Case Study on Student's attentiveness detection

Chetna Khandagle
Department Of IT and MCA
Vishwakarma Institute Of Technology
Pune,India

Rishabh Jaiswal
Department Of IT and MCA
Vishwakarma Institute Of Technology

Ajay Shelke
Department Of IT and MCA Vishwakarma Institute Of TechnologyPune,India

Abhishek Nangre D
epartment Of IT and MCA
Vishwakarma Institute Of TechnologyPune,India

Prashant Vanse
Department Of IT and MCA
Vishwakarma Institute Of TechnologyPune,India

ABSTRACT

The importance of student attention in the educational system has increased. For the purpose of monitoring classrooms and giving teachers feedback, automated and effective solutions are required. In this study, we describe architecture for detecting students' attention. By examining the student's head pose, Eye gaze and facial expressions. Additionally, In this paper we have done a thorough comparison of current systems that are documented in the literature.

Keywords: Face Recognition, Head pose estimation, eye Gaze detection, Body Gesture Classification

1. Introduction

Choosing and concentrating on what is With the help of a wide range of materials, e-learning gives students a quick and comfortable way to access online lessons and learn at their own pace. It has experienced remarkable growth in recent years, particularly with the rise in popularity of learning platforms like Udemy and Coursera. The COVID-19 global pandemic has also compelled educational institutions to swiftly alter how they impart knowledge to students, increasing interest in e-learning. In Order to learn, one must first pay attention. Avoiding distractions is necessary to

Complete this. Although many authors claim that student pay close attention to the teacher at the beginning of the lecture, the majority of them claim that after about 10 minutes, they start to lose interest [1]. This causes educational institutions to view students' attention not only as a tool to enhance learning but also as an essential element that must be assessed and examined The primary goal of this work is to utilize various technologies for the detection of facial and bodily expressions in order to identify and analyze the student's attentiveness. The ultimate aim is to know exactly how much attention a learner is paying at any one moment. In order to identify and measure attention,

this study suggests a multimodal architecture. This architecture is adaptable enough to support a potential evolution, such as the addition of additional components. On the other hand, it was made to be flexible enough to work in accordance with the changing circumstances of a classroom.

In this paper we first discussed the current systems. The architecture of system is then

thoroughly described. Then, we go over the primary advantages and disadvantages of current systems as well as the features that method provides. We draw some conclusion at end of the paper.

2. Current Systems

A technique for tracking student concentration in an online learning environment has been suggested by Krithika et al. [5]. Their task is to use two crucial measurement head rotation and eye movement to determine a student's level of attention. One of the three degrees of concentration that the authors define high, medium, or low level of concentration is obtained from the examination of these components.

By monitoring student participation through the classification of face features, the developers of [6] hope to integrate emotional calculus to assist teachers in evaluating the effectiveness of their lessons. They also strive to build a connection between facial expressions and learning ability while taking into account that the student's participation has three components: behavioral, cognitive, and emotional.

Zaletelj [7] The fundamental concept behind their technology is to covertly gather behavioral data from students during lectures by utilizing the sophisticated capabilities of the Kinect One sensor. They put forth a method for calculating characteristics from the Kinect data that match to actions that can be seen with the naked eye. Then, they use machine learning techniques to create models that estimate each student's level of attention. Their main concern was establishing a link between the students' attention and the teacher's observations. They compare the attention scores obtained from human observers with the students' perceptible behaviors (activities, gestures, etc.). This enables the differentiation of the many behaviors related to a level of attention. The model for estimating the student's level of attention is then created using a machine learning technique using the features offered by the Kinect One sensor.

