance of students as well because the teaching standards of the lecturers.

1.1 BRIEF DESCRIPTION OF THE PROJECT

The demand and need for online education are increasing rapidly. Almost all the schools and colleges throughout the world have shifted to the online mode of lectures and exams due to the recent corona virus outbreak, and this trend will most likely continue in the upcoming years. The increasing demand for online education opens the gate to automation in the field. One major issue in the online mode of lectures is that students tend to lose their concentration after a certain period and there is no automated mechanism to monitor their activities during the classes. Some students tend to just start a lecture online and move away from the place, or might even use a proxy to write online tests for them. This situation also takes place in online course platforms such as EdX and Coursera where the student tries to skip lectures just for the sake of completion and certification. The loss in concentration not only affects the student's knowledge level but also hurts the society by producing low-skilled labourers

2. WRITING REVIEW

The purpose of being able to focus location during online classes is to gather data and break down the condition of the understudy, to evaluate his presentation based on fixation level rather than just scholastic scores. According to an individual's typical squint pace is between 8 and 21 flickers per second, but when the individual is intensely focused on

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96120

a specific visual task, the rate of flickering drops to a normal of 4.5 flickers per second. When the singular's fixation level is low, the flicker rate increases to over 32.5 squints per second. The evaluation looks into how understudies' close-to-home situation develops over time and how deep input might help to further increase learning experiences. Sentiments like bitterness, dread, wrath, and loathing indicate a pessimistic meeting, whereas feelings like satisfaction, happiness, shock, and unbiased indicate a positive beneficial potential for growth. The review discusses how mind wandering can impair performance and how eye gaze data collected with a dedicated eye tracker can naturally indicate lack of thought during PC-based tasks. The eye's behaviour was observed, and it was discovered that the instances shifted, especially during mind meandering. Article explains what a lack of consideration means in terms of understudies' learning abilities. Sluggishness, caused by rest, fretfulness, and mental tension, is one of the fundamental points that leads to a lack of thought, according to the audit study . The Haar classifier and Support Vector Machine (SVM) were found to have superior precision in escalating circumstances when compared to other cutting-edge weariness recognition systems. The review establishes that understudies' focusing capacity is influenced by natural commotion settings, and the findings suggest that clamour levels more than 75dB have a significant impact on understudies' precision. The research offers a Convolutional Neural Network (CNN) and Principle Component Analysis-based computerised facial recognition model (PCA). To acquire familiar with the significant level meagre and specific face component maps, the review offers a two-layer CNN. Scanty Representation Classifier (SRC) uses a haphazardly chosen inclusion extractor to work on the exhibition. The review looks at the highlights of body attitude and makes a beeline for predicting the client's level of regard by detecting examples of consideration-related behaviour. We discovered from the writing audit that the five boundaries - squint rate, look, eye stare, foundation commotion, and body position - compose a good list of skills to assess the understudies' attentiveness level.

3. METHODOLOGY

This examination employs five boundaries to assess the understudy's abilities to concentrate in a web-based lesson. The understudy's participation is approved by facial acknowledgment. The capacity to focus score is calculated using flicker rate, glance, eye gaze, foundation uproar, and body act, and it is updated continuously for a 5-second window. Rather than being executed in order, once the internet-based address begins, all of the models anticipated to compute the ability to focus are executed in the same order. This is performed by multithreading all of the capacities, which plays a vital role in reducing the time spent on each model as well as the complete framework. The model will construct the ability to focus score and deliver constant criticism to the understudies as live charts that are drawn for every boundary, as well as the determined ability to focus score, as if it were a clockwork. The sections that follow will go over each of the models used in this review in detail, as well as their significance in determining one's capacity to focus. The suggested framework's overall design is depicted in (Fig. 1), and the operation of each module is depicted as flowcharts in (Fig 2).

Fig. 1. The architecture of the proposed system.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⋇ Impact Factor 7.105 ⋇ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96120

