International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 08 | Aug - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

AI based Online Proctoring Remote Monitoring Intruder, Emotion Detection and Distance Estimation

Subrahmanyam BHVSP #,
#B.Tech, B.Sc (Student), Department of Computer Science,
IIT Madras,
Chennai -600036, India,
Pavansubbu03@gmail.com

*Chandrasekhar Balabhadrapatruni**
**M.Sc.M.Tech.MBA, CEO, Magnocode Tech Pvt.Ltd.*
A leading Software Company ,
Hyderabad, India,500068
Bchsekhar65@gmail.com

***Abstract*— AI based online learning has undeniably surged in popularity over recent years. The COVID-19 pandemic has further accelerated the transition to online education, heightening the need for secure methods to authenticate and proctor online students. Today, a range of technologies offers varying levels of automation. In this paper, we present a comprehensive analysis of a specific solution that integrates multiple automated authentication technologies with an automatic proctoring system. The parameters that we used to achieve our goal are face detection, eye gaze tracking, multiple person detection, emotion detection, distance estimation and background noise detection. All these components help to maintain the exam integrity and to mitigate the existing limitation of e-exam proctoring software's.**

***Keywords*—** AI based online learning, Exam remote Monitoring integrity, Gaze tracking, Intruder, distance estimation.

I. INTRODUCTION

The advent of online learning platforms marks a transformative era in education, offering unprecedented flexibility and accessibility. The migration from traditional classrooms to virtual spaces has redefined the dynamics of assessment, placing the responsibility for maintaining the integrity of examinations squarely in the hands of educators and institutions. This research endeavours to delve into the multifaceted dimensions of online exam cheating, seeking to comprehend its trends, motivations, and countermeasures.

The surge in online education has fundamentally altered the learning landscape, necessitating an adaptation of evaluation techniques for digital environments [1]. While the transition to online exams brings forth numerous advantages, it concurrently exposes vulnerabilities that compromise the credibility of assessments. This study endeavours to unravel the complexities surrounding online examination cheating by addressing issues related to current trends, underlying motivations, various cheating strategies, and the most effective approaches to detection and prevention.

Our motivation stems from the observed dearth of a comprehensive literature review and classification in this field, hindering educators, institutions, and policy makers from accessing valuable insights. We aspire to contribute a refined understanding of the publication trends surrounding cheating in online examinations, the psychology of cheating motivations, the tactics employed, and the procedures for detection and prevention [2], [5], [6]. This endeavour enables us to systematically map the landscape of online exam cheating, particularly crucial in the development of proactive measures aimed at enhancing security in online assessments and fortifying integrity within e-learning.

This investigation is poised to employ rigorous research methods, underscoring our commitment to acquiring new knowledge that will empower education stakeholders to effectively navigate challenges associated with online exam cheating. Additionally, our research aims to contribute to the ongoing discourse on upholding the sanctity of assessments within the dynamic landscape of internet-based education, offering a comprehensive analysis of academic honesty levels in today's digital classrooms.

TABLE 1. COMPARISION OF EXISTING AND PROPOSED PROCTORING SYSTEM FEATURES COMPARISON

Features	ProctorU	Kryterion	Mercer Mettle	Proposed System
Physical Proctor for Monitoring	*Yes*	*Yes*	*Yes*	*No*
Usage of Webcam	*Yes*	*Yes*	*Yes*	*Yes*
Voice Recognition	*No*	*No*	*No*	*Yes*
Emotion Recognition	*No*	*No*	*No*	*Yes*
Distance Estimation	*No*	*No*	*No*	*Yes*

