Student Attention Monitoring System using Webcam

Haritha T[1], Aaisha Sultaana Z[2], Bharani S[3]
Department of Information Technology,
Vivekanandha College of Engineering for Women,
Namakkal, India.

Abstract: This project is aimed to monitor the students while attending online class through web cam. This project is developed by using python 3.11 under machine learning. With the widespread adoption of online education, ensuring student engagement and attentiveness during virtual classes has become a paramount concern. This project introduces a Student Attention Monitoring System that utilizes webcam technology to assess and enhance student participation during online learning sessions.

The system employs computer vision algorithms to analyze facial features and eye movements, providing real-time insights into student focus and attention levels. By detecting signs of distraction or disengagement, such as prolonged gaze diversion, the system generates automated alerts to both educators and students, fostering an interactive and attentive virtual learning environment.

The proposed system not only serves as a tool for instructors to gauge the efficacy of their teaching methods but also encourages students to maintain an active and focused presence during online classes. Ethical considerations regarding privacy and data security are integral components of the system design, ensuring a balance between monitoring effectiveness and the protection of students' privacy.

I. INTRODUCTION

In the rapidly evolving landscape of education, fostering student engagement is paramount for effective learning outcomes. With the advent of remote and online learning, the challenge of maintaining student attention becomes even more pronounced. To address this, we introduce a cutting-edge solution – the Student Attention Monitoring System (SAMS) using a webcam and powered by Python.
SAMS leverages the capabilities of computer vision to monitor and analyze students' attentiveness during virtual classes or self-paced learning sessions. By utilizing the webcam, this system discreetly observes facial cues and movement patterns, providing valuable insights into the level of focus exhibited by students throughout the learning process.

Traditional methods of assessing student engagement, such as quizzes or surveys, are limited in capturing real-time attention dynamics. SAMS, on the other hand, offers a non-intrusive and objective approach to gauge attentiveness. The system employs facial recognition and tracking algorithms to detect key indicators, including facial expressions, eye movements, and head orientation.

Key Features of SAMS:

1. Facial Recognition: SAMS utilizes advanced facial recognition algorithms to identify individual students and track their facial expressions.
2. Eye Tracking: The system monitors eye movements to assess visual attention, ensuring that students are actively engaged with the content.

This innovative solution not only assists educators in optimizing their teaching methods but also empowers students to be more conscious of their own learning habits.

As we delve into the details of the system's implementation using Python, we embark on a journey to revolutionize the way we understand and enhance student engagement in the digital era.

II.RELATED WORKS

Facial Detection and recognition research has been widely studied in recent years.
The facial recognition applications plays an important role in many areas such as security, camera surveillance, identity verification in modern electronic devices, criminal investigations, database management systems and smart card applications etc. This work presents deep learning algorithms used in facial recognition for accurate identification and detection. The main objective of facial recognition is to authenticate

and identify the facial features. However, the facial features are captured in real time and processed using haar cascade detection. The sequential process of the work is defined in three different phases where in the first phase human face is detected from the camera and in the second phase, the captured input is analyzed based on the features and database used with support of keras convolutional neural network model. In the last phase human face is authenticated to classify the emotions of human as happy, neutral, angry, sad, disgust and surprise. The proposed work presented is simplified in three objectives as face detection, recognition and emotion classification.

In support of this work Open CV library, dataset and python programming is used for computer vision techniques involved. In order to prove real time efficacy, an experiment was conducted for multiple students to identify their inner emotions and find physiological changes for each face. The results of the experiments demonstrates the perfections in face analysis system. Finally, the performance of automatic face detection and recognition is measured with Accuracy.

Human computer interaction is a common trend and innate ability to distinguish among multiple faces. Until recent past computer vision problems were quite challenging but advent of modern technologies has trivially improved from the problems of varying light, changed by age, hair and other accessories [1]. However, face recognition applications are used improve access to identify and verify the people by their face features. Hence interpreting the facial features and their actions is much required. As these features and expressions helps in classify the emotions of human face. Recent advances in technology has resulted in the use of Artificial intelligence system as these systems are capable to understand and realize the emotion recognition through facial features. Hence this is an attempt to prove the existence of latest technological developments for human-computer interaction using deep learning or Convolution neural network models [2].

To recognize and classify the human face various methods are required but deep learning technique outperforms other methods by its large capabilities of different datasets and fast computation capabilities. Usually the process of face recognition and classification involves various steps such as preprocessing, detection, orientation, extraction of features and classification of emotion.

