iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** V **Month of publication:** May 2021

DOI: https://doi.org/10.22214/ijraset.2021.34707

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue V May 2021- Available at www.ijraset.com

A Study on Automatic Attention Span Detection of Students

Aditya Parashar[1], Hritwik Biyani[2], Sanchita Biswas[3], Chudaman Sukte[3], Akshay Saigal[3]
[1, 2, 3]*Department of Computer Science and Technology, Dr. Vishwanath Karad Maharashtra Institute of Technology World Peace University, Pune, India*
[4]*Professor, Department of Computer Science and Technology, Dr. Vishwanath Karad Maharashtra Institute of Technology World Peace University, Pune, India*
[5]*Solutions Consultant, ZS Associates India Pvt. Ltd.*

Abstract: Due to the arrival of the coronavirus, the education system has completely transformed. In the wake of this medical emergency and keeping the students' safety in mind with their academic concern, the schools and colleges have endorsed online classes. In online lectures, the teacher is unable to track attentiveness of the students. Academic institutions may not have the human resources or time required to manually analyze each student's video feed. The task of manually analyzing attention of students can be solved by automating it using Computer Vision and Machine Learning, by creating an application that can track the attention levels of each student. An automated solution would be a much cheaper and easier solution to this problem, along with the fact that it can eventually be applied to self-paced online courses as well.
Keywords: Computer Vision, Machine Learning, Deep Learning, Attention Span, Gaze Detection, Pose Detection

I. INTRODUCTION

A. *Attention Span*

Attention Span is defined as the duration that one can focus on a specific task. Distraction or loss of focus is a commonly occurring experience for many students. While in a classroom setting, a skilled teacher may be able to react to students' loss of focus, in online classroom environments/MOOCs such intervention is not very practical.

Student attentiveness measures the time and effort that students devote to studying and other curricular activities. It also involves the planning of resources/material for education that promote learning. Furthermore, the enormous number of attendees who enroll in MOOC education along with the fact that these courses are often self-paced, rules out using people to track engagement levels. Therefore, it is vital to design and develop automated engagement tracking methods based on rapidly evolving computer vision (CV) and machine learning (ML) technologies.

B. *Machine Learning and Computer Vision*

Computer Vision is used in programming to enable machines to gain high-level understanding from digital images or videos. It has a variety of applications from image reconstruction to cosmetic filters in social media apps. One of the fundamental topics in computer vision is about object detection. Essentially, it is about locating a given set of objects (in this case students) in an image and a variety of approaches have been tested on it. Object detection is also essential in surveillance for tracking uncharacteristic or atypical activities. Machine Learning is equipped to provide a reliable, noninvasive, and scalable way of monitoring student attention span within the domain of online education.

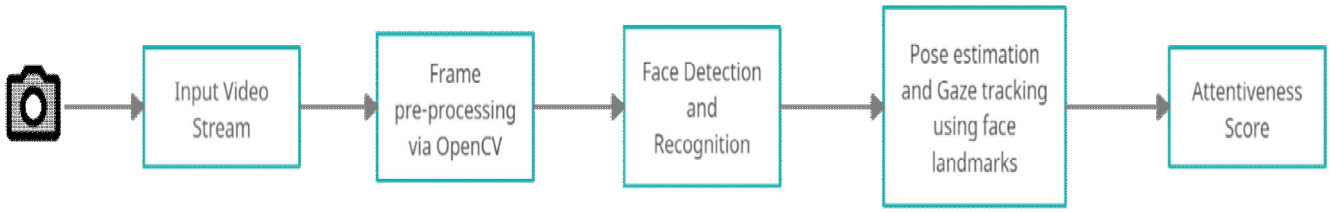


Figure 1 above, is a block diagram of a general Computer Vision based attention span tracking system.

II. AN OVERVIEW OF CURRENT STATE OF ATTENTION SPAN DETECTION OF STUDENTS

An easy way to comply with IJRASET paper formatting requirements is to use this document as a template and simply type your text into it.