An autonomous agent that can track students' attention and provide output data for this component for each student has been proposed by Canedo et al. [8].According to [9], the orientation of the head contributes 68.9% to establishing the direction of the gaze and reaches an accuracy of 88.7% when determining the focus of attention. This justifies their decision to use the head approach as a potent indicator of pupil attention. As they take into account the fact that pupils who are paying attention typically respond to a stimulus in the same manner.In other words, it is thought that pupils are paying attention if their movements are in rhythm with those of the majority. When the teacher instructs the class to record something significant, that is an illustration of this synchronization in action. [8], [10]. They made the decision to combine the head orientation approach with a second technique based on the body gesture in order to enhance the performance of their system.

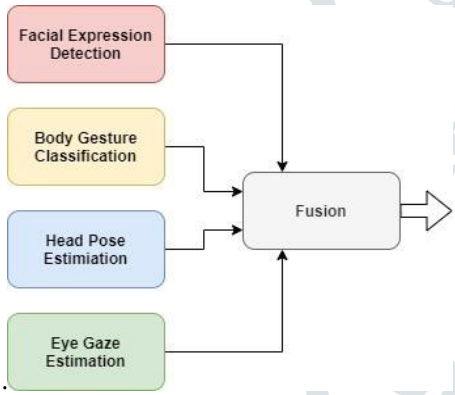
Whitehill paper's [11] studied the creation of real-time automated engagement recognition using student facial expressions .The first stage of their work involved gathering experimental data for the engagement study of 34 participants who took part in the Cognitive Skills Training study in the spring 2011 at the Historically Black College/University(HBCU) and the Cognitive Skills Training study in the summer 2011 at a university in California. The pupils were captured on camera while using an I Pad to perform cognitive activities. In order to select the most dependable and practical solution, various scenarios for picture la belling were investigated. One binary engagement classifier was developed for each of the four levels the authors had previously established. After assessing the precision of their model, they went on to reverse engineer the human labelers' decision-making process. Finally, they look into the relationship between student test performance and learning and human and automatic assessments of engagement.

3. STUDENT ATTENTION DETECTION SYSTEM

The essential processing nodes of the framework are depicted in the basic schema of Figure 1. A camera will supply a stream of high

definition images encoded in video format as the data for this system to process. Computing power is quite expensive in such a system. In order to achieve real-time execution, the various functionalities must be carried out appropriately. The proposed architecture for the detection and estimation of attention is thus split into four primary components (or "nodes") that must operate as a distributed processing. Each node is in charge of carrying out a variety of therapies independent

The tasks related to facial expression detection, body gesture categorization, eye gaze estimation, and head position estimation are handled by the nodes, as shown in Figure 1. To get a clear depiction of the level of attention, the outputs from these numerous nodes' varied outcomes must be combined (Fusion).



[12]Figure1

Webcam is used to acquire student images, and it is placed above the projector to catch the entire class. All of the students' faces must be included in the precisely taken image in this manner. At regular intervals (15 frames/second), a series of picture snapshots are taken.

The importance of utilizing a high-performance face detection algorithm can be seen in the fact that face detection is the first stage in our biometric analysis system and that its accuracy has a substantial impact on the efficiency of following operations. Then, every face that is visible on the image is retrieved and saved in the cropped format necessary for the analysis activities to be successful. In fact, the following devices will receive it:

Head posture estimation, facial expression recognition, and Eye gaze .

3.1Facial Expression Detection

The ability of a person to perform daily chores, from important and attention-demanding activities to enjoyable activities, depends heavily on his emotions. Immordino et al.[8] assert that there is a deeper connection than many educators recognize between learning, emotion, and bodily state and that the initial goal of brain evolution was to alter physiology in order to improve survival and promote flourishing [11].

Emotions play a significant role in nonverbal communication, and they have a variety of effects on cognition, including how we process information, how we pay attention, and how we perceive information [10].

We were motivated by the circumplex model of affect [9] to address the mechanism of emotion detection. According to the circumplex model of affect, excitement (activation) and valence are the two dimensions of a two-dimensional circular space in which emotions are distributed (pleasure).

