Saudi Journal of Educational and Psychological Studies
Edition 13th
Vol. (5) Issue (2)
(1-23) 2024

From Pixels to Perception: Next-Gen Learner Interaction through a Real-time Engagement Detector

Salwa Mrayhi, Mohamed Koutheair Khribi, and Mohamed Jemni
LaTICE Research Laboratory, National Higher Engineering School of Tunis (ENSIT), University of Tunis, ***Tunisia***

Published on: 4 Sept. 2024

Abstract

In the sphere of e-learning, particularly in massive open online courses (MOOCs), accurate, real-time evaluation of learner engagement is increasingly crucial for ensuring the efficacy of education and promoting an inclusive learning environment. Existing techniques for measuring engagement often suffer from subjectivity, inefficiency, and a notable absence of prompt feedback. The system developed has the capability to detect and analyze learner emotions in real time and hence compute the level of engagement throughout the online learning experience in a MOOC environment, providing immediate insights for educators accordingly into learner responses toward MOOC materials. This approach is designed to identify the areas that require enhancement and facilitate the development of personalized and engaging learning experiences for all people, particularly those with disabilities. Furthermore, this research contributes to the advancement of affective computing in education by examining the benefits and challenges of monitoring emotions and engagement in real time. This endeavor comprises a collaborative, interdisciplinary approach, leveraging

knowledge from computer science, psychology, education, and other related fields to enhance the quality of education and enrich learners' learning experiences.

Keywords: MOOCs, engagement detector, facial recognition, emotion, eye gaze, disabilities, artificial intelligence, neural networks, CNN algorithms

* INTRODUCTION

In a world in flux, ensuring access to high-quality, tailored, and inclusive education is crucial. At the same time, learners with disabilities face considerable challenges accessing the educational opportunities they need [1]. Alarmingly, a staggering 240 million learners with disabilities worldwide are among those most likely to be excluded from education, often plaster discrimination and challenges [2]. In addition, conventional learning strategies do not cater to everyone's needs, and learners with different abilities encounter challenges in achieving appropriate learning outcomes [3]. This research focuses on massive open online courses (MOOCs), which present an unparalleled opportunity for learners worldwide to gain access to a myriad of suitable educational resources according to their learning objectives [4]. Owing to their flexibility and extensive course offerings, MOOCs have emerged as indispensable tools in the process of democratizing education. This enables individuals from diverse backgrounds to engage in learning at their own pace and in accordance with their unique educational needs. Nevertheless, despite such benefits, MOOCs also present significant contests that impede their potential impact. One of the most prominent challenges is the high dropout rate, with studies revealing that only approximately 10% of participants typically complete their courses [5]. These alarming statistics raise questions about the effectiveness of MOOCs in retaining and engaging learners. Moreover, learners with disabilities often encounter additional barriers in accessing these educational opportunities, such as navigating online platforms or accessing course content that is not adapted for their needs, such as audio lectures without captions for the deaf or image-based content without alternative text for the visually impaired [6].

Furthermore, the COVID-19 pandemic has motivated accessible, inclusive, and personalized learning experiences, hence sparking new approaches and solutions [7]. In this regard, the work of Huang et al. [8] becomes particularly relevant. The handbook on flexi-learning facilitation during educational disruptions and, in particular, the review of the Chinese experience to guarantee continuous learning during the COVID-19 disaster are highly instrumental. However, at the same time, it is important to ask if the work they are doing is addressing issues of accessibility and inclusivity of persons with disabilities (PWDs). This focus has become increasingly important in the wake of the pandemic, which has led to the enormous need for education systems to be flexible and, at the same time, inclusive, as all learners, in particular, disabled individuals, are not edged out during such times.

This research project contributes to an ongoing endeavour to develop an AI-powered system for personalized learning paths (PLPs) within MOOCs for learners of different abilities. Recognizing the profound influence of learner engagement on the success of MOOCs [9], our study seeks to harness the potential of sensor technology and machine learning to meticulously track and quantify engagement levels across diverse course activities. This intention is driven by the following critical question: How can learner engagement in MOOCs be effectively tracked and quantified using sensor technology and artificial intelligence?

The structure of this article is as follows: Section 2 presents a literature review discussing the existing research on enhancement using sensor technology and machine learning within MOOC environments. A description of the engagement detector system is provided in Section 3, which explains how machine learning algorithms are applied to identify learning behavior and how to develop and implement sensor-based interactions to track learner engagement. The fourth section presents the main results of this research, outlining the outcomes of the experiments and data analysis. The final section discusses the study's contributions and explores possible avenues for future research. The paper discusses how the findings can be further developed and applied to create universally accessible MOOCs.

*** Literature Review**

MOOCs have revolutionized the way learners access education, with millions of online learners from around the world taking these courses [10]. However, MOOCs still face challenges in providing effective and personalized learning experiences for learners with different abilities and needs [11]. This study proposes the use of an engagement detector artifact to address this challenge by detecting learner engagement in personalizing the learning experience. Previous studies have shown that personalized learning can significantly impact learning outcomes [12]. Learners tend to be more engaged, motivated, and successful when the learning experience caters to their specific needs and abilities [13]. However, providing personalized learning experiences [14]–[16] can be challenging, particularly in MOOCs, which have large and diverse user populations [17]. These include the need for adaptive content that responds to each learner's progress and feedback, the difficulty of accommodating different learning styles, and the difficulty of keeping learners engaged in a noninteractive environment. A spark-based personalized MOOC recommendation system is one potential avenue. This system can recommend courses to learners based on their interests and learning goals, helping them find relevant and engaging content [18]. Furthermore, knowledge concept recommendation in MOOCs employs advanced technologies such as knowledge graphs, graph convolutional networks, and language model encoders, demonstrating substantial effectiveness in offering personalized, diverse, and precise course suggestions in MOOC environments. The study findings indicate high user agreement on the accuracy, novelty, and diversity of Concept Graph Convolution Network GCN-based recommendations, suggesting its potential to improve the MOOC learning experience. However, it is important to remember that user perception does not guarantee a direct impact on learning outcomes. Additionally, limitations such as sample size and specific user demographics might restrict generalizability [19]. Another proposed approach for adaptive learning utilizes wearable technology and electroencephalogram (EEG)-based engagement monitoring to tailor educational experiences to each learner's unique needs. By analyzing real-time EEG data, this system evaluates cognitive and emotional engagement levels, offering a more personalized learning experience. Its ability to predict engagement may enhance retention and completion rates in educational settings. This innovative system also paves the way for an automated teaching platform that dynamically adjusts to changes in learners' attention and interest. Through continuous EEG monitoring, the approach aims to revolutionize learning, ensuring active engagement and motivation throughout the learning process [20]. These innovations are part of a broader spectrum of studies focusing on personalization in MOOCs, encompassing engagement detection, recommendation systems, and emotion recognition, among other aspects.