A. Involvement of Machine Learning and Computer Vision

Studies show a direct correlation between the human body posture and their respective attention state. One such study [1] explores a method in which the system takes in a series of images of the student, extracts pose-related key-points using Open Pose, and uses a simple CNN model to classify the image into one of five prominent postures (like Attentiveness, Head rested, leaning back, Writing and Not looking at the screen. The model takes care of momentary outliers by performing frame average sampling. The paper also uses a facial emotion recognition system to classify the emotions of the student which are mapped to probable cause (like sadness or disgust can be attributed to personal distractions, laughter or boredom can be attributed to social media and so on). In combination with the above methods, it also employs a drowsiness detection system by calculating the eye aspect ratio.

This paper [2] focuses on detecting the attention state of the student by calculating the drowsiness of the student. It uses the Lucas-Kanade tracker to keep a track of where the eyes of the student are in the frame and generates an ROI covering both eyes. It then tracks the intensity of pixels in the ROI as the attention metric (higher intensity meaning fully opened eyes, lower intensity meaning half-opened/closed eyes). It takes a sliding window median of the value over multiple frames to eliminate false classification due to blinking. It has an active mode with an alerting system and a passive mode without alerts for content evaluation. For content evaluation, the attention metric is calculated between logically placed bookmarks in the video.

The proposed system [3] employs multiple modules in an effort to quantify students' attentiveness. It first analyses an image of the classroom and calibrates the camera parameters to maintain the uniformity of intensity of the images i.e. change the exposure of the image to deal with different lighting conditions. The system then employs a multi-task cascaded convolutional neural network (MTCNN) to detect the face and identify facial landmarks. It uses these facial landmarks to estimate the head pose and subsequently gaze direction of the students. To track each student in the classroom, the system recognizes faces by using a ResNet-101 layers CNN. The system also uses OpenPose to track the students' pose and measures the degree of affinities between multiple students' body part motion (higher affinities imply more synchronized motion of multiple students indicating student engagement).

The work done by the authors of this paper [4] includes creating a dataset for eye-tracking with around 13 users watching MOOC related videos in a with self-reported mind-wandering feedback. Video and gaze data recorded by a professional eye tracker along with a webcam is also included. They also evaluated the automatic detection of loss of focus based on the gaze data - (i) tracked using a professional eye-tracking device (ii) using a standard webcam and processed by an open-source gaze processing library. They evaluate and weigh both approaches against each other and conclude that the performance of a standard webcam and a specialized device for gaze and attention tracking are on par with each other.

The system outlined in this paper [5] utilizes Facial Expression Recognition by way of Ensemble Learning along with CNN. The system tries to recognize human expressions and assigns the respective state the student is in(focused/distracted). The face of a student is extracted from each frame by a Haar Cascade algorithm. If the expression is 'neutral' for a pre-defined duration, the student is assumed to be 'focused'. The model used is an ensemble of two similar models applied to the original image and a mirrored image. It achieves an accuracy of 70% on the FER-2013 dataset. The system is designed to use the model and send appropriate notifications to effectively monitor a classroom.

The proposed agent in this paper [6] uses an ensemble of four models to track student engagement. It uses OpenFace to extract a 31-dimensional feature set for each frame which contains eye gaze movement, head movement and facial action units(AUs) features. The ensemble comprises three cluster-based conventional models and one attention-based neural network model(bi-directional LSTM), boosted with heuristic posture-based rules(like hand fidgeting and body fidgeting) that it extracts from the frame using OpenPose. The attention weights are ultimately used to generate insights into time periods where the student is not engaged, thus assisting in better feedback for the students and the instructors.

This paper [7] proposes a method using computer vision and machine learning to monitor and detect the drowsiness of a person in real-time through a web camera. The detection is done by evaluating the eye aspect ratio by classification with a support vector machine. In industries, where operating specific roles for example control rooms, this technique will be helpful at capturing the fatigue-related characteristics using video processing.

First, the eye blinks are computed using a metric called (EAR) introduced in [8]. Post which, an optimized SVM [9] was applied for classification. And finally, the decision rule was adopted to detect the fatigue of the operator.

