A MODEL OF VIDEO WATCHING CONCENTRATION LEVEL MEASUREMENT AMONG STUDENTS USING HEAD POSE AND EYE TRACKING DETECTION

[1]AYAT ALRAWAHNEH, [2] DR. SUHAILAN BIN SAFEI

Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin Gong Badak, 21300 Kuala Terrengganu, Malaysia

E-mail: [1] si3170@putra.unisza.edu.my, [2] suhailan@unisza.edu.my

ABSTRACT

E-learning has become popular across countries in recent years because of its flexibility, availability, and accessibility. During the COVID-19 pandemic, most schools and universities have switched to embrace E-learning for teaching purposes. One of the big challenges in online teaching is to monitor student's learning engagement especially in watching video learning materials. The concentration level of students in watching the video is among the important factors for having effective learning. While this aspect is essential, there has also been limited work to develop it. Although there have been a variety of studies and implementations for student concentration monitoring and evaluation in pervasive learning, the majority of the current works are either based on questionnaires, self-reports, instructor introspective assessments, and assignments or are assisted by commercial eye tracking devices/software. In this paper, we propose a model for measure a student's concentration level in an e-learning environment, using hybrid methods combining head pose and eye tracking detection. Specifically, obtain the information through the webcam of the student's head pose rotation and eye-tracking technique to discriminate the behavior of the students during the E-learning classes. For the performance evaluation experiment, we used a 5-minute E-learning video watched by 10 participants. We observe the scenario from an expert perspective and evaluate the students' high and low concentration levels. The video is then benchmarked with our model to see the model accuracy with respect to the expert decision. Based on this experiment, 80% of the result were similarly matched with the expert's opinion. Our extensive evaluations, which included different scenarios, and an expert perspective, demonstrate that the proposed method is effective and usable. This technique will enhance the student's concentration and make the session more interesting and updated. This approach could readily be expanded to generate real-time alerts for remote live monitoring in E-learning classes.

Keywords: *E-learning, Concentration Level, Facial landmark, Head Pose estimation, Eye Aspect Ratio.*

1. INTRODUCTION

The COVID-19 pandemic impacted people worldwide, forcing most activities, including learning, to switch online. In March 2020, the majority of colleges and schools moved from face-to-face to e-learning. E-learning has become increasingly popular and widely adopted by educational institutions and universities. It enables information to be provided whenever students need it at anytime and anywhere on the web. Due to the massive development of the E-Learning environment, there is an increasing demand to evaluate and measure the student's behavior and personalize E-learning (1, 2).

To ensure that these E-Learning environments are interactive and comparable to traditional classrooms, it is essential to maintain a high level of student concentration during online classes (3). The detection of student engagement has risen in importance as a method for evaluating each student's learning progress and, as a result, gaining indications for enhancing teacher instruction and student learning (4). One of the main concerns of instructors, among other concerns related to e-learning, is measuring student concentration (5).

Although concentration and learning are closely associated, concentration is recognized as one of the most major assessments for assessing learning efficiency. It is a significant metric that plays a significant role in the learning process (6). Compared to traditional classrooms, lack of concentration cannot be monitored dynamically or in real-time in e-learning environments (7). As a result, providing each student with a personalized learning environment is impossible (8).

Computer vision and machine learning techniques have been widely used in recent years to automate the personalization process for each student and improve teaching and learning quality.
Computer vision has significantly improved, and various complex problems can now be resolved using different aspects of computer vision(9). One of the aspects is face detection, detection of facial landmarks, rotation of head pose, eye center direction, and other features by facial landmarks over time. In addition, outstanding research is being conducted in facial recognition, landmark detection, emotion detection, etc. However, there are few works on human interaction and behavioral analysis using facial expressions and body movement. Additionally, it is still an open-topic whether could use additional features such as the dynamics of the eye, head movement, and how those features might be related to the estimated concentration value (9, 10).

In E-learning environments, eye-tracking devices effectively monitor effective attributes such as concentration (11). However, most eye-tracking techniques require extra sensors or a calibration process, both of which are expensive and computationally inefficient (12). Instead, our proposed model considers as a low-cost and efficient alternative for eye tracking because it relies only on a single webcam.

Head pose estimation techniques can be divided into two categories: wearable sensors and computer vision techniques. However, wearable sensors are difficult to apply to daily learning environments because of the annoyance caused by wearable sensors and their cost for many students in E-learning environments. On the other hand, most heads pose estimation techniques limited by high computational software and hardware requirements. Even though our proposed method is based on standard computer vision technology, the results are encouraging, and it is lightweight, making it easy to integrate into edge devices. This research exclusively focuses on computer vision techniques, which are more challenging as they must handle variable factors such as facial expressions, lighting conditions, and lens distortion.

Computer vision measures of subtle human behavior, cognition, concentration, or emotion are relatively little researched; being aware of the many limits of expert view is still one of the best accessible criteria for measuring concentration (13). Thus, this research decided to adopt an expert's perspective views for concentration by simultaneously monitoring many kinds of computer vision measures and comparing them to expert perspectives.