The different state of arousal are represented on the vertical axis while the valence occupies the horizontal axis, the center of the circles signifies a neutral valence and an average level of excitation. In this model, each emotion can be represented by a level of valence and excitement, or at a neutral level of on other of these factors are illustrated in the Figure3.

camera provides a continuous non-intrusive way to capture images of students' faces. facial information can beused to understand certain facets of the student's current state of mind, and there are several techniques to automate this measurement process [21], [22].Knowing the affective state of the student can lead to deduce his level of attention or atleast gives partial information that can be combined with others to calculate his level of attention. That excitement (activation) and valence are the

two dimensions in which emotion(pleasure).

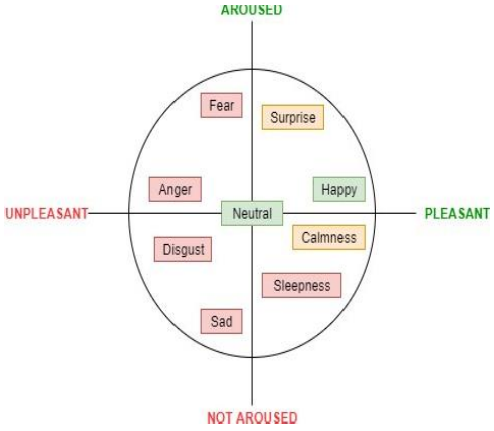
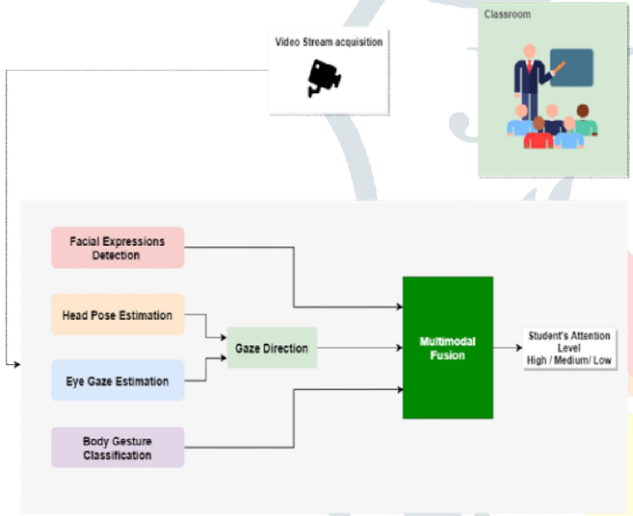


Figure 2



[12] Figure3

3.2Estimation of a head stance

Estimating head posture is necessary to determine where each student is looking in the classroom. Technical difficulties arise while estimating multiple people's head poses in a large classroom [7]. A camera that continuously outputs 2D images serves as our sole data source. We are particularly interested in the investigation of approaches for estimating the student's head position based on 2D photographs. These methods have demonstrated some unreliability, particularly when a person's eyes, nose, and mouth are covered. In fact, it has been noticed that during an instructional scene, pupils adopt common mannerisms like holding their heads up with one hand or scratching their hair. These brief interruption in the tracking of head posture are prevalent with these partial facial occlusions. In the real world, a student can concentrate on several topics in the classroom. An attentive student will face the instructor or the slideshow, showing that he is paying attention to the explanations. He can also be focused and looking at his notepad, but he must adopt the stance of someone transcribing notes. When a student looks

Found, fixes the ceiling, or stares at his notepad for an extended period of time without writing anything, this shows a lack of attention or a distraction.