Delving deeper into the nuances of student engagement and emotion detection in educational settings, it becomes imperative to examine the spectrum of existing research. This involves categorizing key studies into thematic areas to offer a well-rounded view of their methodologies, techniques, tools, and findings. The analysis is structured into groups A-E to categorize and highlight each study's approaches and contributions to improving digital learning experiences.

1- Eye Tracking and Facial Emotion Analysis

Studies [20], [22], [23], and [24] have used eye tracking and facial emotion analysis techniques to create systems capable of detecting negative emotions in e-learning environments and alerting educators to unengaged learners. These studies used tools such as OpenCV and MATLAB. However, these approaches often require controlled environments to ensure accuracy and can be invasive, which may affect the natural learning behavior of users. Additionally, they may not fully capture the complexity of emotional states, as they can be influenced by various factors beyond visible expressions.

2- Engagement and Concentration Analysis

Research [21], [25], and [26] has focused on assessing learner engagement and concentration levels. These studies leveraged techniques such as the local binary patterns (LBP) algorithm and deep learning architectures, demonstrating the effectiveness of these approaches in MOOCs and e-learning environments. However, these methods are computationally intensive and require extensive datasets for training, potentially limiting their scalability and accessibility for broader educational applications. Furthermore, they predominantly focus on quantitative metrics of engagement, often overlooking critical qualitative aspects such as motivation and personal interest. This gap presents an opportunity for research aimed at developing more efficient, holistic models of engagement assessment that can capture both the depth and breadth of learner engagement.

3- Advanced Detection Techniques and Models

[27], [28], and [29] explored sophisticated methods for facial feature extraction and engagement detection algorithms. By implementing techniques such as the Facial Action Coding System (FACS) and Local Gray Code Patterns (LGCP), these studies achieved significant strides in measuring engagement through decoded action units. Nevertheless, the reliance on complex algorithms such as FACS and LGCP necessitates high computational resources and specialized expertise, which may not be readily available in all educational settings. There is also a risk of oversimplification, where nuanced emotional expressions are reduced to predefined categories.

4- Biometric and Physiological Signal Analysis

In [30] and [31], innovative approaches using biometric sensors and physiological signals, including EEG signals, were applied to understand emotional and behavioral engagement. These methods revealed the potential of biometrics as dependable indicators of student engagement. However, these approaches can be costly and require specialized equipment that may not be widely available. Moreover, accurately interpreting these signals necessitates a deep understanding of the biological basis of emotional and cognitive processes, which can be intricate and varied among individuals.

5- Multimodal Data and Online Training Analysis

In [32] and [33], multimodal data analysis and advanced facial analysis were conducted in online training environments. These studies were crucial for revealing varied engagement patterns among teachers and validating the effectiveness of facial expression recognition in gauging emotional states and engagement levels. The integration of diverse data types, from video to biometric signals, introduces analytical challenges that require advanced tools for accurate interpretation. This approach, while insightful, significantly increases the demands on data processing infrastructure and storage solutions, highlighting the critical balance between depth of analysis and resource efficiency.

This research advances previous methodologies, such as eye tracking, facial emotion recognition, and physiological signal analysis, as well as traditional methods such as self-reports and basic activity tracking, by integrating them with improved algorithms for enhanced accuracy in real-time engagement detection. Unlike “traditional” e-learning environments, where engagement detection methods are more commonly applied, MOOC platforms have limited the application of such comprehensive, real-time approaches [21]. The contribution of this work lies in its adaptation and optimization of these methodologies for the scalable, diverse audience of MOOCs, addressing specific challenges such as large participant numbers, their specific needs and varied learning contexts. This distinction underscores the importance of developing specialized engagement detection techniques tailored to the unique dynamics of MOOCs, thereby filling a notable gap in current educational technology research. Additionally, the research focused on real-world applicability, available equipment and extensive user feedback, ensuring practical relevance and addressing potential limitations in existing e-learning systems by enhancing privacy and ethical standards. It contributes to responsible data handling practices in MOOC environments, focusing on improving privacy protections and ethical considerations. Following this, Affectiva Classroom AI and Emotuit were compared based on their alignment goals in educational technology and their distinct and representative approaches in the domain (Table I). Both are focused on real-time engagement analysis, yet they differ in terms of the methods and insights they offer. This comparison highlights how each tool uniquely contributes to understanding and enhancing learner engagement.