II. RELATED WORK

The existing body of research includes extensive work in the field of online exam proctoring systems. Each study incorporates various components designed to detect cheating during e-exams. A solution incorporating various biometric technologies, such as typing recognition, facial recognition, and voice detection and recognition, along with an automatic proctoring system that includes system workflow and AI

algorithms, aims to achieve 100% accuracy [3]. Another online proctoring system uses deep learning to monitor physical locations continuously, eliminating the need for a physical proctor. It employs biometric techniques like face recognition with the HOG face detector and OpenCV algorithm, and incorporates eye-blinking detection to identify stationary images. The software-based system, evaluated with the FDDB and LFW datasets, achieves up to 97% accuracy for face detection and 99.3% for face recognition [4]. A convolutional neural network (CNN) framework was developed for designing real-time CNNs. To validate the models, the author created a real-time vision system that simultaneously performs face detection, gender classification, and emotion classification using the proposed CNN architecture. After evaluation, the system achieved accuracies of 96% on the IMDB gender dataset and 66% on the FER-2013 emotion dataset [5]. Prathish, S. et al. proposed a system that detects the examinee's face and extracts feature points to estimate their head pose. Misconduct is identified by monitoring variations in yaw angle, the presence of audio, and the active window capture [6]. Y. Atoum et al. proposed a two-phase online exam process: preparation and exam phases. In the preparation phase, the test taker authenticates using a password and face recognition. During the exam phase, the Online Exam Proctoring (OEP) system monitors for real-time cheating, using multimedia analytics and hardware like a webcam, wearable camera, and microphone. The system's six core components track user verification, text detection, voice detection, active window detection, gaze estimation, and phone detection [7].

III. PROPOSED APPROACH

In this work, we aim to create a multimedia analysis system that can identify different types of cheating during online exams. Our system uses monitoring tools like webcams to observe the candidate's behaviour, including facial movements, system interactions, and audio signals. We use six components to detect cheating during exams: face detection, eye gaze tracking, multiple person detection, emotion detection, distance estimation, and background noise detection. The webcam captures input for the first five components, while noise detection is performed using an audio input device, such as a microphone.

A. Face Detection and Recognition using MTCNN

We implement Face Detection using Multi-Task Cascaded Convolutional Networks [10]. We start by evaluating the effectiveness of our proposed hard sample mining strategy. We then compare our face detection and alignment methods with leading techniques using the Face Detection Data Set and Benchmark (FDDB) [8], which contains annotations for 5,171 faces in 2,845 images. Finally, we assess the computational efficiency of our face detector. Our CNN detectors are trained using three tasks: *classifying faces and non-faces, performing bounding box regression, and localizing facial landmarks*.

B. Eye Tracking

Eye tracking can be implemented using a face landmark model that identifies the eyes based on coordinates from facial landmarks. This model tracks and records the movements of these coordinates, outputting the position as an integer. A pre-processing function is applied to threshold images, taking the image as a Numpy array [12]. This function compares the endpoints with the actual eye point values from the processed image to determine if the gaze is centred, looking left, or looking right. The function returns a value between 0.0 and 1.0 for the horizontal gaze direction, with 0.0 indicating the extreme right, 0.5 the centre, and 1.0 the extreme left. Similarly, for the vertical direction, the function returns a value between 0.0 and 1.0, where 0.0 represents the extreme top, 0.5 the centre, and 1.0 the extreme bottom.

C. Emotion-Detection

Our next component is a fully convolutional neural network architecture consisting of four residual depth-wise separable convolutions, with each convolution followed by batch normalization and a ReLU activation function [5]. The final layer uses global average pooling and a softmax activation function to generate predictions. We present the confusion matrix results for our emotion classification mini-Xception model, noting some common misclassifications such as predicting "sad" instead of "fear" and "angry" instead of "disgust." We also validated this model on the FER-2013 dataset, which contains 35,887 grayscale images, each classified into one of the following categories: "angry," "disgust," "fear," "happy," "sad," "surprise," and "neutral". The results indicate that the proposed models can be stacked for multi-class classification while still maintaining real-time inference capabilities.

D. Distance Estimation

Another feature of our proctoring software is a distance estimation network developed using Tensor Flow, trained on four NVIDIA TITAN Xp GPUs, each equipped with 12GB of memory [13]. This convolutional neural network generates depth maps from single RGB images by leveraging recent advancements in network architecture and high-performance pre-trained models. A well-constructed encoder, initialized with meaningful weights, can surpass state-of-the-art methods that either rely on expensive multistage depth estimation networks or require the design and integration of multiple feature

encoding layers. To estimate distance from the camera, we used the same assumption of the focal length of the camera to be 640mm, we have used assumed the width of an average human face to be 15cm, using these assumptions, we have initially detected the face using the MTCNN model, then calculation of the distance of the subject is done through the given formula.