3.3Eye gaze estimation

Although analytically modeling audience involvement or attention appears challenging, there are helpful indications that can be utilized to

estimate an audience's level of attention. A person's gaze is one very obvious cue. Particularly, useful measures of attention include where a person focuses their gaze and how long they keep it there [8].The length of an individual's eye fixations is directly proportional to how much of their cerebral processing resources is being devoted to a certain task, which can then be directly tied to how attentive they are[8], [2].There are numerous methods for precisely estimating the gaze position in the literature. Singh et al[3] .'s comparative analysis of eye gaze estimation methods is quite strong. camera provides a continuous non-intrusive only after the eyes are recognized could the eye gaze estimation be measured. Actually, a number of things, such the existence of occlusive objects or an excessively brilliant image, etc., might impede the detection of the eyes. The estimation of the eye gaze will be combined with the head pose to complete the data. In fact, good gaze direction prediction becomes a need when the eyes are visible. According to physiological studies [1, 2], a person's ability to predict where someone else

will look depends on both their head posture and eye direction. The eyes are depicted in the same arrangement in both of Figure 4's views of a head, which are presented at various angles.

looking at this image, it is evident that the head position has a significant impact on the perceived direction of gaze. The perceived direction is similar to that when the head is arranged frontally if the head is completely removed, leaving only the eyes[2].

figure 4: The Wollaston illusion: Despite the fact that the eyes are identical in both pictures, the direction of the apparent glance is determined by the position of the head [12, [33]].

3.4 Classification of body gestures

The issue of identifying attentiveness from body motions is exceedingly difficult. We looked at how the students behaved during the lecture in order to address this problem. In this instance, the upper
body is of particular interest. In a scenario where a student is seated behind a table, this later portion is what is seen. Each action has a corresponding amount of attention, such as taking notes or standing with the hand on the head and looking away from the slide show.

The authors of [4] provide excellent surveys on detecting body position. The survey article examines the several electronic gadgets available that enable body posture detecting jobs. It creates a standard against which to compare body posture data sets required for training the detection model and automatic recognition systems. As was previously noted, this module's goal in terms of our methodology is to identify postures that indicate a person is

paying attention. The output of the body posture detection node will accept the following values: leaning back, writing notes, holding up one's head, and sitting up straight.

	Emotinal	Gazedirection		Body gesture	Attention level
		Head Pose	Eyegaze		
1	Neutral	Slide		Upright sitting	High
2	-	Fixes the ceiling		Lean back	Low
3	Sad	Slide		Upright sitting	Medium
4	-	-		Writing notes	High

Table 1.The body and facial characteristics of students' attention states

3.5 Intermodal fusion

Our approach is designed to assist professors in identifying potential student attention lapses during lectures. Such a system must handle real-world situations flawlessly. In fact, it must rely on a variety of sources of knowledge that have been thoroughly examined together. To provide a multidimensional picture of the student's attention and to lessen the errors associated with the retrieved features, a fusion phase is required.

This system sends the student photographs to the various analysis units, which then extract features from them. Due to their heterogeneity, combining them is necessary to acquire all available information.

An artificial neural network is used to accomplish multimodal fusion (ANN). From the many face and bodily traits that were extracted by all of the system's units, the ANN was utilised to estimate the student's level of attention. The dataset of student attention states, which we have partially produced and will be regularly supplied with the observations of the teachers, will thus need to be learned with great precision. Table 2 provides a brief description of this dataset.

Our ANN's architecture consists of three layers:

an input layer I which gets data from the face and body features, a hidden layer (h), which processes the input layer's data, and an output layer (o), which allows for one of three outcomes: high, medium, or Low degree of attention. Figure 5 depicts the architecture of our neural network. We can perform the training that will enable us to ascertain the weights of the neurons in each layer of the network once the dataset of facial and body traits has been corrected by the experts

.