Table 1. Comparison of two related programs

Program	Affectiva Classroom AI	Emotuit
Approach	Facial recognition technology + Machine learning algorithms	Machine learning algorithms
Focus	Emotion detection	Behavior analysis
Real-time Analysis	Yes	Yes
Data Analyzed	Facial expressions	Mouse movements, keyboard inputs, etc.
Insights Provided	Learner emotions	Learner behavior
Purpose	To measure engagement and well-being, and provide insights to educators	To measure engagement, identify areas for improvement, and optimize teaching strategies

The analysis showed that Affectiva focuses on visual emotional cues, while Emotuit emphasizes behavioral data. This highlights a crucial gap where each system may miss key engagement indicators captured by the other system. Therefore, a synergistic approach that combines the strengths of both systems and addresses their limitations is needed. To achieve this goal, our proposed solution advocates for a nuanced integration of emotional and behavioral analytics using cutting-edge machine learning techniques. Our approach aims to combine the emotional depth of Affectiva with the detailed behavioral insights from Emotuit, resulting in a comprehensive understanding of student engagement.

*** The proposed approach**

This research focuses on creating an engagement detection system for MOOCs by employing sensors and machine learning techniques to analyze patterns of learner engagement. The goal is to enhance the understanding of how learners engage during their educational experience and to quantify their learning behavior. The methodology adopted for this purpose is the Cross-Industry Standard Process for Data Mining (CRISP-DM), known for its comprehensive and iterative approach [22]. The effectiveness of CRISP-DM lies in its ability to discover involved patterns in large datasets at the intersection of machine learning methods. This method is especially suited to the dynamic and evolving demands of educational technology research because of its comprehensive nature and adaptability to various research scenarios.

The CRISP-DM model is a reference guide for data mining projects, providing an overview of their life cycle. It consists of six phases, each with its own set of tasks and interrelationships [23].

1- Business Understanding

The research begins by establishing a comprehensive understanding of the problem. Traditional methods of measuring student engagement and emotion in online education, such as self-reports and basic activity tracking, have limitations. These methods, as highlighted in the literature review, often involve self-reports or basic activity tracking; however, real-time analysis is lacking, and the results may not capture the full spectrum of learner engagement [24]. Delayed feedback, potential biases in self-reports, and a lack of contextual data can limit these methods, which might prevent educators from effectively adapting their teaching strategies to individual learner needs in a timely manner [25].

The goal is to develop an advanced system that can provide real-time insights into learner emotions and engagement levels through MOOCs. This approach will enable educators to enhance teaching effectiveness and create personalized learning experiences, providing automatic guidance for learners, especially those with disabilities.

2- Data Understanding

A- Data Description

Various datasets of video recordings from online courses are gathered to train emotion and engagement detection models. The data undergo careful preprocessing to address noise, variations in lighting, and potential occlusions. Data augmentation techniques are applied to increase dataset diversity and prevent overfitting. The images are resized and normalized for better model training. Many datasets are used for facial emotion recognition (FER) [26]. The facial expressions consisted of seven basic facial expression classes: anger, surprise, fear, happiness, disgust, sadness, and neutral. The famous FER [27] datasets are outlined in Table II.

Table 2. Comparison of popular FER datasets

Dataset	Database Information	Size	Number of fearful examples	Expression
FER-2013	- Created using the Google image search. - Consider gender, age and ethnicity variation	35,887 Images	5,121	Spontaneous
CK+	- 123 subjects - 69 female,31 male - Age: 18-50 years	5,876 images	423	Posed Spontaneous
JAFFE	- 10 female Japanese models	213 images	33	Posed
MMI	- 25 subjects - 12 female, 13 male - Age: 20-32 years	238 images and sequences	28	Posed Spontaneous
MUG	- 86 subjects - Age: 20-35 - 35 female, 51 male	324 sequences	51	Posed
ExpW	Downloaded from Google search and manually annotated	91.793 images	51	Spontaneous
Raf-DB	Collected from the internet for diverse ages and races. 29,672 posed images. 70 subjects.	29.672 images	355	Posed
KDEF	-70 subjects - 35 female , 35 male -Age 20-30 years	4.900 images	35	Posed

To address learner engagement measurement accurately, we opted to build a CNN model from scratch rather than use. Pre-trained models such as YOLO, ensuring that the model directly addresses our research focus, including the nuanced needs of learners with disabilities. This approach allows for the precise personalization required for recognizing subtle facial expressions and body language across a diverse range of learners. First, to train a CNN model, we require a substantial dataset that not only shows diversity in geography, ethnicity, gender, and age but also specifically includes learners with disabilities. Importantly, selecting a dataset rich in significant emotional data, such as fear, anger, and surprise, is crucial for effective emotion detection, ensuring the model's applicability in diverse educational contexts, particularly for learners with disabilities. Therefore, the table above displays the number of examples of fearful instances. It is evident that certain extensive datasets, such as ExpW, have only a few instances of fear, making them unsuitable for the intended use. Therefore, the FER-2013 dataset is chosen for developing the facial emotion recognition model because it fulfills the aforementioned requirements.

The FER-2013 dataset detailed in Table III is an open-source collection of facial images created by Pierre Luc Carrier. It consists of 35.887 grayscale images of faces with a resolution of 48x48 pixels[28]. It is selected for training the facial emotion recognition model due to its realistic set of grayscale images that mirror a wide range of human expressions found in everyday scenarios. The dataset, sourced from Google, provides a rich variety of facial expressions, enhancing the ability of the model to generalize across different contexts. Its preprocessed images, with faces centrally aligned and occupying consistent space, streamline the initial stages of model training and contribute to the development of more accurate algorithms attuned to the refinements of human emotions.

Table 3. Data Description

Name	Year	Size	Type of Variations
CK +	2010	593	angry, sad, disgust, frustration, happy surprise , neutral.
FER-2013	2013	35.877	angry, sad, disgust, frustration, happy surprise , neutral
Wider Face	2016	32.203	Scale , Pose , Expression , Make up ....

The dataset labels can be divided into 7 categories: 0: angry, 1: disgust, 2: fearful, 3: happy, 4: neutral, 5: sad, and 6: surprise, as shown in Figure 1.