The proposed work carries out in three sequential steps as Face Detection, Face Recognition and Face Classification. In the first step a video camera is used to capture the human face and detect the exact location of face by a bounding box coordinates for the face detected in real-time. This step involves face detection using Haar cascade detection with open CV library. Viola jones algorithm and haar cascade features are combined to detect human face. The images detected have shapes, objects and landscapes etc. In this phase human face is detected and face features are extracted and stored in the database for face recognition. The CNN model as shown in figure 4 uses VGG 16 to match the face from the database and recognize with the name for the face detected.

Faces are recognized from the database and are compared to identify or detect the face through embedding vectors. The distribution platform use Anaconda and python 3.5 software in processing face detection, recognition and classification. The image features in the database dlib and other libraries. First face is detected and then recognized with the database features and matching using CNN model training and testing database. Finally the recognized human face is classified based on the expression in real time as Angry, fear, disgust, happy, neutral and surprise. The network architecture VGG 16 is built with CNN model for large database recognition and classification. The designed network model has honey comb 3 x 3 layers where the two connected layers have 4096 nodes with Softmax classification. The local binary model histogram is used as open CV library for detecting the human faces. The image pixels are identified by setting a threshold and the end result is represented in form of a binary number.

One key challenging issues of facial expression recognition (FER) in video sequences is to extract discriminative spatiotemporal video features from facial expression images in video sequences. In this paper, we propose a new method of FER in video sequences via a hybrid deep learning model.The proposed method first employs two individual deep convolutional neural networks (CNNs), includinga spatial CNN processing static facial images and a temporal CN network processing optical flow images,to separately learn high-level spatial and temporal features on the divided video segments.

These two CNNsare fine-tuned on target video facial expression datasets from a pre-trained CNN model. Then, the obtainedsegment-level spatial and temporal features are integrated into a deep fusion network built with a deep beliefnetwork (DBN) model. This deep fusion network is used to jointly learn discriminative spatiotemporalfeatures. Finally, an average pooling is performed on the learned DBN segment-level features in a videosequence, to produce a fixed-length global video feature representation. Based on the global video feature representations, a linear support vector machine (SVM) is employed for facial expression classification tasks.The extensive experiments on three public video-based facial expression datasets, i.e., BAUM-1s, RML,and MMI, show the effectiveness of our proposed method, outperforming the state-of-the-arts.

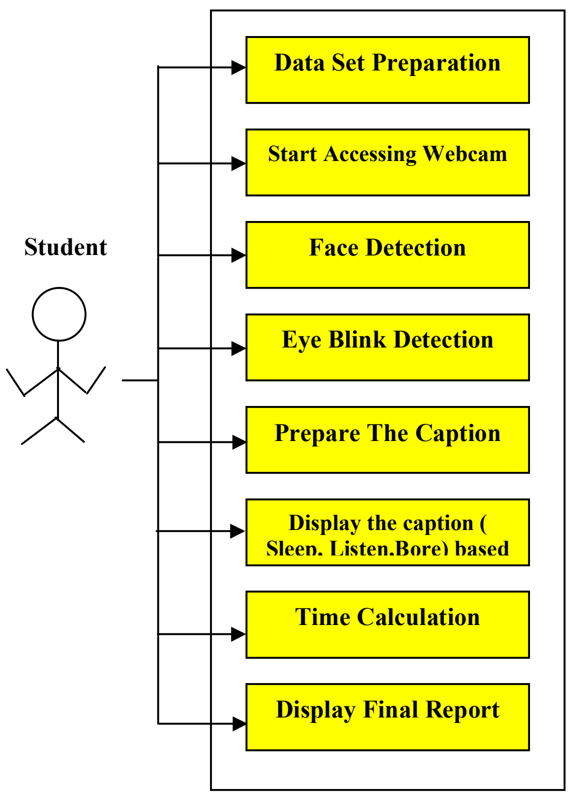
Facial expression is one of the most natural nonverbal ways for expressing human emotions and intentions. In recent years, automatic facial expression recognition (FER), which aims to analyze and understand human facial behavior, has become an increasingly active research topic in the domains of computer vision, artificial intelligence, pattern recognition, etc. This is because FER has many potential applications such as human emotion perception, social robotics, human- computer interaction and healthcare [1]–[5].
Inspired by the strong feature learning ability of deep neural networks, this paper proposes a new deep neural network- based FER method in video sequences by using a hybrid deep learning model. Our hybrid deep learning model contains three deep models. The first two deep models are deep Con- volutional Neural Networks (CNNs) [16], including a spatial CNN network processing static facial images and a tem- poral CNN network processing optical flow images. These two CNNs are separately used to learn high-level spatial features and temporal features on the divided video segments. The third deep model is a deep fusion network built with a Deep Belief Network (DBN) [17] model, which is trained to jointly learn a discriminative spatio-temporal segment-level feature representation. When finishing the joint training of a DBN, an average-pooling is applied on all the divided video segments to produce a fixed-length global video feature representation. Then, a linear Support Vector Machine (SVM) is adopted to perform facial expression classification tasks in video sequences.