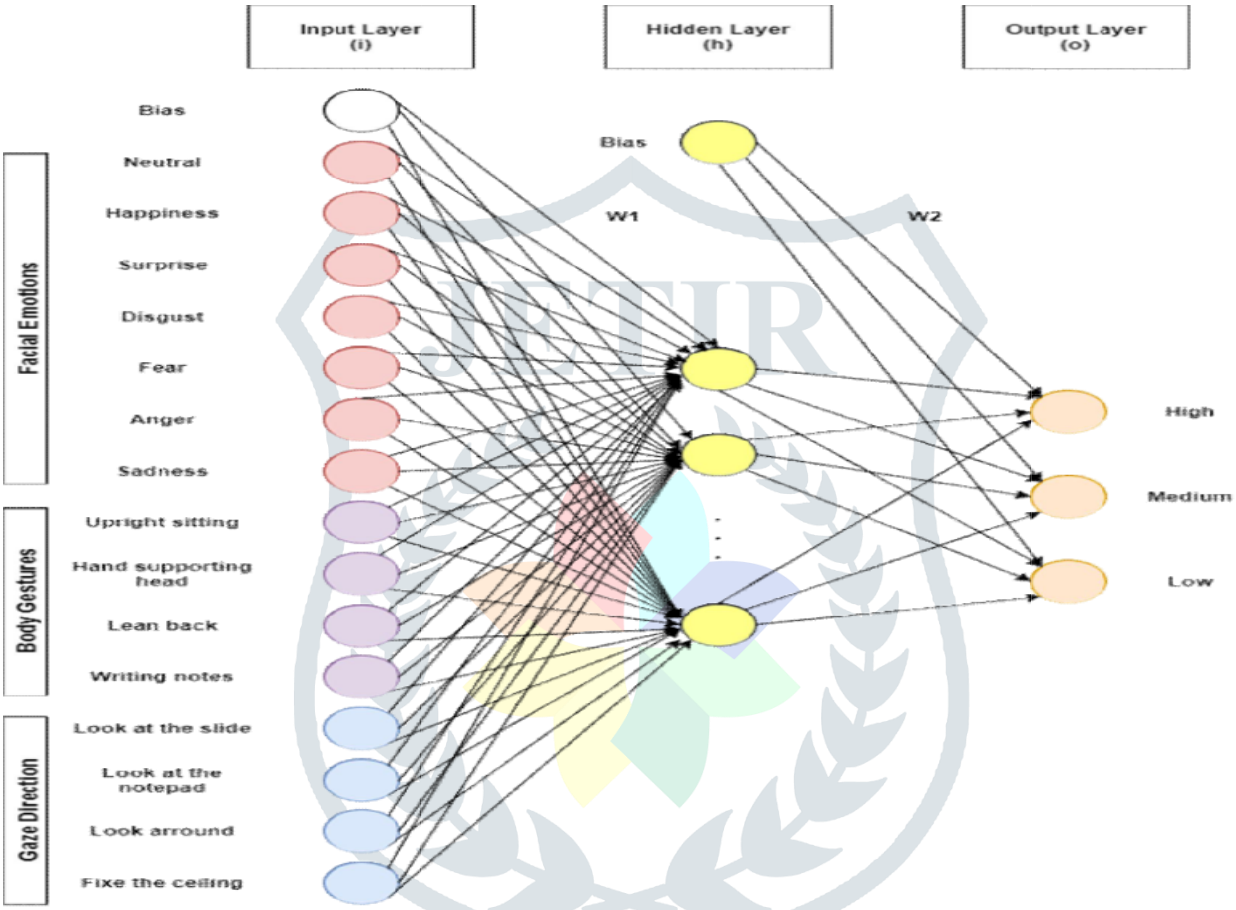


figure5:[12]ANN ensuring Multimodal Fusion

Reference	Eye gaze	Head pose	bodyg esture	Facialex pressions	Proand cons
[5]	non	yes	no	no	The system takes only a few measurements to determine the student's degree of concentration. The real direction of the student's attention cannot be determined by seeing their eyes.
[6]	no	no	no	yes	- One tool (emotions detection) is insufficient to evaluate students' levels of engagement.
[7]	yes	yes	yes	yes	The training dataset is rather small, with a total duration of 122 minutes for 18 people; the variations of human behavior that are present within the dataset are limited and do not cover all possible student behaviors; interpersonal differences in behavior during lecture were clearly visible and influenced the accuracy of a person-independent classifier; and the detection of all students' faces. Similar problems occur with skeletons of people, which are inaccurate because of obscured views of people seated behind tables.
[8]	yes	yes	no	no	- Despite his apparent concentration on the lecture, the student may not pay much attention.
[11]	no	no	yes	yes	- They adhered to a strict procedure for gathering and identifying photos, which increases the system's accuracy.