Figure 1. FER-2013 dataset sample [29]

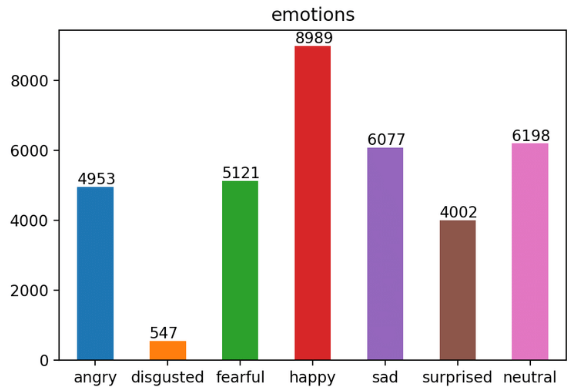

Figure 2. Distribution of FER-2013 dataset classes[30]

The FER-2013 dataset represents several challenges. As shown, the dataset is imbalanced, and the number of examples in each class is not equal, as revealed in Figure 2. For example, the disgust emotion is much less common in the dataset than other emotions, and the happy emotion has many more examples than other classes. Additionally, the dataset exhibits an intraclass variation, which is the difference in nature of examples of the same class. For example, we have both cartoon and human faces with the same emotion. Deep learning architectures are robust enough to address this type of challenge.

Furthermore, FER-2013 includes images with facial occlusion; the most known type of occlusion is a hand-covered face, so only a part of a face is visible. Additionally, some of the images exhibit contrast variation, meaning that they are too black or too white. Another challenge of the FER-2013 dataset is eyeglasses, as they hide the eyes, which are a very important feature of FER. The dataset also contains outliers since some images can be mislabeled.

B- Data Preparation

In the critical phase of data preparation for the project, a structured and multifaceted approach was employed. First, the data involved were understood and recognized via methods such as face detection and cropping, where external disturbances such as background, clothing, and accessories were meticulously removed using dlib’s face detector to enhance facial feature extraction and classification. A subsequent step included face alignment, where the images were systematically centered, rotated, and scaled to create uniformity across the dataset using dlib's 5-face landmark detector. Reshaping and normalization were integral to the preprocessing step, ensuring compatibility with pretrained models such as ResNet and VGG and aligning pixel values between 0 and 1, respectively, for optimized learning within neural networks. Finally, the dataset was judiciously divided into three segments: a training set constituting 80% (28,709 images) for model training and parameter optimization; a validation set comprising 10% (3,589 images) for hyperparameter tuning, model evaluation, and overfitting detection; and a testing set also representing 10% (3,589 images) for final model assessment and comparison. This careful orchestration not only contributed to enhancing the model's performance but also laid the foundation for a robust and nuanced analysis of the FER-2013 dataset, which included 35,887 images (Figure **3**).

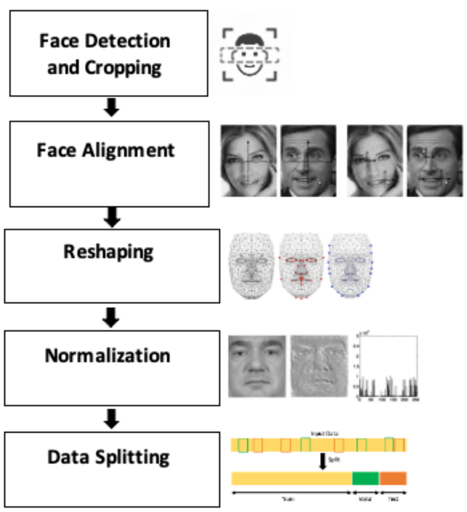

Figure 3**. Data Preparation**

C- Data augmentation

Enhancing module performance often requires an extensive dataset. To achieve this, we commonly implement data augmentation techniques that generate variations in the original data, thereby increasing the volume and diversity of the dataset [40]. The augmentations performed in this study are detailed as follows:-

1- Rotation: Rotation involves orienting the image on an axis between 1° and 359°. The effectiveness of rotation is contingent on the degree of the angle. For instance, minor rotations between 1 and 20 degrees or -1 to -20 degrees have demonstrated utility in digit recognition tasks, such as those found in the MNIST dataset.

2- Translation: Translation is a transformation that shifts images left, right, up, or down, thereby reducing positional bias. This approach is particularly useful when images in a dataset, such as those in face recognition studies, are uniformly centered. Translation necessitates adjustments to the remaining space, which can be managed by filling with constant values (such as 0 or 255) or by inserting random or Gaussian noise. This technique guarantees the preservation of spatial dimensions postaugmentation.

3- Noise Injection: Noise injection involves the infusion of a matrix filled with random values, typically originating from a Gaussian distribution. It has been rigorously tested by researchers such as Moreno-Barea et al. across various datasets in the UCI repository. The addition of noise to images fosters the learning of more resilient features via convolutional neural networks (CNNs).

The successful application of these operations on a sample of the FER dataset is illustrated in Figure 4.