Distance from Camera-

*(Face Width * Focal Length)/(Face Width in Pixels)*

E. Head Pose Estimation Using Deep Architectures

When the code developed for Head Pose Estimation is executed, a window opens and begins capturing real-time webcam footage. Each frame from the live feed is processed, and the output is determined based on the angle values. The possible outputs are "Looking Left", “Looking Right”, "Looking Centre". To achieve head pose detection, we used the reliable shape_predictor_68_face_landmarks.dat model file, which is a standard component of the DLib library in Python. This model identifies 68 facial landmarks, from which we use the nose tip, chin, left eye corner, right eye corner, mouth left corner, and mouth right corner as key coordinates to calculate accurate head pose. The camera's focal length is assumed to be 640mm, which is an estimated average for most webcams in use. We also assume that there is no lens distortion, which is generally the case. Based on the images, we obtain rotation vectors and translation vectors, from which we calculate the pitch, yaw, and roll values. These values are essential for accurately determining the head pose. Figure 1 represent how the proposed system display the output based on the angle of yaw, roll and pitch.

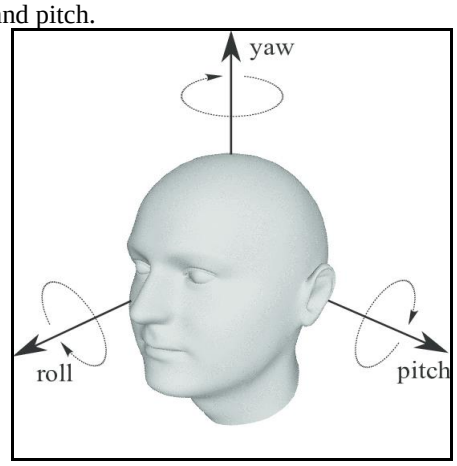

Fig. 1. Head pose estimation

F. Multiple-Person Detection

Identifying the presence of multiple individuals and mobile phones is an essential aspect of the webcam proctoring process. The MTCNN module continuously receives frames from the webcam, storing the face detected in the first frame as the reference "student frame” [10]. In each subsequent frame, the detected face is compared to this student frame to determine if a new person has appeared or if there is potential for cheating. The MTCNN module is capable of detecting multiple faces in a single frame. Head detection and multiple person detection are run on a single thread in parallel with other modules. This can also be accomplished using a YOLOv3 model, which utilizes pre-trained YOLOv3 weights. The model requires input parameters, such as the number of filters for the convolutional layer and the name of the layer. The YOLO convolutional layer is then configured using Dark-Net functions and produced as the output.

G. Background Noise Detection

One of the most common cheating behaviours in online exams is seeking verbal assistance from another person in the same room or remotely through a phone call. Noise Estimation module captures audio through a microphone in chunks, processes the audio data to determine its amplitude, and compares it against a predefined threshold [9]. Which is set to normal talking volume. If the detected audio level exceeds this threshold, it is classified as noise, indicating that there is significant background sound. This helps us classify behavior as cheating/non-cheating

IV. METHODOLOGY

Our proposed methodology involves seven components: Face Detection, Multiple Faces Detection, Head Pose Detection, Emotion Detection, Eye-Gaze Tracking, Distance Estimation, and Noise Detection. The Face Detection and Multiple Faces Detection components run on a single MTCNN process. The face detected at the beginning of the examination session is labeled as the student, and any new faces detected later are labeled as intruders. Based on these MTCNN face coordinates, the width of the face is estimated in pixels. We assume that the student's face is around 14-15 cm wide and that the camera's focal length is approximately 600mm with a resolution of 720p.