Table2: The detailed comparison of existing attention detection sytem

4. Conclusion

In this study, a distributed architecture for detecting student attention in a classroom using an analysis of body language and facial expressions is detailed. The main objective is to maintain the students' interest during the lesson. We have selected a number of features to investigate for our study in order to achieve this. The decision to include these elements was not made at random; rather, it was influenced by the data source we employ, which is the camera. In order to provide a general architecture that will get beyond the drawbacks of existing systems, we have also created a comparison of existing systems.The key drawback of our study is the restricted availability of real-world data; hence, the size of the available samples is insufficient for statistical measurement, which we will take into account in subsequent work. In order to complete our student attention detection dataset and evaluate the effectiveness of our architecture, we must now gather more samples.

REFERENCES

[1] N. A. Bradbury, "**Attention span during lectures: 8seconds,10minutes,ormore?**"AmericanPhysiologic alSocietyBethesda,MD,2016.

[2] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris,"**SchoolEngagement:PotentialoftheConcept, State of the Evidence**," *Rev. Educ. Res.*, vol. 74,no.1,pp.59–109,Mar.2004.

[3] N.Sabri,N.H.Musa,N.N.A.Mangshor,S.Ibrahim,andH.H. M.Hamzah,"**Studentemotionestimationbased on facial application in E-learning duringCOVID-19pandemic**," *Int. J. Adv. TrendsComput.Sci.Eng.*,vol.9,no.1.4SpecialIssue,2020.

[4] J. Zhang, E. Kamioka, and P. X. Tan, "**Emotionsdetectionofuserexperience(Ux)formobil eaugmented reality (mar) applications**," *Int. J.Adv.Trends Comput. Sci. Eng.*, vol. 8, no. 1.4 S1,pp.63–67,2019.

[5] KrithikaL.BandLakshmiPriyaGG,"**StudentEmotionRec ognitionSystem(SERS)fore-learningImprovementBasedonLearnerConcentrat ion Metric**," *Procedia Comput. Sci.*, vol.85,pp.767–776,2016.

[6] R. Manseras, T. Palaoag, and A. Malicdem, "**ClassEngagementAnalyzerusingFacialFeatureCl assclassification**,"no.November,pp.1052–1056,2017.

[7] J.ZaleteljandA.Košir,"**Predictingstudents' attention in the classroom from Kinect facial andbody features**," *EURASIP J. Image Video Process.*,vol.2017,no.1,p.80,2017.

[8] D.Canedo,A.Trifan,andA.J.R.Neves,"**Monitoring Students' Attention in a ClassroomThroughComputerVision**,"2018,pp.371 –378.

[9] R. Stiefelhagen and J. Zhu, "**Head orientation andgaze direction in meetings**," in *CHI'02 ExtendedAbstracts on Human Factors in Computing Systems*,2002,pp.858–859.

[10] M. Raca and P. Dillenbourg, "**System for assessingclassroomattention**,"in*Proceedingsof3rdIn ternationalLearningAnalytics&KnowledgeConferenc e*,2013,no.CONF.

[11] J.Whitehill,Z.Serpell,Y.- C.Lin,A.Foster,andJ.R.Movellan, "**The faces of engagement: Automaticrecognitionofstudentengagementfromfa cialexpressions**," *IEEE Trans. Affect. Comput.*, vol.5,no.1,pp.86–

[12] Hachad, Tarik & Sadiq, Abdelalim & Ghanimi, Fadoua & Hachad, Lamiae. (2020). A Novel Architecture for Student's attention detection in classroom based on Facial and Body expressions. International Journal of Advanced Trends in Computer Science and Engineering. 9. 7357. 10.3053

.