The majority of the studies mentioned in the following topic, while they achieved relatively good accuracy in recognizing concentrations, lacked proof to measure the relationship between these measures with the e-learning video concentration.
In this paper, computer vision measures were measured against the expert perspective view. It tries to solve the problem of measuring concentration levels and determine how factors like Head Pose Estimation, Eye Center Direction, and Eye Aspect Ratio influence the student's concentration .Our major goal is to propose an efficient and lightweight model for automatically measuring concentration level, where on-device feature extraction can be performed without the need for high cost and external devices by using a webcam, a computing device, and computer vision technologies to capture and process recorder video for students in E-learning environments.

The next section will review relevant studies in this field. Then will go over our methodology and propose a low-cost concentration level measurement model. Later will show the results of our experiments to show the effectiveness of the proposed model and validate them with an expert perspective. Finally, we will summarize our findings and make some recommendations for future work.

2. RELATED WORK:

In recent years, e-Learning has become a rising global application area with significant development potential (14). Various studies have revealed a significant increase in online courses compared to traditional techniques (15, 16). With the development of e-learning, there is a high demand for technological methods to analyze student behavior or status in online learning. The attitude and status of students are vital in improving the learning effect. Students' status and attitude have been linked to concentration, which can retain the essence of what they do, read, think, or watch. One of the major advantages of E-learning is that students may study at their convenience. Students had to leave their comfort zones for a long time to attend lectures. This environment change leads students to lack concentration. E-learning enables students, by contrast, to choose the best study environment and therefore enhances their understanding. As a result, students like the learning process more than traditional classroom learning (17).

Concentration is one of the most important factors influencing the results of learning. Furthermore, the student's concentration is critical to the effectiveness of the class for learning(18). Concentration and overt participation, positive behavior, and persistence are considered a component of behavioral engagement (19). Concentration is the ability to monitor thoughts while maintaining attention. Also, the ability to concentrate is the ability to maintain the collected thoughts (6). Several computer vision techniques for modeling the student's head pose, head movements, eye state, and emotions are proposed (20-22).

Several traditional approaches and measurements for assessing the concentration level introduced in the literature have their relevance. First, self-reports, for example, Surveys and quizzes, are two possible approaches for keeping track of students' actions/behaviors. On the other hand, these two approaches are inconvenient and lack objectivity because people may not remember exactly what they did (23). Second, an external expert performed an observational study focusing on student behavior analysis; external experts may evaluate concentration in some cases based on live or pre-recorded videos of educational activities (24, 25).

A traditional concentration measuring methods are insufficient in all contexts, and these methods are not scalable. As a result, for the digitally transformed learning environment, automated concentration assessments are necessary. These approaches use affective computing techniques to evaluate different facial signals, head posture, eye direction, and well-known social cues of concentration and distraction. These techniques are sensitive to changes in concentration levels over time. Furthermore, automatic measures can help prompt intervention to reverse the decline in concentration levels (26). With the development of computer vision technology, recording and analyzing students' classroom behaviors in real time is no longer an impossibility (27).

Computer vision can detect students' engagement and concentration levels due to affective computing techniques, low-cost cameras, and their widespread availability in cell phones, tablets, laptops, and even automobiles(28, 29).
Human-computer interaction needs to understand how an individual's behavior changes constantly. It is also important to determine the visual aspects of behavior change. In an e-learning environment, eye and head movement patterns predict how students feel (30). For example, if a student is rotating head frequently, that could indicate different behavior for different contexts. If a student is looking left or right too frequently, instead of looking at the screen, if each rotation period is comparatively too long or exceeds a threshold, it indicates a lack of concentration.

Face detection has proved to be more effective than reliable biometric methods such as fingerprint and iris; the face detection advantages are based on cost and user convenience (31). To extract the details of the face region, there are various face detection algorithms Viola-Jones (32), Local Binary Pattern (LBP), Ada Boost, and Neural Network are some of the most popular face detection algorithms. In addition, Viola and Jones suggested a fast and robust method for face detection that is 15 times quicker than existing techniques and has a 95 % accuracy rate at the time of release(33). The Viola Jones face detection algorithm is used by Kamath, Biswas, and Balasubramania (29) to analyses the input images, then the Histogram of Oriented Gradients (HOG) is used to represent the face for the patch to get the final vector of features. These features have been used to train the instance-weighted Multiple Kernel Learning-Support-Vector Machine (MKL-SVM) to build a model, and then the system's performance was evaluated. They achieved an average accuracy of 43.98 % and a maximum accuracy of 50.77 %.

Accurate Facial landmark detection within facial images is a critical step toward completing several higher-order computer vision tasks, such as facial recognition and facial expression analysis (34). facial landmarks are extracted using Dlib's implementation of Kazemi and Sullivan’s paper (35), where the main advantage is high detection speed (36). In addition, the importance of accurately identifying facial features such as the location points of the eyes, nose, and mouth is important since the following features extraction and classification stages can be more powerful and efficient if the facial feature points are correctly located(37).

One of the body interaction techniques that has been discovered and used as an input modality is the head gesture. The gestural input identified by the yaw, pitch, and rolls of the user's head movement using 3-dimensional (3D) interfaces as an interaction platform (38). UÇAR (8) detect student engagement in an e-learning environment based on head pose estimation and eye aspect ratio, algorithms used solvePnP (39, 40) and EAR (41) . The result shows that student engagement can be detected by machine learning models with 72.4% accuracy.

The application of eye-tracking techniques in e-learning has enabled the estimation of the student's concentration in real-time. It has been discovered that the stress level, the focus of concentration, and problem-solving capacity of students may be effectively estimated using several eye-tracking measures such as fixation duration, gaze point direction, and blink duration (42). Valente et (43) concentrate on combining head pose and eye location information to compute gaze direction when non-frontal faces or off-center pupils are present. In addition to using cured isotopes to detect eye locations, the researchers use a cylindrical head model to estimate head pose and track head movement. The results show that eye estimations and head pose tracking accuracy has increased by 16% to 23% and 14% to 24%, respectively. However, accuracy remains low due to the dynamics of head pose and eye location and the limitations of model-based gaze estimation approaches. Model-based gaze estimation methods frequently rely heavily on specialized hardware, such as a high-resolution camera, infrared illumination, complexity sensors, and a chinrest. However, free head movements may cause model-based approaches to fail frequently.

Hutt et al. (44) proposed using gaze patterns and locating areas of interest (AOIs) to detect mind wandering while students were taking a massive open online course (MOOC). In detecting student mind wandering, exploiting gaze patterns was more effective than locating AOIs Still, their approach Compared to research-grade eye trackers, consumer versions have a lower sample rate, reducing the accuracy of eye-gaze data. Furthermore, because they can't generalize their findings to other lectures, data were obtained in a calm lab setting; for greater ecological validity, they need to investigate more realistic learning situations (e.g., homes or libraries). Krithika (45) uses eye and head rotation to evaluate students' concentration and generate a low concentration alert. The implementation was carried out in MATLAB, with various functions for face detection and Viola Jones feature detection used. However, their system could not track eye center movement and head pose in different angle scenarios but instead focused on determining if the emotion was positive or negative when only if the face was detected. Besides that, when facial features and eyes status are considered, the detection precision is skeptical. Sharma et al. (46) suggested a real-time system for checking students' concentration in an e-learning environment based on face detection. They expressed facial emotions during an online class and automatically adjusted the contents according to the student's concentration level by analyzing their emotions. Then, the emotions are analyzed to determine the final concentration index. The results indicated that the emotions expressed were related to the student's concentration, and they devised three distinct levels of concentration (high, medium, and low). However, different variations in head pose and subtle facial behaviors are difficult to handle with a 2D-based study comparing to our model study in 3D. Zalatelj and Kosir (47, 48) estimated body pose, facial expressions, and gaze using a Kinect sensor and its commercial SDK. Kinect was then used to calculate behavioral signals (such as yawning, taking notes, and so on) and trained a bagged decision tree classifier to estimate observer-rated concentration levels (low, medium, high). SDK. However, several sensors are required in a normal classroom with 20-30 students, potentially bringing additional costs and device synchronization challenges.

Based on previous studies, most of the research area is not fully covered as integrated factors using hybrid methods combining head pose, eye center direction, and eye aspect ratio (EAR) with results validation from an expert perspective view. In addition, a threshold model is employed to define for each attribute the slandered value. This work defines standard values for each behavioral model attribute rather than manually selecting the reference frame. A threshold model controls the deviation from the standard values. If the deviation for a certain period is lower than the corresponding threshold value, a low concentration level event is detected in the proposed model. Therefore, this research aims to measure the level of concentration as judged by an expert perspective observation. The basic argument is that because teachers modify their teaching behavior based on student concentration, automating concentration levels via computer vision is significant for various educational applications. Furthermore, experts must generate concentration judgments based on various factors such as the student's head pose, eye center direction, and eye aspect ratio.

3. METHODOLOGY

The proposed model focuses on measuring the student's concentration level by continuously monitoring head pose rotation, eye center direction and eye aspect ratio (EAR) based on recorded videos watching the learning material. First, the model's main components were defined. Then, using an input

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

video, our model iteratively evaluates each of these components to measure the appropriate concentration level. Figure 1 shows the proposed model design with its major components.

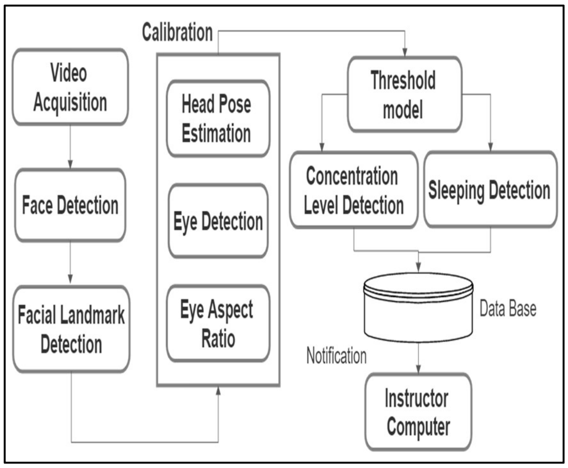


Figure 1: Framework for concentration level recognition model.

3.1 Face Detection

Haar Cascade Classifier implementation from OpenCV is used to detect the face, which Violas Jones has suggested. To detect the face, this model allows images to be processed quickly and achieves high average detection. First, a new image representation called the "integrated lines" is introduced to the face, and a small number of critical images is selected. The model then integrates classifiers into a "cascade" that rejects the image background areas, uses more computation on objects such as regions, and eliminates non-facial regions. Finally, the output will show a bounding box for the face detected.

3.2 Facial Landmark Detection

After face detection, facial landmarks are extracted, the location of 68 (x, y)-landmark coordinates corresponding to facial features is estimated using a pre-trained landmark detection model from the dlib library (i.e., implementation of (35). Recognition of the concentration level does not require all facial landmarks. Furthermore, OpenCV does not always require computing the deviation for all 68 landmarks. As a result, only landmark points evaluated are needed for our model. In this case, only the required points are used. Figure 2 shows the landmarks utilized to detect the head and eye, respectively. Students' concentration level is best defined by their head pose, eye center direction, and eye aspect ratio (EAR). After detecting all 68 landmarks, 12 landmarks were selected for eyes detection and 14 for head detection for our proposed model. Instead of all 68 landmarks, our model uses 26.

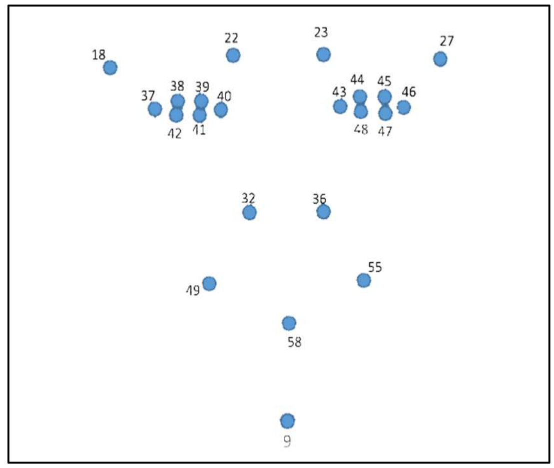


Figure 2: Facial landmarks detection

3.3 Calibration:

One of the main contributions to this work is to estimate a calibration model that identifies the efficient, relevant facial attributes using facial landmarks to monitor students' concentration levels. One of the main contributions to this work is to estimate a calibration model that identifies the efficient, relevant facial attributes using facial landmarks to monitor students' concentration levels. As relevant facial attributes for concentration detection, we have determined three attributes: Head Pose, Eye Center Direction, and Eye Aspect Ratio (EAR). We use the head pose attribute to determine if the student is heading straight to the screen or left or right. Using the eye direction attribute, we can decide if the person is looking straight ahead or in another direction. And the eye ratio (EAR) attribute can determine whether the student blinks frequently and is sleepy. All these essential attributes contribute to student concentration detection and make the model more efficient.

3.3.1 Head Pose Estimation

To estimate head pose, solvePnP function from the OpenCV library's is used. This function used Levenberg-Marquardt optimization to solve the Perspective-n-Point problem. Perspective-n-Point is a problem of estimating the pose of a calibrated camera given a series of n 3D point coordinates in the world and their corresponding 2D projections in the image. 2D facial landmark coordinates are used 14 face landmarks and corresponding approximate

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3D point coordinates. Then, using the solvePnP function, the corresponding rotation and translation vectors were obtained. Using Rodrigues' rotation formula (49), the rotation vector was transformed to a rotation matrix. Finally, the rotation matrix was used to compute Euler angles (pitch, yaw, and roll).

After detecting the Euler angles, our model determines whether the head is rotating left or right by using an appropriate threshold value. When the yaw angle remains between -20 and +20 threshold value, the head remains in the center. A yaw angle more than 20 indicates that the head is facing left, whereas a yaw angle less than 20 means that the head is facing right. Similarly, when the pitch angle is maintained between -20 and +20 threshold value, the head remains in the center. When the roll angle keeps between -25 and +25 threshold value, the head remains in the center. When the head is turned away from the center for more than 10 seconds, then it is determined to have low concentration. In Figure 3, Euler angles are shown.

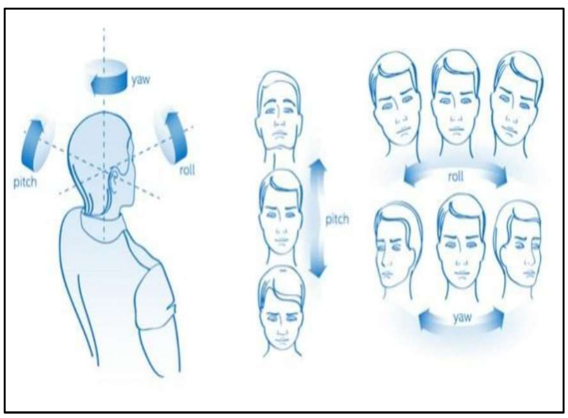


Figure 3: Head Movement at Euler Angles (e.g., roll, pitch and yaw).