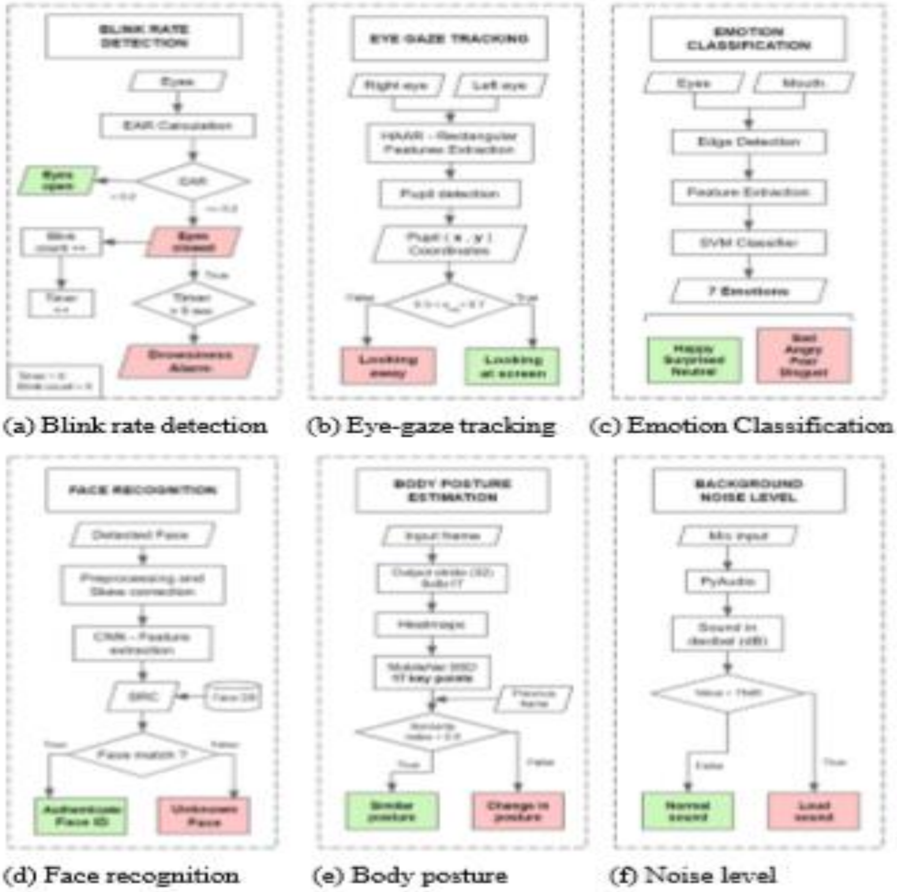


Fig. 2. Flowcharts of each module

A. Facial Landmark Detection

The Viola-Jones computation [13], which employs a windowing approach to analyse images for differentiating highlights of human countenances, is used for face recognition. The study [8] shows how to use OpenCV's module to deal with 68 fundamental concerns from a recognised facial image in a consistent and efficient manner (Fig. 3). To isolate Haar highlights from the rest of the image, we use rectangular regions. The milestones are organised into five facial element classifications: brows, eyes, nose, mouth, and jaw, which are shown sequentially using the central concerns. These unique milestone highlights will be used as contributions to future modules. By processing a larger number of central concerns, we can improve the precision of the face identification module. However, this lengthens the handling time.

Fig. 3. Face detection and Facial Landmarks.

B. Squint Rate Detection

One of the important variables in determining the perspective - whether the understudy is effectively tuning in or drowsy throughout the session - is the blink rate. We crop the localities containing the eye matches in this module and divide each eye into equal pieces. To determine whether the eyes are open or closed, we calculate the Eye Aspect Ratio (EAR) using Euclidean distances (Fig. 4a) for each edge according to Formula (1). We also have a start clock that starts once a squint is detected and keeps track of the number of seconds the eyes are closed. If the eyes are closed for more than two seconds [5], it is extremely likely that the client is weary (lack of consideration), and an alarm will be

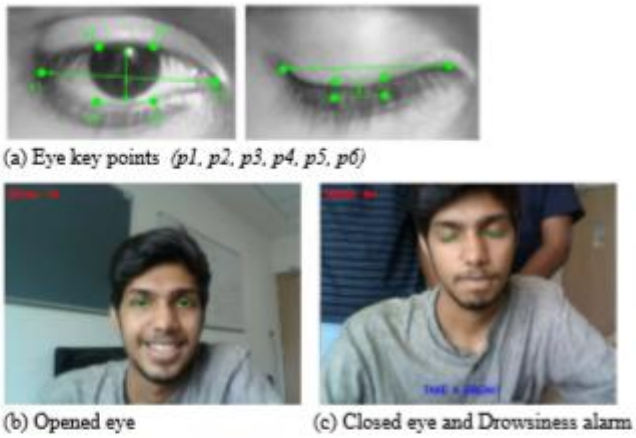
IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96120

delivered both visibly as shown in (Fig. 4c) and with cautioning alert noises. We calculate the number of flickers over a period of seconds to get the client's average squint speed. In light of test results, the EAR edge esteem has been set to 0.2. We used the flicker identification module on 306 images with 156 closed eyes and 150 open eyes, and the squints were ordered with91.02% precision and the open eyes with 92.66% precision.