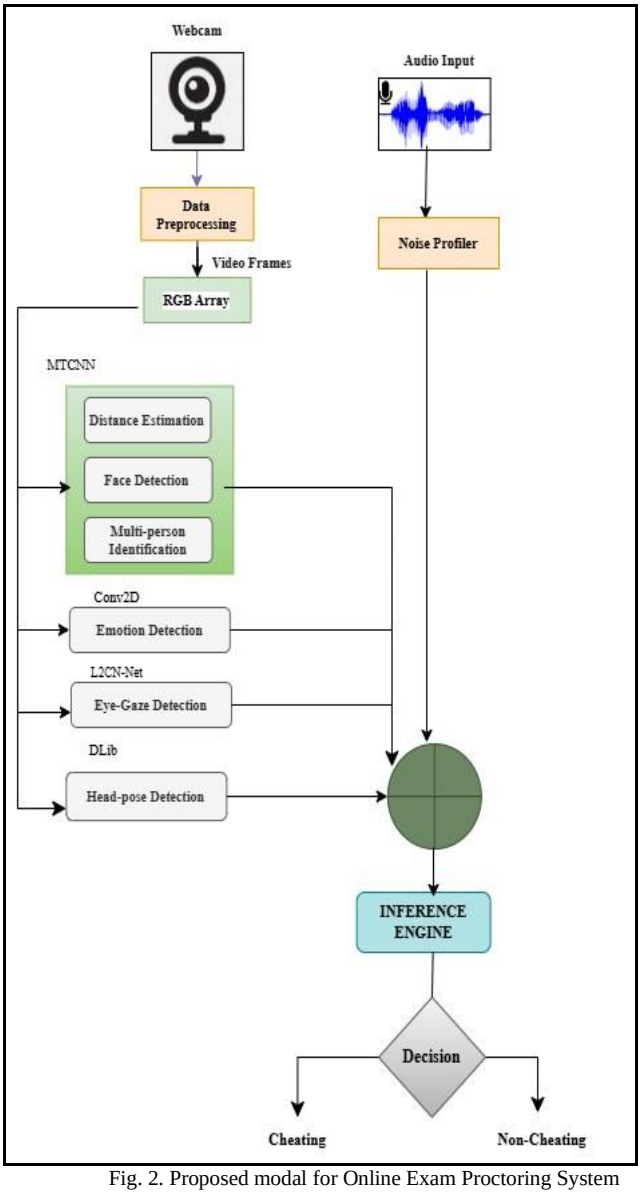

Fig. 2. Proposed modal for Online Exam Proctoring System

Based on these parameters, the distance of the student's face from the camera is calculated in real time on a separate process thread. Head Pose is estimated using the DLib library with the standard 68 facial landmarks model to determine the pitch, yaw, and roll values, also on a separate process thread. Emotion detection, with seven labels, is performed on another CPU thread after loading an emotion detection model. All these inputs are retrieved from the webcam after pre-processing. The audio input is retrieved from the microphone and fed to a Noise Profiler process that checks the sound levels in real time in decibels (dB) to evaluate background noise for signs of cheating. The outputs from all seven components are fed into an inference engine to classify the given time frame as either cheating or non-cheating.

All these components run on various CPU threads in parallel to enhance the overall system's efficiency. While this results in higher resource utilization, the student has no other tasks except writing the examination, which is why this approach has been devised.

V. Experimental Results

We start by detailing the evaluation process. Then, we examine the performance of several key components of our proposed system individually. Afterward, we test the overall performance of the entire system. Lastly, we address the effectiveness of the proctoring system. The various modules utilized in our proctoring system software include:

- Audio Read: Captures audio inputs from the microphones in chunks for sound detection processing.
- Gaze-Tracking: Monitors the student's eye movements to determine if they are looking left, right, or straight ahead.
- Numpy: Processes images as RGB a Numpy arrays, which are used as inputs for all sub processes.
- OpenCV: Handles video capture from the webcam, image processing, and the display of text and rectangles on frames.
- DLib: Detects facial landmarks.
- Keras: Runs deep learning models.
- Tensor Flow: Serves as the backend for Keras.
- Parallel: Executes all features concurrently across different CPU cores.
- Tkinter: Manages the system's graphical user interface.