This paper proposes a hybrid deep learning model, which consists of the spatial CNN network, the temporal CNN network, and the DBN fusion network, to apply for FER in video sequences. We implement our proposed method in two stages. (1) We employ the existing VGG16 model pre-tained on ImageNet data to individually fine-tune the spatial CNN network and the temporal CNN network on target video- based facial expression data. (2) To deeply fuse the learned spatio-temporal CNN features, we train a deep DBN model to jointly learn discriminative spatio-temporal features. Exper- iment results on three public video-based facial expression datasets, i.e., BAUM-1s RML, and MMI, demonstrate the advantages of our proposed method.

In future, we will extend our work to practical applications. For instance, it is challenging to develop a real-time FER system based on our proposed method. In addition, it is also interesting to explore deep compression of deep models so as to reduce the large network parameters of deep models.

.III.METHODOLOGY

1. Video Capture Module:
   - *Purpose:* Capture video frames from the webcam.
   - *Functionality:* Access the webcam feed and retrieve individual frames for analysis.
2. Face Detection Module:
   - *Purpose:* Identify the presence and location of faces in each frame.
   - *Functionality:* Employ face detection algorithms or models to locate faces, typically using bounding boxes.
3. Facial Landmark Detection Module:
   - *Purpose:* Identify key facial landmarks, such as eyes, nose, and mouth.
   - *Functionality:* Use algorithms or models to detect and locate specific points on the face, providing more detailed information for analysis.
4. Eye Gaze Tracking Module:
   - *Purpose:* Track the direction of the eyes to gauge attention.
   - *Functionality:* Analyze the position of the eyes over time to determine where the student is looking, allowing inference about their focus of attention.
5. Attention Assessment Module:
   - *Purpose:* Evaluate the level of attention or engagement.
   - *Functionality:* Combine information from face detection, facial landmarks, and eye gaze tracking to create a metric or model for assessing attention. This may involve analyzing factors like blink rate, head orientation, and eye movement patterns.
6. Visualization Module:
   - *Purpose:* Present the analyzed data in a human-readable format.
   - *Functionality:* Create visualizations, such as charts or graphs, to convey information about student attention levels over time.



The current landscape of online education lacks comprehensive tools for monitoring student attention and engagement during virtual classes. While educators can gauge participation through basic attendance tracking, there is a notable absence of real-time mechanisms to assess individual attentiveness. Traditional methods heavily rely on manual observation or intermittent quizzes, which may not provide a nuanced understanding of student engagement throughout the entire duration of an online class. Consequently, the existing system often faces challenges in promptly identifying and addressing lapses in attention, hindering the ability to create a truly interactive and engaging virtual learning environment.

Limited Real-Time Monitoring: The existing system lacks real-time monitoring capabilities, relying on intermittent methods such as manual observation or periodic quizzes. This limitation hinders the immediate identification of students' attention lapses during the entirety of online classes, potentially allowing disengagement to go unnoticed.
Subjective Assessment: Current methods for evaluating student engagement are often subjective and may not capture nuanced indicators of attentiveness. Manual observation, in particular, introduces the potential for bias, making it challenging to obtain a standardized and objective measure of student attention.
Inability to Detect Subtle Cues: Traditional approaches may struggle to detect subtle cues indicative of distraction or disengagement, especially when students employ passive behaviors such as looking away from the screen or multitasking. These nuanced signs are crucial for a comprehensive understanding of student attentiveness.
Lack of Automated Alerts: The existing system typically lacks automated alert mechanisms to notify educators and students in real-time when signs of reduced attention are detected. This absence of immediate feedback may delay interventions, diminishing the system's effectiveness in promoting active participation and engagement during online classes.

The proposed Student Attention Monitoring System revolutionizes online education by introducing a sophisticated solution that utilizes webcam technology and computer vision algorithms. Addressing the drawbacks of the existing system, this innovative approach enables real-time monitoring of student attention and engagement during online classes.

By analyzing facial features and eye movements, the system can detect subtle cues indicative of distraction, providing a nuanced and objective measure of attentiveness. Automated alerts are incorporated to promptly notify both educators and students when signs of reduced attention are identified, facilitating timely interventions and fostering a dynamic and engaged virtual learning environment.

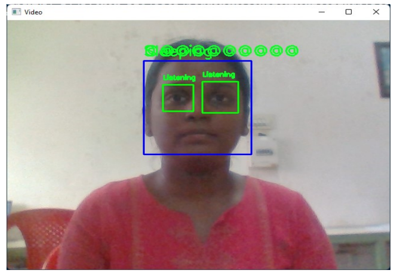
The proposed system not only overcomes the limitations of traditional methods but also introduces a proactive and automated approach to enhance the overall quality of online education by ensuring continuous, real-time assessment of student attentiveness.