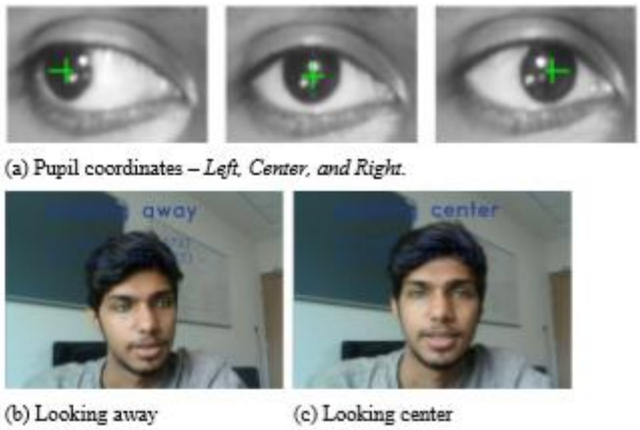
(a) Eye key points *(p1, p2, p3, p4, p5, p6)*

(b) Opened eye (c) Closed eye and Drowsiness alarm

Fig. 4. Blink rate and drowsiness detection

C. Eye-stare Tracking

The understudy's eye-stare is closely monitored to determine where he is glancing out, and it is typically linked to the understudy's interruption level. We dissect the excised eye locale arranges for rectangular elements to distinguish eye portions comprising the student, as recommended in [10]. To define the eye stare path, the understudy facilitates (x, y) of each eye (Fig. 5a) are determined and planned. We've laid out two potential types of eye gazing based on the screen's goal: looking at the screen (Fig. 5c) and turning away (Fig. 5b) (right or left). There are 150 images in all, divided into two categories: direct gaze and turning away.



(a) Pupil coordinates – *Left, Center, and Right.*

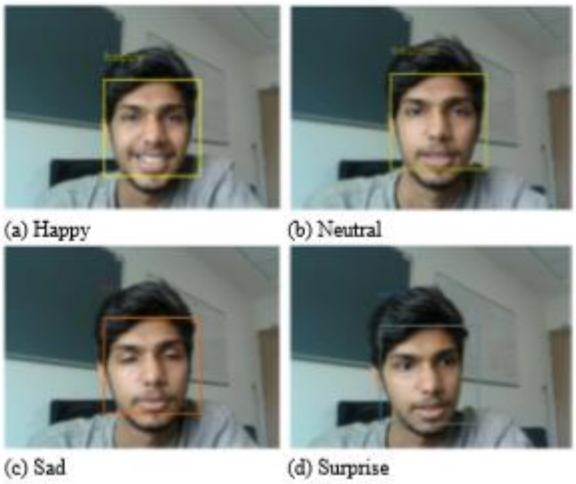
(b) Looking away (c) Looking center

Fig. 5. Eye-gaze tracking

D. Feeling Classification

The individual's perception of the internet-based class plays a key role on his level of consideration. This study uses Haar cascade classifier and face milestone finder to differentiate facial features such as eyes, nose, and mouth. The understudies' feelings are classified into seven categories using the Support Vector Machine (SVM) algorithm: anger, scorn, dread, joyful, miserable, shock, and neutral. Every tendency is given a score based on its impact on the client's level of consideration. Only the trimmed eye is used in the traditional method of sensing characterization pre-handling. In any case, [11] suggested an optional solution for higher precision by using mouth highlights. When compared to the Sobel edge eye finding, this technique only uses Haar fountains to group the inclination, therefore the handling speed was much faster. The model was validated using the JAFFE dataset, which includes 213 images and 7 emotions shown by ten different Japanese women. The preparation package included 42 items and seven different types of feelings. There were 70 images in the test set. On our test set, we got a normal exactness of 82.55 percent (Fig. 6).