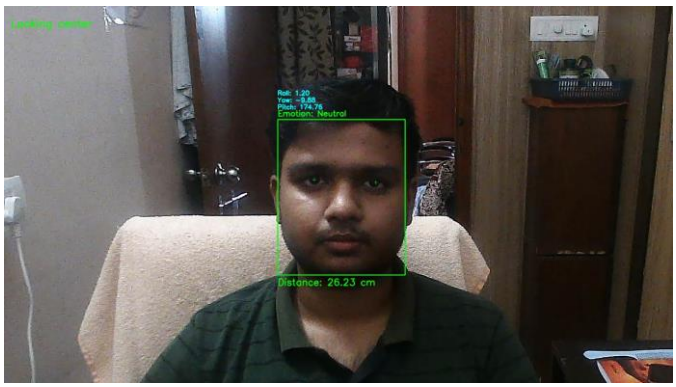

Fig. 3. Results of the provided combined Head-pose, Distance Estimation, Eye Gaze Tracking, Background Noise detection, Multiple Person detection, Face and emotion detection.

Figure 3 shows the resulting screen, efficiently presenting all seven features in a single view. In the following section, we provide additional outputs. These interpretable results have

provided insights into the cheating behaviors linked to all the discussed types of cheating components. The figure below illustrates all the different types of cheating components.

Fig. 4. Proctoring system software captured and resulted image examples.

CONCLUSIONS

There is always the potential to develop better models using improved datasets, which can enhance performance in practical applications. Additionally, the field of Artificial Intelligence is continually evolving, providing ongoing opportunities to explore new methods for implementing these features through research. Research can involve comparing various models built on the same dataset to identify the most accurate one. There is also potential to enhance the frame rate of the live webcam feed obtained using OpenCV.

REFERENCES

[1] Nigam, A., Pasricha, R., Singh, T., & Churi, P. (2021). A systematic review on AI-based proctoring systems: Past, present and future. Education and Information Technologies, 26(5), 6421–6445. https://doi.org/10.1007/s10639-021-10597-x

[2] Kaddoura, S., & Gumaei, A. (2022). Towards effective and efficient online exam systems using deep learning-based cheating detection approach. Intelligent Systems with Applications, 16, 200153.

[3] Labayen, M., Vea, R., Flórez, J., Aginako, N., & Sierra, B. (2021). Online student authentication and proctoring system based on multimodal biometrics technology. IEEE Access, 9, 72398-72411.

[4] I Ahmad, F AlQurashi, E Abozinadah and R Mehmood, "A Novel Deep Learning-based Online Proctoring System using Face Recognition Eye Blinking and Object Detection Techniques", *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol. 12, no. 10, 2021.

[5] Arriaga, O., Valdenegro-Toro, M., & Plöger, P. (2017). Real-time convolutional neural networks for emotion and gender classification. arXiv preprint arXiv:1710.07557.

[6] Prathish, S., & Bijlani, K. (2016, August). An intelligent system for online exam monitoring. In 2016 International conference on information science (ICIS) (pp. 138-143). IEEE.

[7] Y Atoum, L Chen, AX Liu, SDH Hsu and X Liu, "Automated Online Exam Proctoring", *IEEE Transactions on Multimedia*, pp. 1-1, 2017.

[8] V. Jain, and E. G. Learned-Miller, "FDDB: A benchmark for face detection in unconstrained settings," Technical Report UMCS-2010-009, University of Massachusetts, Amherst, 2010.

[9] Irfan, M., Aslam, M., Maraikar, Z., Jayasinghe, U., & Fawzan, M. (2021, December). Ensuring academic integrity of online examinations. In 2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS) (pp. 295-300). IEEE.

[10]Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE signal processing letters, 23(10), 1499-1503.

[11]Harish, S., Rajalakshmi, D., Ramesh, T., Ram, S. G., & Dharmendra, M. (2021). New features for webcam proctoring using python and opencv. Revista Geintec-Gestao Inovacao E Tecnologias, 11(2), 1497-1513.

[12] GazeTracking/gaze_tracking/eye.py at master- antoinelame / GazeTracking – GitHub

[13]Alhashim, I., & Wonka, P. (2018). High quality monocular depth estimation via transfer learning. arXiv preprint arXiv:1812.11941.