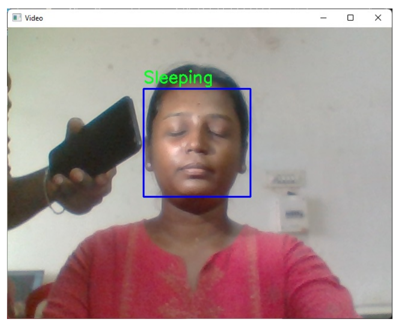
ADVANTAGES

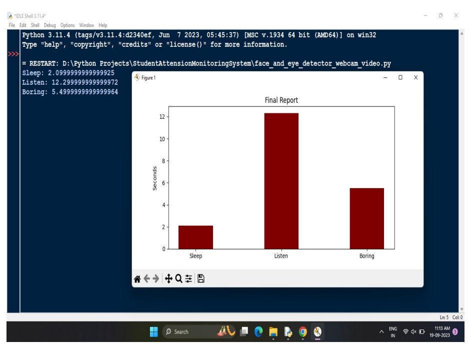
Real-Time Recognition: One of the primary advantages of the proposed Student Attention Monitoring System is its ability to provide real-time feedback on student engagement during online classes. By continuously analyzing webcam data and employing computer vision algorithms, educators can receive immediate alerts, allowing for timely interventions to re-engage students who may be distracted or disengaged.
Objective Assessment: The system introduces objectivity to the assessment of student attentiveness by leveraging computer vision algorithms to analyze facial features and eye movements. This ensures a standardized and impartial measure of attention, mitigating the subjectivity and potential biases associated with manual observation.
Nuanced Detection of Distraction: The proposed system excels in detecting subtle cues indicative of distraction or disengagement that may be overlooked by traditional methods. By analyzing facial expressions and gaze patterns, the system provides a more nuanced understanding of student attention, enabling educators to tailor interventions based on specific behaviors.

Automated Alert Mechanism: Automated alerts generated by the system offer a proactive approach to student engagement. Educators and students receive immediate notifications when signs of reduced attention are detected, enabling timely interventions and fostering a responsive virtual learning environment. This automated feedback loop enhances the overall effectiveness of online teaching strategies and promotes active student participation.

EXPERIMENTAL RESULT







REFERENCES

[1] Y. Alotaibi, M. Noman Malik, H. Hayat Khan, A. Batool, A. Alsufyani, and S. Alghamdi, "Suggestion mining from opinionated text of big social media data," Computers, Ma- terials & Continua, vol. 68, no. 3, pp. 3323–3338, 2021.

[2] S. A. Hussain and A. S. A. Al Balushi, A Real Time Face Emotion Classification and Recognition Using Deep Learning Model, IOP Publishing Ltd, Bristol, UK, 2019.

[3] C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - Journal of ELECTRICAL ENGINEERING, Vol.63 (6), pp.365-372, Dec.2012.

[4] C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.

[5] C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor & Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.

[6] C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.

[7] C.Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. "Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter" Journal of VLSI Design Tools & Technology. 2022; 12(2): 34–41p.

[8] C. Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya "Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay" Asian Journal of Electrical Science, Vol.11 No.1, pp: 1-8, 2022.

[9] C. G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash "Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756

[10] G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, "Perfromance Investigation of T-Source Inverter fed with Solar Cell" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749

[11] C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007

[12] M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022

[13] M Suganthi, N Ramesh, CT Sivakumar, K Vidhya, "Physiochemical Analysis of Ground Water used for Domestic needs in the Area of Perundurai in Erode District", International Research Journal of Multidisciplinary Technovation, pp: 630-635, 2019

[14] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG2017), Washington, DC, USA, May 2017.

[15] S. Zhang, X. Pan, Y. Cui, X. Zhao, and L. Liu, "Learning affective video features for facial expression recognition via hybrid deep learning," IEEE Transactions on Multimedia, vol. 7, 2019.

[16] H. Hayat Khan, M. Noman Malik, Y. Alotaibi, A. Alsufyani, and S. Alghamdi, "Crowdsourced requirements engineering challenges and solutions: a software industry perspective," Computer Systems Science and Engineering, vol. 39, no. 2, pp. 221–236, 2021

[17] C. Fabian Benitez-Quiroz, R. Srinivasan, Q. Feng, Y. Wang, and A. M. Martinez, "Emotionet challenge: recognition of facial expressions of emotion in the wild," 2017, org/abs/1703.01210.