International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified ⋇ Impact Factor 7.105 ⋇ Vol. 9, Issue 6, June 2022
DOI: 10.17148/IARJSET.2022.96120



D. Feeling Classification

The individual's perception of the internet-based class plays a key role on his level of consideration. This study uses Haar cascade classifier and face milestone finder to differentiate facial features such as eyes, nose, and mouth. The understudies' feelings are classified into seven categories using the Support Vector Machine (SVM) algorithm: anger, scorn, dread, joyful, miserable, shock, and neutral. Every tendency is given a score based on its impact on the client's level of consideration. Only the trimmed eye is used in the traditional method of sensing characterization pre-handling. In any case, [11] suggested an optional solution for higher precision by using mouth highlights. When compared to the Sobel edge eye finding, this technique only uses Haar fountains to group the inclination, therefore the handling speed was much faster. The model was validated using the JAFFE dataset, which includes 213 images and 7 emotions shown by ten different Japanese women. The preparation package included 42 items and seven different types of feelings. There were 70 images in the test set. On our test set, we got a normal exactness of 82.55 percent (Fig. 6).

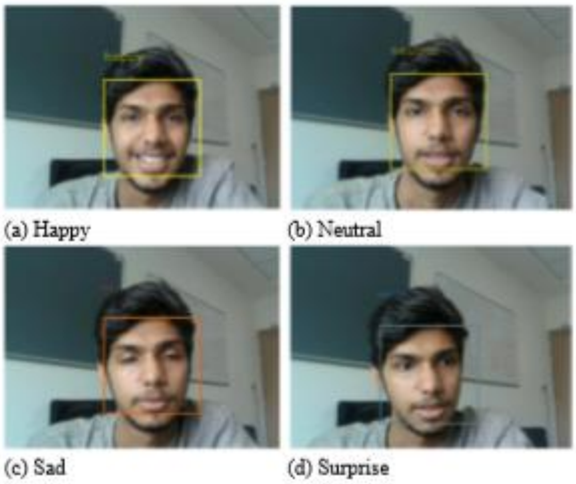


Fig. 6. Emotion Classification (Happy, Neutral, Sad, Surprise)

E. Face Recognition

A solid facial recognition framework is required to authenticate the understudy using biometrics in order to avoid understudy intermediaries and to automate the participation of executives during classes using the webcam stream. In this module, we modified [8]'s architecture and implemented a three-layer Convolutional Neural Network (CNN) with max-pooling, a totally associated layer with dropout, and a Sparse Representation Classifier (SRC) output layer. The dropout layer contributes to the reduction of computing costs. A dataset of 500 photos with four different faces was developed. We divided the dataset into 300 training photos and 200 validation images. After, 15 epochs of training, we attained a training accuracy of 94.8% and a validation accuracy of 90% (Fig. 7).

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁒ Impact Factor 7.105 ⁒ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96120

Fig. 7. Face Recognition

F. Body Posture Estimation

Many scientists have used CNN or R-CNN to assess posture with great accuracy. Nonetheless, the primary goal of our investigation is to continually compute the understudy's consideration level without settling on a handling time. Following that, we use the TensorFlow posture assessor (Pose Net) in conjunction with Mobile net SSD to assess the understudy's stance. The intensity guide is used in this model to determine the posture difference from one casing to the previous edge. Face, shoulders, elbows, wrists, hip, knees, and lower leg are among the 17 primary concerns that Pose Net can detect. We assign a pixel similitude score based on the difference in head posture and body stance between sequential cases to predict whether the understudy is worried or focused during the internet-based address.

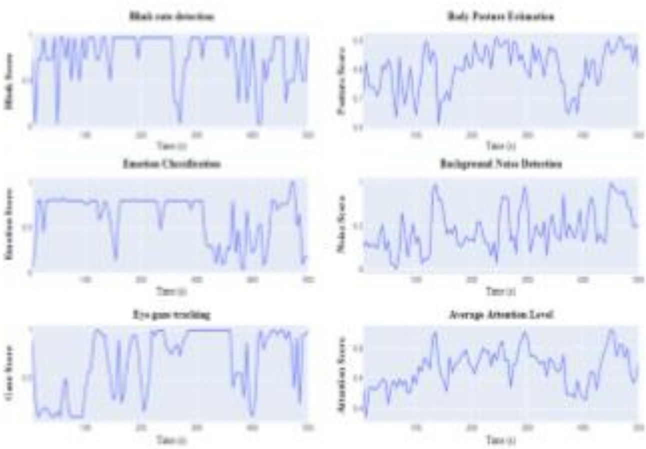
Fig. 8. Body Posture Estimation

G. Foundation Noise Detection

To recognise the information sound from the device's mouthpiece, we use the Python package PyAudio. The understudy's focus level may be influenced by the fundamental commotion during instruction. The average sound level in a school is 50 decibels, with 75 decibels being the limit for clearly audible commotion [6]. Anything beyond 75 decibels is deemed a raucous environment, and the scores will be inversely proportional to the foundation commotion. The model will continuously monitor the foundation turmoil, and we will determine the typical uproar level as if it were a clock.