[10] proposes another methodology of a framework to identify the commitment level of the understudies. The framework uses just the data given by the web camera present in the PC. The system classifies the engagement metrics of participants, into three broad categories 'very engagement,' nominally engaged' and 'no engagement at all'. The detector comprises three submodules, the Viola & Jones algorithm[11],a CNN model and a concentration index.

This paper [12] was a feasibility study of estimating the head pose and emotion of participants using a standard laptop. An application was built using the OpenCv and SVM was used to detect the facial landmarks on the face. The FER2013 model was used as the dataset. The face is first located, taken as an individual image, and finally converted to greyscale [13].

This paper [14] conducted an exhaustive survey comparing both, a manual annotation questionnaire to those that were recorded through machine learning. Instruments developed by Helmke and Renkl(1992) [15] and Hommel(2021)[16] were used as a basis. For the machine learning-based approach, first, the faces were detected (Zhang et at.2 017)[17] and mapped those faces to the students to observe them. The main modules of the detection were head pose, gaze direction and facial expressions using a facial action coding system. (FACS: Ekman and Friesen 2978) [18]

Both the machine learning and manual annotation method were tested, and eventually, the study found that the manual rating was the most effective and could account for the variation in the test results.

The system [19] presented in this system to derive the facial landmarks utilizes Dlib. Once the facial landmarks are determined, a number of other characteristics are measured such as the Blink rate, and a number of other features are drawn from the landmarks such as how far the user is from the screen, duration of blinks, the EAR, the size of the eye etc. This system also comprises feedback for the users to know about their state of engagement. A probabilistic method was used to assess the engagement of the users and later it was proposed to be integrated into the system. This approach can be applied to each feature individually or as a group to assess the level of engagement.

The paper [20] proposes the idea of a framework where the processing unit consists of three significant modules which are facial recognition, system for motion analysis, and behaviour analysis. The body motion analysis is used on the students with recognized face and their activity will be recorded into the database such as entering or leaving the classroom. At the completion of the lecture, the gathered data would be used to calculate their overall attendance. Finally, students' behaviour and performance are evaluated.

The paper [21] talks about an idea of an AI-based - Deep learning model. This model is trained by using pre-existing available generic facial expression data. The first step deals with CNN i.e. Convolutional Neural Network is trained on the dataset named FER-2013 (Facial Expression Recognition Challenge 2013) to provide a rich facial representation model, achieving state-of-the-art performance. Further to this, the curated model is then made to run to initialize the engagement recognition model, which is developed and designed using another CNN, this uses the newly collected dataset in the engagement recognition.

The paper [22] aims to estimate the level of attention of a student in a lecture using the principles of behavioural cues which the experience lecturers use to curate their presentation to the class. The core objective of this thesis was to identify the behaviour of students using: Curating a theoretical framework that would explain non-verbal behaviour. Formulating a set of unobtrusive metrics which would capture it, Design and develop a system for experimental validation of ideas on the collected data set.

The paper [23] aims around building an automatic system for teachers in a classroom-based environment which helps the teachers to capture and summarize student's behaviour. The system makes a recording of the complete lecture session and then figures out the attentiveness of the students. The outcome of the research shows that their built system is very flexible and accurate in terms of attentiveness detection. The system makes uses of a. Face Detection and Face Alignment, b. Face Embedding and Recognition, c. Gaze Estimation, d. Position Estimation.

The paper [24] discusses a tool - ClassACT which stands for e Classroom Attentiveness Classification Tool which is a system with the purpose of detecting and monitoring the attention levels of students in various instructional phases under a given learning environment which could be lectures, group work, assessments etc. This takes place by gathering information on the students via the sensors deployed on a device. This information is then processed in form of data and then is classified using a classifier. The system helps to identify attentive and inattentive behaviour.