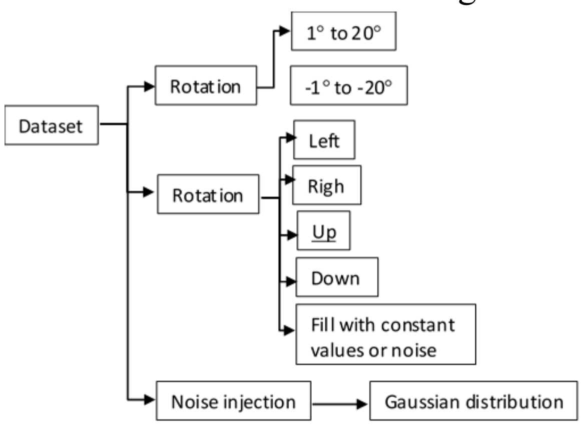

Figure 4. Data augmentation

3- Model development and evaluation

Model B, an advanced variant of the convolutional neural network (CNN) with additional layers or modified hyperparameters, outperforms Model A on emotion and engagement recognition tasks; this model follows a standard architecture. This superiority is evidenced by the higher accuracy and lower loss rates attributed to Model B, highlighting its efficacy in image-based analysis. Deep learning models, especially CNNs, are chosen for their established effectiveness in these domains. A comprehensive evaluation was conducted using metrics such as accuracy, precision, recall, and the F1-score. These metrics assess the overall prediction accuracy, ability to identify true positives, effectiveness in recognizing all positive cases, and balance between precision and recall, thereby confirming Model B's robustness in various performance aspects.

4- System Architecture

The FER system architecture is designed to integrate face and eye detection with real-time emotion detection and engagement index calculations. The single shot detector (SSD) is employed for face detection, and the Haar cascade algorithm is used for eye detection. Model B is integrated into the system for real-time emotion detection. The engagement index is calculated based on positive emotions relevant to learning. The proposed global architecture of the system developed during this project is illustrated in Figure 5.

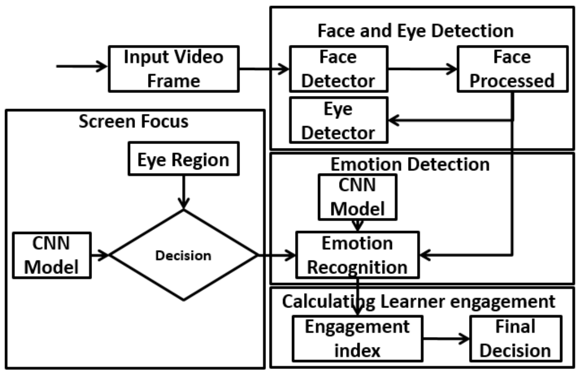

Figure 5. Proposed System Architecture

Extensive testing, incorporating test images and video streams, was pivotal in assessing the system's precision and accuracy. Analysis of the confusion matrices revealed both strengths and areas for improvement in the model. The integration of advanced face and eye detection techniques using SSD architecture and the Haar cascade algorithm enables real-time monitoring of student emotions and engagement. These methods are crucial for analyzing learner interactions effectively in MOOC environments, where the unique format and scale of MOOCs differentiate them from other online content. The variability of course design, interactive components, and necessity for engagement measurement across diverse international learner bases underscore the specific applicability of these techniques to MOOCs, highlighting their importance in the distinct context of MOOC learning experiences. The methodology combines these elements with robust CNN models to achieve high accuracy and provide essential real-time insights for educators, aiming to enhance the learning experience.

*** The results**

The integration of the cvzone face detection model and the Haar cascade eye detection algorithm in the project has resulted in precise and reliable detection of facial and ocular features. This high level of accuracy is crucial for the overall success and effectiveness of the research, underpinning the system's ability to monitor and analyze key aspects of learner engagement and emotion in real time. The accurate identification and tracking of these features are essential components in achieving the desired outcomes of this study, as depicted in Figure 6.

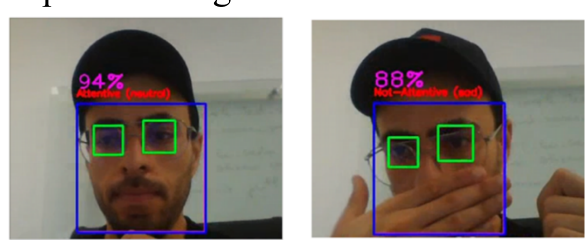

Figure 6. Face and Eye Detection Demonstration

1- Screen Focus

Integrating the Haar cascade algorithm into the system improves its ability to detect when a student's gaze is fixed on the screen. This detection is not simply an observational process but rather an interpretive process, allowing the system to infer a likely state of attention. This inference, based on the algorithm's robust functionality, enables it to respond dynamically to the student's level of engagement. The system uses this information in real time to provide targeted feedback customized to the student's individual needs or to make strategic adjustments to the pace and content of the lesson. These adjustments are not made randomly but are guided by continuous monitoring of the algorithm to ensure that the algorithm meets the learner's requirements in real time.

It represents a significant advance in the system's ability to understand, monitor, and positively influence student engagement and learning outcomes. This integration underscores the commitment to using cutting-edge technology to create an adaptive and empathetic learning environment that sets a new standard for personalized learning.

2- Emotion detection

To enhance the understanding of learner engagement, the engagement detector system incorporates not only face and eye detection but also real-time emotion detection using pretrained Model B. This model rigorously refined and efficiently identified seven key emotions: neutral, happy, sad, angry, surprised, afraid, and disgusted (Figure 7). Its real-time processing capability transforms technical capability into a pedagogical tool by revealing the classroom's emotional dynamics as they occur.

Complex emotional data are rendered accessible and intuitive through dynamic visualizations created with the Plotly library, representing each emotion's likelihood in a horizontal bar chart. This visualization acts as a gateway to students' emotional states, transforming abstract data into practical insights.

This real-time emotional tracking approach represents a proactive step toward engaging with and responding to learner emotions during classroom activities, enabling tailored teaching strategies and personalized support.

This fusion of Model B and Plotly represents a technological advance that serves pedagogical innovation, marking a significant step forward in real-time emotion detection and visualization. This integration directly enhances pedagogical interaction, particularly benefiting learners with disabilities, who may find it challenging to express their emotions. By providing educators with real-time emotional cues, this tool can transform the responsiveness of educational strategies, leading to improved engagement and learning outcomes. This study underlines the holistic approach to education, where emotional understanding is paramount, showcasing the belief in technology's power to empathize and inspire, epitomizing adaptive, responsive, human-centered education.