H. By and large Attention level Detection

To get the consideration score, all of the scores from the previous boundary scores (squint rate discovery, eye stare following, sensation grouping, body act assessment, and foundation clamour location) are standardised using Formula (2). We draw live diagrams, as shown in (Fig. 9), with the understudy's expected consideration level and the scores for each border refreshed in real time. We don't employ facial recognition in the scoring technique because it doesn't contribute to determining the understudy's consideration level; rather, we use it for biometric authentication and computerised involvement of the understudies. $\sum *100$



IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁘ Impact Factor 7.105 ⁘ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.96120

4. EXECUTION EVALUATION (RESULT)

A dataset of 15 college understudies, consisting of nine men and six women, was used to assess the framework's performance. The students were asked to participate in 500-second online conversations on diverse topics, and three human eyewitnesses were asked to give a noted consideration score based on the recorded web camera footage, which was used as the ground truth-value. To analyse the overall exhibition of our framework, we compared the expected and noticed consideration scores. Table (1) presents the framework's exhibition metrics and (Fig. 10) demonstrates the link between predicted and noticed ratings.

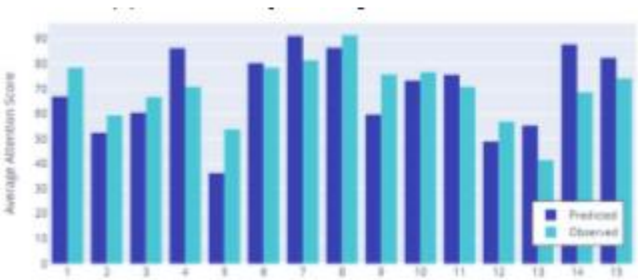


Fig. 1. Predicted Attention Score vs Observed Attention Score Comparison.

TABLE I. PERFORMANCE METRICS

Metric	RMSE	MAE	R2	MAPE
Value	11.152	9.837	0.154	15.248

Given the minimal data utilised to train the models, our system performed admirably. By averaging the accuracies of each module, we were able to determine the overall accuracy of our attention-tracking model. As seen in Table, compiling OpenCV's DNN module and Caffe with CUDA support improved model performance and greatly reduced inference time (2). With a total accuracy of 84.6233%, we were able to complete the task.

5. CONCLUSION AND SCOPE FOR FUTURE ENCHANCEMENT

CONCLUSION

we have implemented a system to tackle the issues involved in online education using five parameters. We used the face recognition model to verify the student attending the online class. We used the other five parameters - blink rate, eye gaze, emotion, posture, and noise level to calculate the attention level of the student throughout the lecture. Since this involves real-time processing, we have implemented and used lightweight models to reduce the processing time. We visualize the scores in the form of a live graph and generate automated reports. The feedback generated can be used for:

1) Evaluating student performance
2) Improving teaching standards
3) Preventing malpractice during online examinations

FURTURE ENHANCEMENT

As an element of future works, we will improve our system's performance further by training our models using more data. Also, the identical attention tracking mechanism can be further optimized to simultaneously work with multiple subjects during a classroom using video footage from the CCTV cameras. Moreover, we've used human observed attention scores as ground truth-values as we currently don't have any dataset for measuring the eye span during online lectures. A standard dataset can help to gauges the system's performance more reliably.

6. REFERENCES

[1] Mark B. Abelson, MD, CM, FRCSC, ORA staff, Andover, Mass, It's Time to Think About the Blink, Review of Ophthalmology June 2011.

[2] Liping Shen, Minjuan Wang, and Ruimin Shen, "Affective e-Learning: Using "Emotional" Data to Improve Learning in a Pervasive Learning Environment", International Forum of Educational Technology & Society 2009.

International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified ⁙ Impact Factor 7.105 ⁙ Vol. 9, Issue 6, June 2022
DOI: 10.17148/IARJSET.2022.96120

[3] Bixler, R., D'Mello, S. Automatic gaze-based user-independent detection of mind wandering during computerized reading. User Model User-Adap Inter 26, 33–68 2016.
[4] Smallwood, J., Fishman, D.J. & Schooler, J.W. Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. Psychonomic Bulletin & Review 14, 230– 236 2007.
[5] M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, "A Survey on State-of-the-Art Drowsiness Detection Techniques," in IEEE Access, vol. 7, pp. 61904-61919, 2019.