The paper [25] consists of an exhaustive study using two famous browser-dependent software frameworks, majorly for detection of gaze and eye - tracking.js as well as WebGazer.js. For the study, the researchers compared three crucial and potential possessing technical solution, which are - Making the use of a top-end and high-quality Tobii eye tracker and then making the use of two software-based solutions based on analysing the visual input stream of a consumer-grade webcam. Here, two open-source libraries were used - Clmtrackr which is used for gaze tracking and tracking.js -the face tracking library.

TABLE I
Features of each Paper

Title -Year	Methodology	Advantages/Disadvantages
Real-time SVM Classification for Drowsiness Detection Using Eye Aspect Ratio [7] – 2018	• Eye blinks are computed using Eye Aspect Ratio (EAR)[8]. • SVM [9] applied for classification. • Decision rule adopted to detect the fatigue of the operator.	This method works well in real-time and can be run along with other programs on a normal computer without slowing down the system. Only relies on eye detection and doesn't incorporate other facial landmarks, thus the accuracy is not at its best as even false positives are detected for eg when a person simply yawns but is not sleepy.
Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning[10] - 2019	• Design and develop a system for experimental validation of ideas on the collected data set. • The detector comprises three submodules:- • Viola & Jones algorithm. • CNN model Concentration index using the confidence score of emotion analysis.	The results showed that students with a higher concentration index scored well in the tests. The dataset used was very limited and showed low levels of accuracy in cases when the students face was covered with the hand as the concentration index was computed as low even in cases when the student was paying attention.
Estimation of Students' Level of Engagement from Facial Landmarks[19] - 2020	'InceptionNet', 'C3D', and Long-Term Recurrent Convolutional Network (LRCN), C3D, InceptionNet – were the three CNN models used to detect engagement.	This system also comprises feedback for the users to know about their state of engagement. This approach can be applied to each feature individually or as a group to assess the level of engagement.
Attentive or Not? Toward a Machine Learning Approach to Assessing Students' Visible Engagement in Classroom Instruction[14] - 2021	Main Modules - Head Pose, Tracking direction of Gaze and recognition of facial expression using FACS.	Both the machine learning and manual annotation method were tested, and eventually, the study found that the manual rating was the most effective and could account for the variation in the test results.
Webcam-based Attention Tracking in Online Learning[25] – 2018	• Comparison of three popular solutions: - Tobii X professional eye tracker - Using Clmtracker with webcam. - Using tracking.js with webcam.	The research has taken a wide area of use cases to consider all face-hit and face-miss scenario which one might encounter in real-life scenarios However, the proposed idea does not prove to be robust in case of high large-scale scenarios.
Classroom Attentiveness Classification Tool (ClassACT): The System Introduction[24] - 2017	• Three classifiers that are used in this paper for comparison: 1. MLP 2. SVM 3. PSVM	The system is well equipped with webcam - audio input functionality to collect the required data to be processed, however misses a crucial feature - gaze detection of the students which might be very vital to identify the attentiveness.