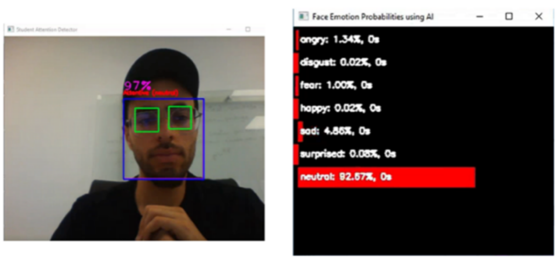

Figure 7. Seven key emotions

2- Engagement Index

Building upon the work of Sharma et al. in "Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning," [31] an engagement index has been developed that functions as a real-time indicator of the emotional dynamics within a learning environment. This index quantifies student engagement by analyzing seven key emotions, each assigned a specific weight based on its significance to the learning process. In layman's terms, a thermometer measures the 'temperature' of a student's level of connection to the educational material, offering educators immediate, actionable insights for enhancing the learning experience.

The Engagement Index (CI) quantitatively assesses emotional states, translating them into measurable scores. Positive emotions such as happiness or surprise, observed during video engagement, correlate with higher CI scores, indicating increased engagement. Negative emotions, such as sadness or fear, yield lower scores, signifying decreased engagement. This approach enables precise quantification of emotional reactions, offering an objective measure of student engagement levels.

The CI is calculated using the following formula:-

$$CI = DEP * EW$$

(1) [32]

Here, CI is the concentration/engagement index, DEP is the dominant emotion probability, and EW is the emotion weight. In essence, each emotion is assigned a weight based on its relevance to learning, and the index is calculated by multiplying the probability of the dominant emotion by its weight. These emotions are assigned weights as outlined in Figure 8, which serves as a visual representation of how different emotions are weighted in relation to their impact on learning engagement. For example, emotions with greater weights on the chart, such as 'Happy' and 'Surprised', are considered more conducive to an engaged learning state, while those with lesser weights, such as 'Sad' and 'Angry', may indicate disengagement.

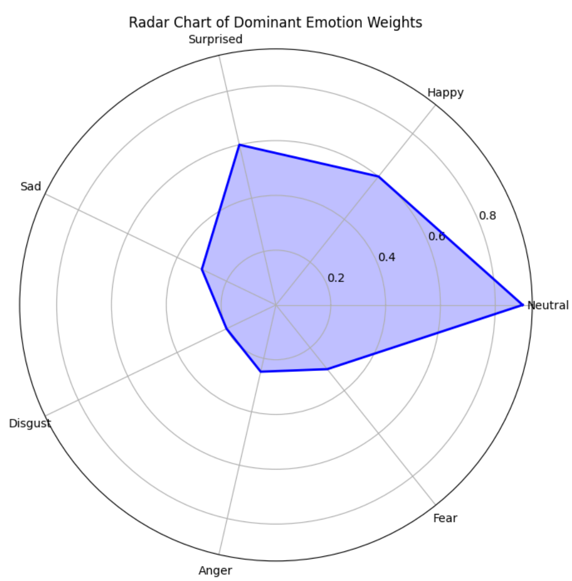

Figure 8. Emotional weights

The system dynamically adjusts to student engagement levels through real-time monitoring of the engagement index. Should this index drop below a certain threshold for an individual or a group, the system triggers tailored interventions to enhance support or adjust the course content. Conversely, an increase in the Engagement Index offers an opportunity to promote positive behaviors and encourage deeper involvement in the learning process. This approach, blending face, gaze, and emotion detection with real-time engagement analytics, is expected to lead to a significant improvement in educational technology ID enhancement and appropriate use. It offers deep insights into learner emotions and behaviors and is synthesized into comprehensive reports for educators. These reports can guide teaching strategies and personalized interventions. Additionally, this system can feed a content-based recommender system, update learner models based on preferences and learning styles to suggest suitable MOOCs, and further personalize the educational experience to meet individual learner needs.

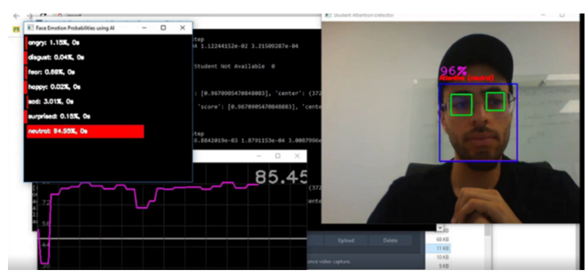

Figure 9. Positive Emotion

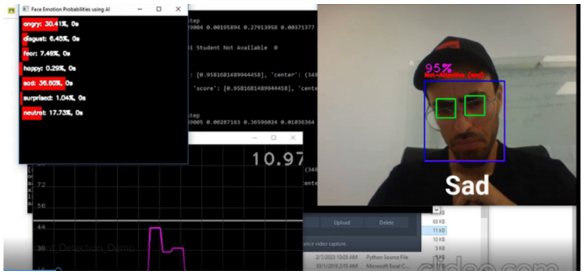

Figure 10. Negative Emotion

The integration of positive and negative emotion detection, as depicted in Figures 9 and 10, can be pivotal in developing a rating system to assess learner satisfaction. The Engagement Index, more than just a measurement tool, aims to foster a supportive educational environment, which is backed by current research. Future enhancements, including the integration of test scores, heart rate, quiz duration, and reinforcement learning, could refine and improve this model. This method is poised to revolutionize how technology supports education, transforming it from a mere tool to an active collaborator in the teaching and learning process.

*** Discussion**

This ongoing research is intended to enhance inclusivity and adaptivity in learning environments, particularly within MOOCs. This highlights the importance of tailoring learning environments to the diverse needs of learners, reflecting a commitment to addressing the educational requirements of a wide range of learners, including those with disabilities.