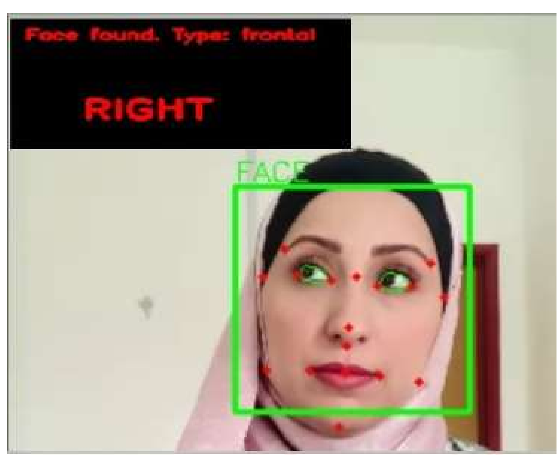
3.3.2 Eye Detection:

Eye detection is done using the Haar Cascade Classifier technique. The first step in determining eye location is to detect faces. Following the detection of the student's face, the next step is eye localization. The eyes are then localized and determined using the 68 coordinated facial landmarks coordinate point mapping. The left eye is accessible through index 37 to 42, and the right eye is accessible through index 43 to 48. The localized eyes will then be displayed using the convex hull method from the OpenCV Library. Next, the eye status monitoring is done using the Blob detection algorithm, which is used to find the image's biggest contour (blob) in the eye region, which means an iris area. The center of this blob is then used as the iris center. After the iris contour is founded, the eye center coordinate can be obtained using moments. It has also been provided in the OpenCV library. The moments will find the center mass of eye contour by counting non-zero-pixel images or the image with color besides black. The formula to get the x and y coordinate of the eye center is shown in equation 1.

$$X = \frac{M10}{MOO} \quad Y = \frac{M01}{MOO} \qquad (1)$$

Where $M10$ is the first moment in X and M00 in all areas, in the case of concentration, the eye's iris is supposed to stay at the center. Therefore, a student is assumed not to concentrate if his eye iris directs looking away from the screen to the left or right direction for a certain period which could be negative indicators of the concentration level. In our model, eyes direction is detected using 12 landmarks for both eyes where the points number 36, 37, 38, 39, 40, 41 and for the left eye frame while the 42, 43, 44, 45, 46, 47 give us the right-eye frame.

To determine the direction of eye movements, thresholds are set. First, the eye area is divided into two sections for each eye. Then count the number of non-zero pixels in and half-section to determine the percentage of white area. Eyes are looking to the right if the average white area of the first half is larger than the average white region of the second half for all eye areas. Eyes are looking to the left if the average white area of the second half is larger than the average white region of the first half of all eye areas. When the eyes are directed to the left or right for more than 10 seconds, the case is identified as not looking forward. When none of the above cases are found, eyes are looking at the center. Figure 4 (a), (b) and (c). shows an eye movement direction with a white region on different cases.



ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

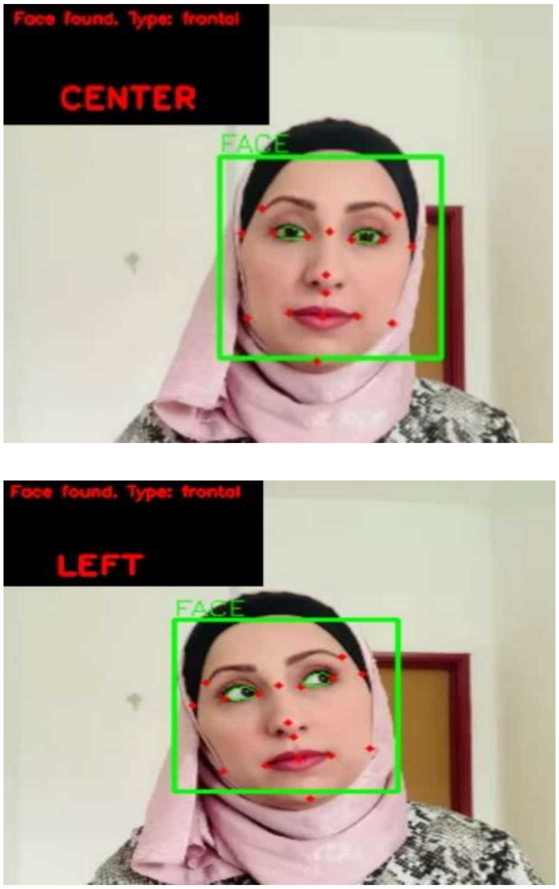


Figure 4: (a), (b) and (c) Eye Center Direction with white region on different cases.

3.3.3 Eye Aspect Ratio (EAR):

Two-dimensional coordinates of six discrete landmarks in each eye were used to measure eye-close or eye-open. P1 and p4 represent the left and right sides of the eye, respectively, while p2 and p3 are two points above the eye, and p5 and p6 are two points below the eye. EAR is only used for the unilateral eye. The range of head ion exceeds 30 since a significant degree of head rotation can result in an eye not being observed; equation 2 was used to calculate EAR.

$$\text{EAR} = \frac{||P2 - P6|| + ||P3 - P5||}{2||P1 - P4||} \qquad (2)$$

As long as the eyes are open, the computed average eye aspect ratio of both eyes remains nearly constant. Therefore, an eye-blink is counted when the average EAR is less than 0.25. Also, if the average EAR is less than our model's threshold for 30 seconds or more, the student is considered sleeping.

3.4 Threshold model for concentration level measurement constraints:

Another important contribution of this research is the identification of a threshold model that indicates the change in three major attributes used in the calibration process. Any deviation from standard values contributes to detecting student concentration behavior. If these deviations continue for a specified period, the case is considered to have a low concentration level. As a result, deviations from the standard threshold value can help detect low concentration behavior. Any low concentration level event detected by the model reduces the concentration level of the corresponding student. OpenCV is used to detect facial landmarks, and our concentration levels are measured by using these landmarks. There has been no manual annotation. Furthermore, our proposed model does not utilize any additional classifier or training. We employed calibration as a classifier and a Threshold model to identify the concentration level from standard behavior.

3.5 Concentration Level Detection:

Student's behavior can be detected at any time using the Threshold model and calibration process. We defined Concentration Level Recognition to correlate student behavior and the calibration process combining head pose, eye center direction, and eye aspect ratio. Our model continuously monitors the input video frame. It also detects required attributes and calculates deviation. In addition, this model detects and saves in a database the estimated values for all attributes with the estimated concentration level for the same values. then extracted and viewed by the instructor. If more than the threshold for students are low concentration, a feedback notification will be sent to the instructor’s computer display. Based on the notification, the instructor will adjust the content style and instructional strategy to the students. This technique will enhance the student’s concentration and make the session more interesting and updated.

3.5.1 High Concentration Level:

To detect high concentration, make sure the head and eyes are in the correct position before proceeding. The head pose, the center of the eyes, and the eye aspect ratio should be determining by a head pose and the eyes facing the screen. Our model assumes that if they do not face the screen correctly or their eye aspect ratio is less than the threshold value (when an eye blink lasts more than a second, it is considered an eye closure), their concentration level is low. The number of frames for which the appropriate level of concentration is not detected is monitored. If the number of frames with low-level concentrations exceeds a predefined threshold, the student is not interested in the topic. For determining the head pose and eye movement direction status, the head pose (-20 degree <yaw angle < +20 degree) and the eye movement direction should be facing the screen (looking to center) where asymmetric average white region over time, and Eye Aspect Ratio value is more than the threshold (open eye), i.e., EAR >=0.25 where these are the standard values for concentration.

3.5.2 Low concentration level:

For determining the head pose status and eye movement direction for low concentration level, it comes in three steps:

1- Head Pose Rotation looking at the different direction (Left or Right) for more than 5 seconds according to these values
   Yaw Angle (-20 degree < yaw angle > +20 degree)
   Pitch Angle (-20 degree < Pitch Angle >+20 degree)
   Roll Angle (-25 degree < Roll Angle > +25 degree)

2- Eye Aspect Ratio <0.25 (Sleeping detection) for more than 30 seconds.

3- Eye Movement Direction for more than 5 seconds (average white region of first half≠ average white region of second half) Looking at right or left.

4 Result and Discussion:

For the performance evaluation experiment, an E-learning video of 5 minutes long was used with a frame rate of 14fps (i.e., frame per second) and 640×480 resolution. Then the video is encoded and processed with Python. Next, it is saved using the AVI format, and in this video, it has one recorded student's learning scenario for e-learning. Finally, the model is tested on a dataset of 5 students for both scenarios (high, low) with 10 total videos. Without loss of generality, participants include males and females, with different illumination conditions and with different types of concentration (high, low); the participants sit comfortably in front of their computers.

During watching the recorded videos, an expert from University Sultan Zainal Abidin's Faculty of Informatics and Computing observed student behavior` during recorded videos for each student to create the known and actual status for concentration level and classified it into two categories: high and low concentration levels. Then The video is benchmarked with our model to determine the concentration level. After that, compare the result with expert observations then calculate the accuracy for our model. Equation 3 was applied to calculate the accuracy of the concentration level detection percentage. Here, accuracy is calculated by

$$\text{Accuracy} = \left(\frac{\text{Total number model results exact with the expert}}{10}\right)100 \quad (3)$$

Based on this experiment, 80% of the result were similarly matched with the expert's opinion. Thus, our model provides good outcomes. It measures head rotation and eye center movements by detecting and tracking the student's face in real time. According to the results, our model was able to measure the state of the student concentration level in real time.

5.CONCLUSIONS

In the e-learning environment, there is an increasing demand to measure student concentration. As a result, Therefore, this study proposes a model for measuring a student's concentration level in an e-learning environment and also providing a feedback mechanism to improve E-learning supports for better content delivery. Specifically, obtain the information through the webcam of the student's head pose rotation and eye-tracking technique to discriminate the behavior of the students during the E-learning classes. A feedback notification will be sent to the instructor's computer display. Based on the notification, the instructor will adjust the content style and instructional strategy to the students. First, the model detects and tracks the face, head, and eyes. Then, it analyses the state of the student's head pose rotation, eye center direction, and eye aspect ratio (EAR). Then compute the difference between the standard threshold value to determines the student's level: high concentration or low concentration. After that, an observation from an expert perspective to determine concentration levels for each student. Finding that our model results matched 80% of the expert decision. This technique will enhance the student's concentration and make the session more interesting and updated.

This study has some limitations. If several faces appear in the window, the webcam can detect more student's faces, and unwanted output may appear

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

because of the different states of concentration for different faces. As a result, only one student's face must ensure that it is visible to the webcam. As potential future work, this approach could readily be expanded to generate real-time alerts for remote live monitoring in E-learning classes.

REFRENCES:

[1] Essalmi F, Ayed LJB, Jemni M, Graf S. Generalized metrics for the analysis of E-learning personalization strategies. Computers in Human Behavior. 2015;48:310-22.

[2] Essalmi F, Ben Ayed LJ, Jemni M, Kinshuk, Graf S. Automating the E-learning Personalization. Springer Berlin Heidelberg; 2013. p. 342-9.

[3] Bhardwaj P, Gupta P, Panwar H, Siddiqui MK, Morales-Menendez R, Bhaik A. Application of Deep Learning on Student Engagement in e-learning environments. Computers & Electrical Engineering. 2021;93:107277.

[4] Li Y-Y, Hung Y-P, editors. Feature Fusion of Face and Body for Engagement Intensity Detection. 2019 IEEE International Conference on Image Processing (ICIP); 2019: IEEE.

[5] Whitehill J, Serpell Z, Lin Y-C, Foster A, Movellan JR. The faces of engagement: Automatic recognition of student engagementfrom facial expressions. IEEE Transactions on Affective Computing. 2014;5(1):86-98.

[6] Mejbri R, Essalmi F, Qaffas AA, editors. Towards modeling the learner's concentration in e-learning environments. 2019 7th International conference on ICT & Accessibility (ICTA); 2019: IEEE.

[7] Yue J, Tian F, Chao K-M, Shah N, Li L, Chen Y, et al. Recognizing multidimensional engagement of e-learners based on multi-channel data in e-learning environment. IEEE Access. 2019;7:149554-67.

[8] M. U. UÇAR1 EO. Detecting Student Engagement in e-learning Environment Based on Head Pose Estimation and Eye Aspect Ratio. International Conference on Advanced Technologies, Computer Engineering and Science 2019;ICATCES 2019.

[9] Khan R, Debnath R. Human distraction detection from video stream using artificial emotional intelligence. International Journal of Image, Graphics and Signal Processing. 2020;12(2):19.

[10] Peng S, Chen L, Gao C, Tong RJ, editors. Predicting Students' Attention Level with Interpretable Facial and Head Dynamic Features in an Online Tutoring System (Student Abstract). Proceedings of the AAAI Conference on Artificial Intelligence; 2020.

[11] Horvat M, Jagušt T, editors. Emerging Opportunities for Education in the Time of COVID-19-Adaptive e-Learning Intelligent Agent Based on Assessment of Emotion and Attention. Central European Conference on Information and Intelligent Systems; 2020: Faculty of Organization and Informatics Varazdin.

[12] Bruno A, Moore M, Zhang J, Lancette S, Ward VP, Chang J. Toward a head movement-based system for multilayer digital content exploration. Computer Animation and Virtual Worlds. 2021;32(2):e1980.

[13] Vanneste P, Oramas J, Verelst T, Tuytelaars T, Raes A, Depaepe F, et al. Computer Vision and Human Behaviour, Emotion and Cognition Detection: A Use Case on Student Engagement. Mathematics. 2021;9(3):287.

[14] Yücel Z, Koyama S, Monden A, Sasakura M. Estimating Level of Engagement from Ocular Landmarks. International Journal of Human–Computer Interaction. 2020;36(16):1527-39.

[15] Sharin AN. E-learning During Covid-19 A Review of Literature. Jurnal Pengajian Media Malaysia. 2021;23(1).

[16] Wang Y, Kotha A, Hong P-h, Qiu M, editors. Automated student engagement monitoring and evaluation during learning in the wild. 2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom); 2020: IEEE.

[17] IvyPanda. Impact of Online Courses on Education 2019 [Bibliography]. Available from: https://ivypanda.com/essays/impact-of-online-courses-on-education/.

[18] Ahn H, Nam S, Song G. Application of bio-signal measurement to identify learning concentration in e-learning environment. the Korean association of computer education. 2012;16(2):125-30.

[19] Sümer Ö. Multimodal Visual Sensing: Automated Estimation of Engagement: Eberhard Karls Universität Tübingen; 2021.

[20] Fan Z, Xu J, Liu W, Liu F, Cheng W, editors. Kinect-based dynamic head pose recognition in online courses. 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC); 2016: IEEE.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[21] MacInnes JJ, Iqbal S, Pearson J, Johnson EN. Wearable Eye-tracking for Research: Automated dynamic gaze mapping and accuracy/precision comparisons across devices. bioRxiv. 2018:299925.
[22] Correa E, Jonker A, Ozo M, Stolk R. Emotion recognition using deep convolutional neural networks. Tech Report IN4015. 2016.
[23] Grafsgaard JF, Fulton RM, Boyer KE, Wiebe EN, Lester JC, editors. Multimodal analysis of the implicit affective channel in computer-mediated textual communication. Proceedings of the 14th ACM international conference on Multimodal interaction; 2012.
[24] Ocumpaugh J. Baker-Rodrigo Observation Method Protocol (BROMP) 1.0 Training Manual version 1.0 Oct. 17, 2012. 2012.
[25] Kapoor A, Picard RW, editors. Multimodal affect recognition in learning environments. Proceedings of the 13th annual ACM international conference on Multimedia; 2005.
[26] D'Mello S, Dieterle E, Duckworth A. Advanced, analytic, automated (AAA) measurement of engagement during learning. Educational psychologist. 2017;52(2):104-23.
[27] Ngoc Anh B, Tung Son N, Truong Lam P, Le Chi P, Huu Tuan N, Cong Dat N, et al. A computer-vision based application for student behavior monitoring in classroom. Applied Sciences. 2019;9(22):4729.
[28] Monkaresi H, Bosch N, Calvo RA, D'Mello SK. Automated detection of engagement using video-based estimation of facial expressions and heart rate. IEEE Transactions on Affective Computing. 2016;8(1):15-28.
[29] Kamath A, Biswas A, Balasubramanian V, editors. A crowdsourced approach to student engagement recognition in e-learning environments. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV); 2016: IEEE.
[30] Talele AK, Chourasia B. STUDENT'S EMOTIONS IDENTIFICATION USING CNN.
[31] Sufyanu Z, Mohamad FS, Yusuf AA, Mamat MB. Enhanced Face Recognition Using Discrete Cosine Transform. Engineering Letters. 2016;24(1).
[32] Viola P, Jones M, editors. Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition CVPR 2001; 2001: Ieee.
[33] Kumar A, Kaur A, Kumar M. Face detection techniques: a review. Artificial Intelligence Review. 2019;52(2):927-48.
[34] Johnston B, de Chazal P. A review of image-based automatic facial landmark identification techniques. EURASIP Journal on Image and Video Processing. 2018;2018(1):1-23.
[35] Kazemi V, Sullivan J, editors. One millisecond face alignment with an ensemble of regression trees. Proceedings of the IEEE conference on computer vision and pattern recognition; 2014.
[36] Khabarlak K, Koriashkina L. Fast Facial Landmark Detection and Applications: A Survey. arXiv preprint arXiv:210110808. 2021.
[37] Iqtait M, Mohamad F, Mamat M, editors. Feature extraction for face recognition via active shape model (ASM) and active appearance model (AAM). IOP Conference Series: Materials Science and Engineering; 2018: IOP Publishing.
[38] Zolkefly NN, Ismail I, Safei S, Shamsuddin SNW, Arsad MAM. Head gesture recognition and interaction techniques in virtual reality: a review. International Journal of Engineering & Technology. 2018;7(4.31):437-40.
[39] Moré JJ. The Levenberg-Marquardt algorithm: implementation and theory. Numerical analysis: Springer; 1978. p. 105-16.
[40] Mallick S. Head pose estimation using OpenCV and Dlib. 2016.
[41] Soukupova T, Cech J, editors. Eye blink detection using facial landmarks. 21st computer vision winter workshop, Rimske Toplice, Slovenia; 2016.
[42] Wang C-C, Hung JC, Chen S-N, Chang H-P. Tracking students' visual attention on manga-based interactive e-book while reading: an eye-movement approach. Multimedia Tools and Applications. 2019;78(4):4813-34.
[43] Valenti R, Sebe N, Gevers T. Combining head pose and eye location information for gaze estimation. IEEE Transactions on Image Processing. 2011;21(2):802-15.
[44] Hutt S, Hardey J, Bixler R, Stewart A, Risko E, D'Mello SK. Gaze-Based Detection of Mind Wandering during Lecture Viewing. International Educational Data Mining Society. 2017.
[45] Krithika L, GG LP. Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric. Procedia Computer Science. 2016;85:767-76.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[46] Sharma P, Esengönül M, Khanal SR, Khanal TT, Filipe V, Reis MJ, editors. Student concentration evaluation index in an e-learning context using facial emotion analysis. International Conference on Technology and Innovation in Learning, Teaching and Education; 2018: Springer.

[47] Zaletelj J, Košir A. Predicting students' attention in the classroom from Kinect facial and body features. EURASIP journal on image and video processing. 2017;2017(1):1-12.

[48] Zaletelj J, editor Estimation of students' attention in the classroom from kinect features. Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis; 2017: IEEE.

[49] Sorgi L, editor Two-view geometry estimation using the rodrigues rotation formula. 2011 18th IEEE International Conference on Image Processing; 2011: IEEE.