Automated classroom monitoring with connected visioning system[20] - 2017	A face recognition algorithm is used for detection and then to determine the face of the student, then we deal with the body movement analysis their attendance is evaluated using the recorded activity.	This framework uses the specific techniques for detection of face along with utilizing upper body detector as well as full-body detector which increases the accuracy of the system considerably. However, the system in an IoT based framework is highly influenced by the classroom environment factors and tremendously affects the performance of the same.
Camera-based Estimation of Student's Attention in Class[22] - 2015	• Curating a theoretical framework that would explain non-verbal behaviour. Formulating a set of unobtrusive metrics which would capture it and Design and develop a system for experimental validation of ideas on the collected data set.	One of the major limitations of this study was the presence of inconsistent format of the lecture being analysed since it's an ever-varying technique and can not be limited.
Attention Analysis in E-Learning Environment using a Simple Web Camera[2] - 2012	Detection of the attention state of the student by calculating the drowsiness of the student.	It is real-time as it has a very quick execution time (frame processing measured in 10-36 milliseconds). It is scale-invariant as well as rotation invariant. It is also tolerant to blink-related variations. Although, reflective spectacles and unfavourable lighting conditions may cause the system to misclassify.
Scalable Mind-Wandering Detection for MOOCs: A Webcam-Based Approach[4] - 2017	Evaluate and weigh both approaches against each other and conclude that the performance of a webcam of laptop and a special device for gaze along with attention tracking are on par with each other.	It takes into account where the person is looking at on the screen (speaker's face, slides, subtitles, etc.), it establishes a scalable and low-cost alternative to specialized eye-tracking hardware. It has a very limited pool of participants and the number of evaluated MOOC videos is very small. It also implements simple classifiers (Naive Bayes/Linear SVM) with lower accuracy.