Addressing the educational requirements of learners with disabilities remains a paramount challenge. Patiño Toroa et al. [33] highlighted the importance of tailoring MOOCs for the deaf and hard of hearing people by incorporating sign language and visual aids. This aligns with the current endeavors aimed at accessing sensor detectors and machine learning to address diverse learner needs. Moreover, understanding the behaviors associated with autism spectrum disorder (ASD), for example, is crucial. Banire et al. [34] supported the development of machine learning models that evaluate emotional engagement customized to each learner's unique profile. This approach is particularly relevant for ASD learners, where tailored interventions can be more effective.

This research complements Sharma et al.'s [23] work in using technology to boost student engagement. Sensors and webcams are being integrated to capture nuanced indicators such as head movements and facial expressions, aligning with the goal of adapting educational technology with practical hardware solutions to learner evolving profiles. Furthermore, Sharma et al.'s approach to classifying students' head and eye movements aligns with the intention of developing personalized machine learning models. These models are designed to adapt to individual engagement patterns, a key factor in enhancing the effectiveness of learning experiences and outcomes.

Our research aligns with the academic consensus on inclusion, personalization, and multimodal methods in educational tech. The aim is to make MOOCs accessible and engaging all, including addressing biases, fairness, privacy, and ethics in AI within education. Despite the limitations of sensor detectors and machine learning models, their potential in promoting inclusive education has been recognized [35]. Efforts are geared toward developing solutions that address these challenges, contributing to equitable education advancements through technology.

*** Conclusion**

This ongoing project contributes to educational technology by incorporating specific technologies such as facial recognition, eye gaze tracking, and emotion detection algorithms. It revolutionizes the monitoring and enhancement of student engagement in both traditional classroom settings and MOOCs. The use of such technologies has the potential to reshape learning engagement, offering more interactive and personalized educational experiences.

The real-time analytics capability of the system offers profound insights into student learning behaviors, enabling the creation of more effective and adaptive educational strategies. In the future, the plan includes incorporating reinforcement learning algorithms to further enhance the system's adaptability in meeting the dynamic engagement levels of students. Additionally, a key focus is on augmenting the system's accessibility, ensuring that it is inclusive for all learners, particularly those with disabilities. This approach goes beyond mere technology integration; it is about shaping an equitable and inclusive learning environment.

Integrating this system into MOOC platforms could lead to a ground-breaking shift in online education, potentially heightening user engagement and satisfaction. The commitment to continuous refinement ensures that the system remains at the cutting edge of educational technology.

The transformative potential of technology in education, particularly the role of sensor detectors, machine learning and analytics, is acknowledged. Future extensions of this work will focus on integrating more refined algorithms to enhance real-time emotion detection and expand the system's applications, with a commitment to personalized and accessible learning platforms, designs and content for all learners, including those with disabilities. This research is a step toward realizing more engaged, empathetic, and inclusive education.

*** Acknowledgments**

The authors would like to thank Avaxia for providing the staff and support required to complete this project. The authors would also like to thank Amine Ayari, who was involved in many aspects of this project as an intern. Their contributions made this work possible.

*** References**

[1] "Inclusive education | UNICEF." https://www.unicef.org/education/inclusive-education (accessed May 14, 2023).

[2] S. R. Lambert, "Do MOOCs contribute to student equity and social inclusion? A systematic review 2014–18," *Comput. Educ.*, vol. 145, p. 103693, Feb. 2020, doi: 10.1016/J.COMPEDU.2019.103693.

[3] D. Aggarwal, "Integration of innovative technological developments and AI with education for an adaptive learning pedagogy," *China Pet. Process. Petrochemical Technol.*, vol. 23, no. 2, 2023.

[4] R. Rulinawaty, A. Priyanto, S. Kuncoro, D. Rahmawaty, and A. Wijaya, "Massive Open Online Courses (MOOCs) as Catalysts of Change in Education During Unprecedented Times: A Narrative Review," *J. Penelit. Pendidik. IPA*, vol. 9, no. SpecialIssue, pp. 53–63, 2023.

[5] H. Huang, L. Jew, and D. Qi, "Take a MOOC and then drop: A systematic review of MOOC engagement pattern and dropout factor," *Heliyon*, vol. 9, no. 4, p. e15220, Apr. 2023, doi: 10.1016/J.HELIYON.2023.E15220.

[6] S. Mrayhi, K. Mohamed Koutheair, and Mohamed Jemni, "Ensuring Inclusivity in MOOCs: The Importance of UDL and Digital Accessibility," *ICALT* , 2023, Accessed: Jul. 15, 2023. [Online]. Available: https://icalt.online/webpub/pdfs/ICAL