An Ensemble Model Using Face and Body Tracking for Engagement Detection[6] - 2018	The proposed agent in this paper three clustering orthodox models and one bi-directional LSTM neural network model	The system has a very low MSE (~0.04) and high accuracy. The Bi-directional LSTM ensures that both forward and backward information flow in time is maintained. On the other hand, the dataset is not large enough, so the model faces overfitting issues, and the model is computationally expensive.
Monitoring Students' Attention in a Classroom Through Computer Vision[3] - 2018	Uses MTCNN to detect facial landmarks, uses CNN to classify facial expressions, and incorporates ResNet-101 layers CNN for facial recognition	The agent aggregates multiple states of the art tracking mechanisms to increase the overall accuracy of the attention measure. It can also handle unfavourable lighting conditions. It doesn't work with different targets i.e. students don't usually look into the camera in classrooms, they look at the instructor.

III. CONCLUSION

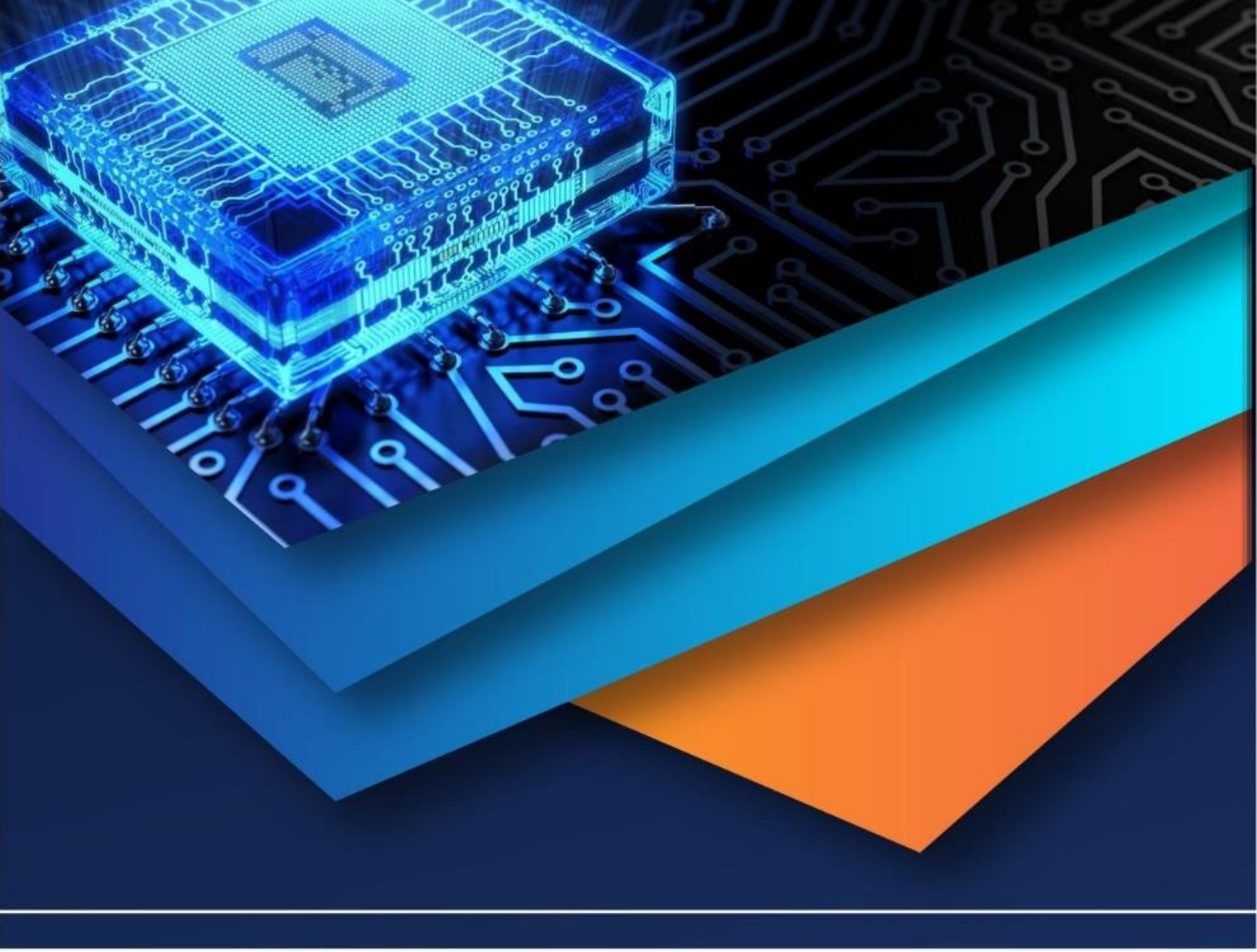
In this paper, we surveyed and outlined the current stage of interdisciplinary studies and the challenges they face before bringing about desired outcomes. We carefully analyze the building blocks of computer vision-based attentiveness tracking mechanisms, which is the measure of the time students study and the effort that they put in. Researchers working separately in different domains comprise most of the solutions. This had always hindered the research community from discovering the best possible approaches which usually involved interdisciplinary research. With this paper, we highlighted how experts from different domains can come together to develop an optimal solution to the problem. The findings presented in this paper give serious insight that provides background for computer scientists to work with the online education culture. We conclude that an all-inclusive non-invasive solution would involve multiple independent techniques and mechanisms of attention tracking as described in this paper, working together to provide an analysis of learners' engagement. As a part of future scope, classroom attentiveness detection can be implemented using a combination of pose detection, face detection, drowsiness detection paired with a user-friendly interface for improvement of attentions span detection.