T2023-39Kl8aG3vUgRQSB3RG54hu/005400a044/005400a044.pdf
[7] M. Laufer et al., “Digital higher education: a divider or bridge builder? Leadership perspectives on edtech in a COVID-19 reality,” *Int. J. Educ. Technol. High. Educ.*, vol. 18, pp. 1–17, 2021.
[8] R. H. Huang, D. J. Liu, A. Tlili, Yang, and J. F. Wang, “Handbook on Facilitating Flexible Learning During Educational Disruption: The Chinese Experience in Maintaining Undisrupted Learning in COVID-19 Outbreak Rights and Permissions Handbook on Facilitating Flexible Learning During Educational Disruption: The Chinese Experience in Maintaining Undisrupted Learning in,” 2020, Accessed: Jan. 07, 2024. [Online]. Available: http://creativecommons.org/licenses/bysa/3.0/igo/
[9] Y. Sun, Y. Guo, and Y. Zhao, “Understanding the determinants of learner engagement in MOOCs: An adaptive structuration perspective,” *Comput. \& Educ.*, vol. 157, p. 103963, 2020.
[10] J. M. Weinhardt and T. Sitzmann, “Revolutionizing training and education? Three questions regarding massive open online courses (MOOCs),” *Hum. Resour. Manag. Rev.*, vol. 29, no. 2, pp. 218–225, 2019.
[11] A. Papadimitriou, “Adaptive and Intelligent MOOCs: How they Contribute to the Improvement of the MOOCs’ Effectiveness,” *Int. J. Res. E-learning*, pp. 1–21, 2023.
[12] Y. Deng, “Application of edge computing and artificial intelligence in English MOOC assisted teaching instruction,” *Int. J. Syst. Assur. Eng. Manag.*, pp. 1–9, 2023.
[13] C. Strauss et al., “What is compassion and how can we measure it? A review of definitions and measures,” *Clin. Psychol. Rev.*, vol. 47, pp. 15–27, Jul. 2016, doi: 10.1016/J.CPR.2016.05.004.
[14] Khribi, M. K., Jemni, M., and Nasraoui, O. “Automatic recommendations for e-learning personalization based on Web usage mining techniques and information retrieval,” *Proc. - 8th IEEE Int. Conf. Adv. Learn. Technol. ICALT 2008*, pp. 241–245, 2008, doi: 10.1109/ICALT.2008.198.
[15] Jemni, M., Kinshuk, and Khribi, M. K., Eds., “Open Education: from OERs to MOOCs,” 2017, doi: 10.1007/978-3-662-52925-6.
[16] Kinshuk and R. Huang, Eds., “Ubiquitous Learning Environments

and Technologies," 2015, doi: 10.1007/978-3-662-44659-1.
[17] Ben Daamech and Khribi, M.K., "Toward a framework for building automatic recommendations of answers in MOOCs' Discussion Forums," *2019 7th Int. Conf. ICT Access. ICTA 2019*, Dec. 2019, doi: 10.1109/ICTA49490.2019.9144850.
[18] N. Mansouri, M. Soui, and M. Abed, "SFS feature selection with decision tree classifier for massive open online courses (MOOCs) recommendation," *J. Comput. Educ.*, pp. 1–22, 2023.
[19] R. Alatrash, M. A. Chatti, Q. U. Ain, Y. Fang, S. Joarder, and C. Siepmann, "ConceptGCN: Knowledge concept recommendation in MOOCs based on knowledge graph convolutional networks and SBERT," *Comput. Educ. Artif. Intell.*, vol. 6, p. 100193, 2024.
[20] Ay¸se Saliha Sunar, "Prediction of Course Completion based on Participants' Social Engagement on a Social-Constructivist MOOC Platform," 2017. Accessed: May 15, 2023. [Online]. Available: https://eprints.soton.ac.uk/419583/1/Final_Thesis.pdf
[21] A. Dubbaka and A. Gopalan, "Detecting learner engagement in MOOCs using automatic facial expression recognition," in *2020 IEEE Global Engineering Education Conference (EDUCON)*, 2020, pp. 447–456.
[22] J. Bokrantz, M. Subramaniyan, and A. Skoogh, "Realizing the promises of artificial intelligence in manufacturing by enhancing CRISP-DM," *Prod. Plan. \& Control*, pp. 1–21, 2023.
[23] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, Jan. 2021, doi: 10.1016/J.PROCS.2021.01.199.
[24] C. Samani, A. Atif, and K. Musial-Gabrys, "Using Emotional Learning Analytics to Improve Students' Engagement in Online Learning," *ASCILITE Publ.*, pp. e22129--e22129, 2022.
[25] L. Saunders and M. A. Wong, "Critical pedagogy: Challenging bias and creating inclusive classrooms," *Instr. Libr. Inf. centers*, 2020.
[26] P. Naga, S. Das Marri, and R. Borreo, "Facial emotion recognition methods, datasets and technologies: A literature survey," *Mater. Today, Proc.*, vol. 80, pp. 2824–2828, 2023.
[27] "FER-2013." https://www.kaggle.com/datasets/msa

mbare/fer2013 (accessed Oct. 08, 2023).
[28] M. I. Georgescu, R. T. Ionescu, and M. Popescu, “Local learning with deep and handcrafted features for facial expression recognition,” *IEEE Access*, vol. 7, pp. 64827–64836, 2019, doi: 10.1109/ACCESS.2019.2917266.
[29] C. M. A. Ilyas, R. Nunes, K. Nasrollahi, M. Rehm, and T. B. Moeslund, “Deep Emotion Recognition through Upper Body Movements and Facial Expression.,” in *VISIGRAPP (5: VISAPP)*, 2021, pp. 669–679.
[30] K. Sarvakar, R. Senkamalavalli, S. Raghavendra, J. S. Kumar, R. Manjunath, and S. Jaiswal, “Facial emotion recognition using convolutional neural networks,” *Mater. Today, Proc.*, vol. 80, pp. 3560–3564, 2023.
[31] P. Sharma et al., “Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning”.
[32] P. Sharma et al., “Student engagement detection using emotion analysis, eye tracking and head movement with machine learning,” in *International Conference on Technology and Innovation in Learning, Teaching and Education*, 2022, pp. 52–68.
[33] O. Nidia Patiño-Toro, A. Valencia-Arias, A. Fernández-Toro, A. Jiménez-Guzmán, and C. Augusto Puerta Gil, “Proposed methodology for designing and developing MOOCs for the deaf community ☆,” 2023, doi: 10.1016/j.heliyon.2023.e20456.
[34] B. Banire, D. Al Thani, and M. Qaraqe, “One size does not fit all: detecting attention in children with autism using machine learning,” *User Model. User-adapt. Interact.*, pp. 1–33, 2023.
[35] S. Vincent-Lancrin and R. der Vlies, “Trustworthy artificial intelligence (AI) in education: Promises and challenges,” 2020.