IV. ACKNOWLEDGMENT

We wish to acknowledge Prof. Chudaman Sukte and the Department of Computer Science Education, for guiding us and helping us every step of the way. We also wish to thank Mr. Akshay Saigal in co-operating with us and supporting us in this endeavour.

REFERENCES

[1] A. Revadekar, S. Oak, A. Gadekar and P. Bide, "Gauging attention of students in an e-learning environment," 2020 IEEE 4th Conference on Information & Communication Technology (CICT), 2020, pp. 1-6, doi: 10.1109/CICT51604.2020.9312048.
[2] S. A. Narayanan, M. Prasanth, P. Mohan, M. R. Kaimal and K. Bijlani, "Attention analysis in e-learning environment using a simple web camera," 2012 IEEE International Conference on Technology Enhanced Education (ICTEE), 2012, pp. 1-4, doi: 10.1109/ICTEE.2012.6208618.
[3] Canedo, Daniel & Trifan, Alina & Neves, António. (2018). Monitoring Students' Attention in a Classroom Through Computer Vision. 10.1007/978-3-319-94779-2_32.
[4] Zhao, Yue & Lofi, Christoph & Hauff, Claudia. (2017). Scalable Mind-Wandering Detection for MOOCs: A Webcam-Based Approach. 330-344. 10.1007/978-3-319-66610-5_24.R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.
[5] Kolkur, S. et al. "Effective Classroom Monitoring by Facial Expression Recognition and Ensemble Learning." (2019).
[6] Chang, Cheng & Zhang, Cheng & Chen, Lei & Liu, Yang. (2018). An Ensemble Model Using Face and Body Tracking for Engagement Detection. 616-622. 10.1145/3242969.3264986.
[7] Souto Maior, Caio & Moura, Marcio & Santana, João & Nascimento, Lucas & Macedo, July & Lins, Isis & Droguett, Enrique. (2018). Real-time SVM Classification for Drowsiness Detection Using Eye Aspect Ratio.
[8] Soukupová T., Cech J. "Real-Time Eye Blink Detection using Facial Landmarks". 21stComputer Vision Winter Workshop. (2016)
[9] Pedregosa F., Varoquaux G. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research (Vol. 12). (2011)
[10] Sharma, Prabin & Joshi, Shubham & Gautam, Subash & Filipe, Vítor & Reis, Manuel. (2019). Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning.
[11] Viola, P., Jones, M.: 'Rapid object detection using a boosted cascade of simple features', Computer Vision and Pattern Recognition, 2001, pp. 511 -518
[12] Farrell, Conor Cohen et al. "Real Time Detection and Analysis of Facial Features to Measure Student Engagement with Learning Objects." (2019).
[13] [Ekman, P., & Friesen, W. V., 1976], Measuring Facial Movement. Environmental Psychology and Nonverbal Behavior.
[14] Goldberg, Patricia & Sümer, Ömer & Stürmer, Kathleen & Wagner, Wolfgang & Göllner, Richard & Gerjets, Peter & Kasneci, Enkelejda & Trautwein, Ulrich. (2021). Attentive or Not? Toward a Machine Learning Approach to Assessing Students' Visible Engagement in Classroom Instruction. Educational Psychology Review. 33. 10.1007/s10648-019-09514-z.
[15] Helmke, A., & Renkl, A. (1992). Das Münchener Aufmerksamkeitsinventar (MAI): Ein Instrument zur systematischen Verhaltensbeobachtung der Schüleraufmerksamkeit im Unterricht. Diagnostica, 38, 130–141.
[16] Hommel, M. (2012). Aufmerksamkeitstief in Reflexionsphasen- eine Videoanalyse von Planspielunterricht. Wirtschaft und Erziehung, 1-2, 12–18.
[17] Zhang, S., Zhu, X., Lei, Z., Shi, H., Wang, X., & Li, S. Z. (2017). S3FD: single shot scale-invariant face detector. Paper presented at, The IEEE International Conference on Computer Vision (ICCV), Venice, Italy.
[18] Ekman, P., & Friesen, W. V. (1978). Facial action coding system: manual. Palo Alto: Consulting Psychologists Press
[19] Yücel, Zeynep & Koyama, Serina & Monden, Akito & Sasakura, Mariko. (2020). Estimating Level of Engagement from Ocular Landmarks. International Journal of Human-Computer Interaction. 36. 1-13. 10.1080/10447318.2020.1768666.
[20] J. H. Lim, E. Y. Teh, M. H. Geh and C. H. Lim, "Automated classroom monitoring with the connected visioning system," 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017, pp. 386-393, DOI: 10.1109/APSIPA.2017.8282063.
[21] Mohamad Nezami, Omid & Dras, Mark & Hamey, Len & Richards, Deborah & Wan, Stephen & Paris, Cécile. (2020). Automatic Recognition of Student Engagement Using Deep Learning and Facial Expression. 10.1007/978-3-030-46133-1_17.
[22] Raca, Mirko. (2015). Camera-based Estimation of Student's Attention in Class.
[23] Lam, Phan & Chi, Le & Tuan, Nguyen & Dat, Nguyen & Nguyen, Trung & Anh, Bui & Aftab, Muhammad Umar & Tran, Van Dinh & Son, Ngo. (2019). A Computer-Vision Based Application for Student Behavior Monitoring in Classroom. Applied Sciences. 9. 4729. 10.3390/app9224729.
[24] T. P. Negron and C. A. Graves, "Classroom Attentiveness Classification Tool (ClassACT): The system introduction," 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), 2017, pp. 26-29, DOI: 10.1109/PERCOMW.2017.7917513.
[25] Robal, Tarmo & Zhao, Yue & Lofi, Christoph & Hauff, Claudia. (2018). Webcam-based Attention Tracking in Online Learning: A Feasibility Study. 189-197. 10.1145/3172944.3172987.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY