Advances in Artificial Intelligence and Machine Learning; Research 5 (2) 3646-3662 Received 19-01-2025; Accepted 17-04-2025; Published 23-04-2025

Advanced Online Proctoring: Facial Emotion Monitoring with Attentive-Net

Sangeeta Lamba army.sangeeta46@gmail.com
Research Scholar, Department of Computer Science,
Banasthali Vidyapith
Jaipur, Rajasthan, India, 304022.

Neelam Sharma neelamsharma@banasthali.in
Associate Professor, Department of Computer Science,
Banasthali Vidyapith
Jaipur, Rajasthan, India, 304022.

Corresponding Author: Sangeeta Lamba

Abstract

The Attentive Proctoring System proposed in this paper addresses the growing need for reliable remote examination solutions amid the global shift toward online learning. Traditional methods of human proctoring are often constrained by scalability issues and resource demands, rendering them inefficient in the face of large-scale online assessments. Leveraging advanced deep learning techniques, our framework aims to ensure exam integrity through a multi-phase approach. By preprocessing video frames captured from students' webcams and employing techniques such as background subtraction and face detection with YOLOv7-SGCN, we establish a robust foundation for identifying potential irregularities. YOLOv7-SGCN is ideally suited for real-time applications since it offers reliable and effective identification of questionable activity with no processing overhead. However, Attentive-Net improves attention-based feature learning, increasing the precision of recognizing tiny behavioral cues. Further enhancing security measures, our system integrates multi-modal liveness detection and head pose estimation, providing comprehensive monitoring capabilities. Emotion detection, facilitated by a Faster R-CNN, enables the identification of unauthorized aids like mobile devices or books. The integration of Attentive-Net allows for dynamic focus adjustment based on various component outputs, ensuring a thorough examination of pertinent areas within the image. With mechanisms in place for alert and intervention, our system offers a proactive approach to maintaining exam integrity, thereby fostering trust and confidence in the online examination process.

Keywords: Proctoring system, Face detection, Multi-Modal liveness detection, Head pose estimation, and Emotion detection.

Citation: Sangeeta Lamba and Neelam Sharma. Advanced Online Proctoring: Facial Emotion Monitoring with Attentive-Net. Advances in Artificial Intelligence and Machine Learning. 2025;5(2):207.

1. INTRODUCTION

The practice of supporting instructors and students in teaching and learning around the globe through online resources is known as distance learning. With the use of technology and online learning materials, students can successfully acquire self-directed learning skills. They might decide what they need to learn, find and use online resources, use the knowledge to complete assignments in class, take tests, and even evaluate the feedback they receive as a consequence [1–3]. One of the tools used in online education to identify exam cheating and other student infractions during remote learning is a facial recognition system. Through ID verification services, a facial recognition technology is frequently utilized to verify users [4–6]. It functions by recognizing and quantifying the features of a student's face in an image or video. It does this by comparing a human face from an electronic photo or a video frame with a face database. Because they are more engaged than students using a more traditional method, students using distant learning technology spend more time on foundational learning tasks. Lecturers can monitor students through distance learning to ensure they fulfill their potential and do well on tests without engaging in dishonest behavior [7–9].

Even if distance learning has many benefits, it still has several drawbacks, like the inability to identify students' emotions, which are crucial to their performance. Because of the distance separating them, instructors in online courses are unable to relate to their students' feelings or even the issues they are facing [10, 11]. Additionally, they currently lack the mechanisms that raise the legitimacy of online tests. Due to the importance of this subject, a variety of approaches and techniques—such as face reader and X press engine—have been developed to assist instructors in precisely identifying the feelings of their students. Convolutional neural networks (CNN) have been instrumental in the development of numerous effective artificial intelligence algorithms, particularly in the area of deep learning, which have gained popularity in the computer vision sector [12, 13]. This study provides important new approaches to distance learning, which has gained popularity recently. Many have questioned the efficacy and efficiency of the system, even though many educational institutions use remote learning as an alternative to traditional classroom instruction. This study's main findings include the absence of a method to identify and confirm students' identities during exams, online courses, and remote learning [14, 15]. This is on top of the absence of a recognizable method to identify attempts by students to cheat on online exams and quizzes. Scalability, computing efficiency, and institutional policies must all be carefully considered before implementing an AI-driven proctoring system on a broad scale. While on premise systems provide more control over data protection, cloud-based deployments provide flexibility and scalability, enabling dynamic resource allocation for thousands of concurrent users. Adoption depends on low latency performance, real-time processing, and adherence to privacy laws. Infrastructure costs, AI model training and maintenance, operational costs for server maintenance and human supervision, and faculty and student training requirements are all factors that affect costs. On-premise configurations have greater upfront costs but lower ongoing fees, while cloud-based solutions offer scalability but may result in long-term costs. There are extra difficulties in integrating smoothly with learning management systems (LMS) like Moodle, Canvas, and Blackboard. It is crucial to guarantee safe authentication, API interoperability, and compliance with data protection laws such as FERPA and GDPR. The system must support students with disabilities and provide a seamless user experience across devices. Institutions can also need customization options for exam settings, reporting capabilities, and monitoring criteria. By resolving these issues, the suggested system will become more feasible, affordable, and accessible, increasing the likelihood that it will be

successfully implemented in actual educational settings. Additionally, past studies have indicated that these problems pose serious barriers to distance learning. The goal of this research is to alleviate these issues by creating a new system that helps instructors oversee and manage students during online classes and tests using computer vision and deep learning methods. Facial characteristics of the students are detected, measured, and output by the system. The primary contributions of the paper are as follows,

- The proposed YOLOv7-SGCN model enhances face detection accuracy by incorporating Spatial Pyramid Pooling (SPP) and Graph Convolutional Networks (GCNs). Additionally, self-supervised learning for landmarks improves face alignment, which is crucial for the accurate analysis of facial features.

The organization of the paper is as follows, section 2 discusses the recent existing papers related to the students' misbehavior in online exams, section 3 gives a clear explanation of the proposed methodology, section 4 discusses the results obtained for the proposed model with the existing techniques and section 5 gives a clear conclusion.

2. LITERATURE REVIEW

In 2023, Ferdosi, et al. [16], performed the process of modeling and categorizing the behavioral patterns exhibited by students taking an online exam. The article outlined a proctoring system for online pen-and-paper exams. Specifically, it examined the examinee's head, eye, and lips frame-by-frame, attempted to discern any patterns in their movements, computed chunk scores (100 chunk), and computed a concluding cheating score for the entirety of the examination. To obtain the coordinates of the chosen face landmarks, used the MediaPipe package. K-NN, the top-performing ML model for orientation organization, was then applied.

In 2022, Kaddoura and Gumaei [17], suggested an approach to deep learning-based cheating detection that leads to the development of efficient and successful online exam systems. With the use of deep learning models, the suggested research seeks to create an efficient and successful method for online exam systems that detects cheating in real-time from speech and recorded video. Deep convolutional neural networks (CNNs) and the Gaussian-based discrete Fourier transform (DFT) statistical method automatically extract valuable information from audio and visual images, allowing it to identify and identify whether the examinee was cheating during the exam. In 2022, Mahmood, et al., [18] have made the development of a deep learning algorithm-based intelligent exam supervision system. Using deep learning techniques, namely Faster Regional Convolution Neural Network (RCNN), that work proposes an automated exam invigilation system. Multi-task Cascaded Convolutional Neural Networks, or MTCNN, were utilized for student identification about face detection and recognition. Faster RCNN was an emotion detection technique used to identify students' questionable behavior during exams based on head movements.

In 2024, Ndungu & Chepsergon [19], proposed a study that looks at how AI may both support and prevent academic dishonesty in higher education in Sub-Saharan Africa. Platforms for content creation encourage plagiarism and cheating, yet AI-powered plagiarism detection and proctoring work to discourage wrongdoing. Universities must make investments in AI training and infrastructure,

put AI-based integrity standards into place, and encourage digital literacy to guarantee that AI is used responsibly in the classroom to meet these issues. In 2024, Cholakov & Stoyanova-Doycheva [20], provided a study that improves the Fraud Detector using AI device agent in the Distributed eLearning Centre (DeLC), turning it from a simple fraud detection tool into a sophisticated system that uses ChatGPT to increase accuracy. The addition of ChatGPT considerably improves detection accuracy, according to the results. The system is nevertheless accessible for outside AI partnerships, guaranteeing its scalability and adaptability. In 2024, Sevnarayan & Maphoto [21], investigated a study looking at first-year second-language learners at an open-distance e-learning institution in South Africa who cheated in two English modules. It looks at the reasons why students cheat on online tests, the variables that affect cheating, and methods for minimizing misbehavior. Findings from qualitative techniques, including focus groups, professor interviews, and marker evaluations, show that cheating is common in distance learning, with students confessing to a variety of unethical behaviors. The study emphasizes the cognitive rationales that motivate immoral behavior and is based on the social cognitive theory of moral disengagement.

In 2023, Alsabhan [22], performed detection of student cheating in higher education, completing the use of LSTM and machine learning algorithms. With an accuracy of 90%, the suggested model outperformed all previous three-reference efforts. It employed a Long Short-Term Memory (LSTM) technique with dense layers, a dropout layer, and an optimizer named Adam. Accuracy gains are attributed to the use of hyper-parameters and a more complex, optimized architecture. Furthermore, the discussed methods for cleaning and preparing our data may have contributed to the improved accuracy. In 2023, Yulita, et al. [23], utilized deep learning to detect online exam cheating as one educational innovation in response to COVID-19. Utilizing HAR, the work employed a deep learning technique built on the MobileNetV2 framework. The information came from a webcam-captured video clip of a person completing an online test. After determining which model was best, the main goal of that research was to develop an online application in Indonesian.

In 2023, Banzon, et al. [24], utilized facial expression for identification in the classroom, raised ethical questions, and offered recommended procedures for affect detection. The purpose of the essay is to develop a typology of proactively reflexive ethical implications by using a Reflexive Principlism technique to track students' feelings using variations in their facial expressions. Using that approach, the authors make a distinction between applications in applied education and research, and they argue that the latter should be restricted until more is understood about the ethical implications of affective computing in educational contexts. In 2022, Nurpeisova et al. [25], studied mathematical models and methods for proctoring system-based facial recognition in images with Python. The paper examines the viability and logic of employing proctoring technology to identify pupils through remote monitoring of their academic progress. Face recognition technology is a component of proctoring technology. Facial recognition is a subfield of biometric and artificial intelligence. In 2022, Pang et al. [26], used the creation and execution of an anti-cheating monitoring system for worldwide Chinese online teaching exams. The fuzzy evaluation method served as the groundwork for the hybrid system that was discussed in that study, which was intended to identify and alert users to potentially suspicious behavior related to cheating in international Chinese online teaching assessments. The objective of hybrid technology used in international Chinese online instruction is to assist instructors in monitoring and assessing student performance to guarantee the validity and efficacy of the exam or evaluation findings and make sound decisions regarding the cheating behavior of international students in online assessments or exams. In 2022, Saraff and Tripathi [27], used facial expressions to deduce emotions require emotional intelligence. The connection between

accurately recognizing emotions and emotional intelligence (EQ) was covered in that research. The emotional intelligence of the participants was assessed online using the Schutte Self-Report Emotional Intelligence Test (SSETT). The capacity of participants to identify emotions from photos of people's faces was investigated using a Google Form. Due to COVID-19 regulations, only a restricted sample and no in-person data collection were done for that study. In 2023, Hossen and Uddin [28], utilized an XGBoost classifier, and students' attention was tracked throughout online classes. The method considers the various individual factors and contextual variations that influence how students react to the virtual learning environment. It uses face recognition technology for user verification and effortlessly incorporates essential features like posture estimation, hand tracking, facial detection, and mobile phone detection.

3. PROPOSED METHODOLOGY

The proposed approach starts with gathering and preparing student webcam video frames, then uses YOLOv7-SGCN and self-supervised learning algorithms for face detection and alignment. SRI-Net's multi-modal liveness detection improves security measures, and the head posture estimate gives information about the test-taker's attention span. The detection of objects, enabled by an optimized Faster R-CNN, guarantees the identification of unapproved assistance. Attentive-Net integration dynamically modifies emphasis according to different component outputs. Lastly, the integrity of the online examination environment is guaranteed by systems for timely alerts and intervention in the event of abnormalities. The goal of this all-encompassing strategy is to offer a solid remedy for upholding exam integrity in online learning environments. FIGURE 1 displays the block diagram for the suggested Online Proctoring System.

In the context of AI-driven proctoring, the following compares Faster R-CNN, EfficientDet, and Vision Transformers (ViTs) to YOLOv7-SGCN and Attentive-Net:

3.1 Data Collection and Preprocessing

To preprocess video frames taken with a student's webcam, the video is divided into individual frames, and every frame is resized to a consistent dimension. Additionally, pixel values are normalized to guarantee uniformity in a range of lighting conditions, and background subtraction techniques are applied to detect abrupt objects' appearance or disappearance.

3.1.1 Captured video frames

The suggested approach makes use of a camera to record and store the video of the examinee in real-time, then generate frames from that footage. Additionally, motion detection or modifications to the display are taken into account when extracting keyframes. Vulnerabilities can be discovered even in the absence of significant modifications to movement and display. Therefore, in this work, all the frames are processed rather than just the keyframes being extracted. The procedure that follows is used to recognize faces in each frame.

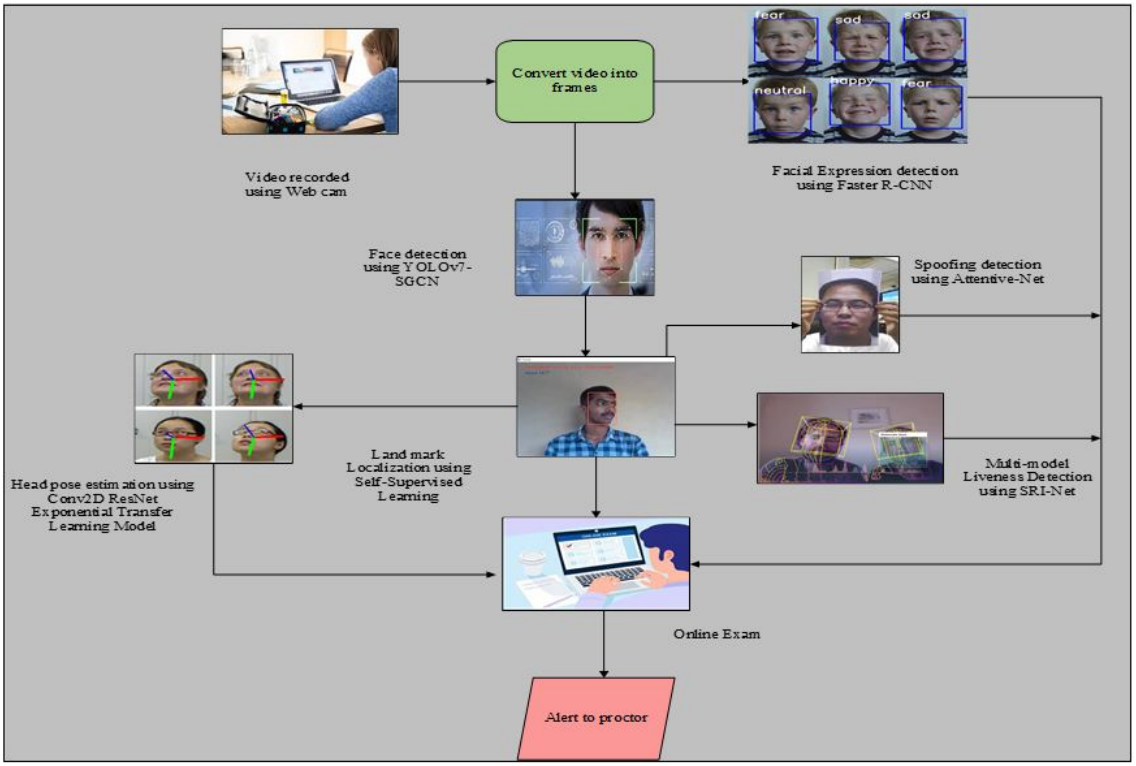


Figure 1: Block diagram of the Online Proctoring System

- **Frame Resizing:** To ensure efficient processing, resize frames to a consistent dimension. Resizing or cropping an image essentially means selecting a section of it and saving it as a fresh training instance. The cropping area may be chosen at random or by a preconceived scheme. To make an image square, its current dimensions must either be resized to fit inside a square or the current aspect ratio must be maintained while extra pixels are added to fill in the newly generated empty spaces.

- **Color Normalization using CLAHE:** For uniformity across lighting differences, normalize the values of the pixels. In CLAHE, the contrast limiting process needs to be applied to each neighboring pixel where the transformation function is produced. The transformation slope function is used to increase the contrast of a specific pixel value. This is proportionate to both the pixel-level histogram value and the cumulative distribution function (CDF) of the neighborhood slopes. Before assessing CDF, CLAHE [29], reduces amplification by histogram clipping at a predetermined value. The size of the adjacent zone establishes the clip limit or value at which histograms are clipped. Eq. (1) assesses the clip point:

$$\beta = \frac{M}{N}\left(1 + \frac{\alpha}{100} S_{max}\right) \tag{1}$$

In this case, the variables M, N, S_{max}, α, and clip factor denote the number of pixels in each block, block range, and maximum slope, respectively. The block's pixels stay constant when "α " approaches 0, resulting in M/N for the clipping point (CP). As "α" approaches 100, there is a noticeable rise in contrast. Conversely, CP is consequently a crucial component in

Table 1: Comparison of other Deep Learning models with the proposed technique

Model	**Strengths**	**Weakness**	**Reason for choosing Yolov7-SGCN and Attentive Net**
Faster R-CNN	• High accuracy due to detection based on location. • Excellent performance for obscured or small things.	• Computationally expensive. • High inference latency, unsuitable for real-time proctoring.	YOLOv7-SGCN offers real-time processing, while Faster R-CNN is too slow for live monitoring.
EfficientDet	• Accuracy and computation are balanced via efficient scaling • more effective than Faster R-CNN	• Slower than YOLO-based models. • May struggle with fine-grained behavior detection.	YOLOv7-SGCN specializes in fast human activity detection, whereas EfficientDet is optimized for general object detection.
Vision Transformers (ViTs)	• Strong feature learning with global attention. • Captures long-range dependencies for	• Requires large-scale data and high computational power. • Less efficient than CNN-based models for real-time tasks.	Attentive-Net retains attention mechanisms but is optimized for real-time efficiency, unlike ViTs, which are computationally demanding.

controlling improvement. The mapping function in CDF [30], is accomplished in Eqs. (2) and (3) to remap a block picture with grey levels as trails.

$$cdf\ (l) = \sum_{k=0}^{l} pdf(l) \tag{2}$$

$$T\ (l) = cdf\ (l) * l_{max} \tag{3}$$

Where $T(l)$ is the remapping function and l_{max} is the maximum pixel value permitted in a block. Several remapping functions about the CDF with the redistributed histogram in the block are obtained. Mapping functions break every pixel value to avoid artifacts. The pixel "p" is randomly divided into blocks, with "a," "b," "c," and "d" serving as the centers of each block. Bilinear interpolation is used to obtain the remapped "p" pixel, as shown in Eq. (4).

$$T\ (p\ (i)) = m.\ (n.T_a.p\ (i) + (1-n)\ .T_b.p\ (i)) + (1-m)\ .\ (n.T_c.p\ (i) + (1-n)\ .T_d.p\ (i)) \tag{4}$$

The remapping function $T(.)$ and the value of pixel "i" with coordinates (x, y) are represented by $p(i)$. Artifacts are eliminated by interpolation. Because blocks are processed, E-CLAHE achieves decreased computational complexity for improvement.

3.2 Face Detection and Alignment

This section introduces a YOLOv7-SGCN, a modified version of YOLOv7 that incorporates Spatial Pyramid Pooling (SPP) and Graph Convolutional Networks (GCNs) for advanced small emotion

detection SPP captures objects at multiple scales well, while GCNs between image regions improve face recognition Relationship capture enhances feature representation. For face matching, it adopts a self-supervised learning approach for landmark recognition, which allows the automatic learning of spatial relationships of faces without manual cues All these methods provide together accuracy and robustness of face recognition and matching, which is important for computer vision applications.

3.2.1 Face detection using YOLOv7-SGCN model

i) SPP layer

The YOLOv7 network is divided into two main parts by the SPP module. By increasing the perceptual field, the SPP component enables the algorithm to adjust to images with varying resolutions. This is accomplished by obtaining maximum pooling for various perceptual fields. Four maximum pooling processes—5, 9, 7, and 1—are applied to the pictures in the SPP module's first branch. Target size differentiation and improved localization of small targets are made possible by these four distinct maximum pooling layers. The target size is obtained by the SPP module after processing the input feature map four times, allowing it to locate and process the target in the maximum pooling channels of various scales.

ii) Backbone layer

The basis for YOLOv7's signature collection process is the Backbone framework. It makes it easier to extract features at the beginning for target detection, which leads to the creation of feature layers. These feature layers that were taken from the backbone are known as "effective feature layers" since they are crucial to the development of the network that follows. Surprisingly, YOLOv7's Backbone feature extraction network leverages the E-ELAN module, which is defined by a last stacking module with four branches. Because of the many stacks produced by this architecture, the residual structure is denser, which simplifies the optimization process and allows for improved accuracy through deeper network penetration.

iii) GCN

Initially, the incoming data is organized into a dimension matrix [m × n], where each column represents a randomly chosen feature. Give a straightforward technique for visualizing the associations between features using the feature correlation graph G. This technique will allow you to explore and understand the correlations between the selected traits effectively. As this work focuses on network assault detection for IDS rather than creating intricate node embedding graph models, get the correlation between feature columns using the conventional GCN method [31]. This graph embedding technology is fairly sophisticated and has been used to solve many graph-based challenges. Eq. (5) defines the GCN's layer structure.

$$H^{(l+1)} = f(H^{(l)}, A) \tag{5}$$

This instance uses A as the graph G adjacency matrix, $H^{(l)}$ as the feature element matrix, and $H^{(0)}$ as the $M * N$ dimensional input data. The multilayer GCN model with layer-by-layer transmission, as defined by Eq. (6), employs the graph's rapid approximation convolution.

$$f\left(H^{(l)}, A\right) = \sigma(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}\, H^{(l)}W^{(l)}) \tag{6}$$

The activation function is denoted by σ, the trainable weight matrix in the layer connections is $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$, the unit matrix is IN, and for a graph without direction G with extra self-connections, the adjacency matrix is $\hat{A} = A + I_N$. With n characteristics per sample, the batch size determines the value of m. A matrix of size $[m \times n]$ results from this. This matrix is in the form of $[m \times 1]$, with F_1, F_2, and F_n representing each unique feature column. These columns of extracted features taken together are called nodes and make up the dynamic network G. Every feature node undergoes this procedure once more, creating a collection of feature nodes denoted as $[f_1, f_2, ..., f_n]$. The resulting feature node f, which is the average of these feature nodes, more compactly depicts the $[m \times n]$ data matrix.

iv) Neck module

The crucial task of feature fusion on the previously created successful feature layers is carried out by the YOLOv7 Neck module. This module is a significant advancement in the YOLOv7 layer-extraction network, carefully designed to tackle the problems caused by different target sizes in deep learning. It also successfully addresses the problem of picture noise. Notably, the Panet structure from YOLOv7's previous series is still present. This entails an extra cycle of feature-down sampling to achieve thorough feature fusion in addition to extending the architecture's built-in features for enhanced synergy.

v) Feature Pyramid Network (FPN)

The essential elements of YOLOv7, which handle regression and classification, are called the YOLO Head as a whole. Notably, better effective feature layers are now contributed by the FPN and Backbone. The next step is to map the feature map to a set of feature points and then combine those points with feature frames. The YOLO Head is based on the strategic strategy of ensuring that every previous frame correlates to several feature channels. Essentially, the YOLO Head assesses the feature points, determining the relationship between the target entities and earlier frames. Similar to its earlier iterations, YOLOv7 does regression and classification using a 1×1 convolution with separated Heads. The whole YOLOv7 network processes the following tasks: input picture processing, feature extraction, feature enhancement, and prediction to anticipate object cases matching the prior frame.

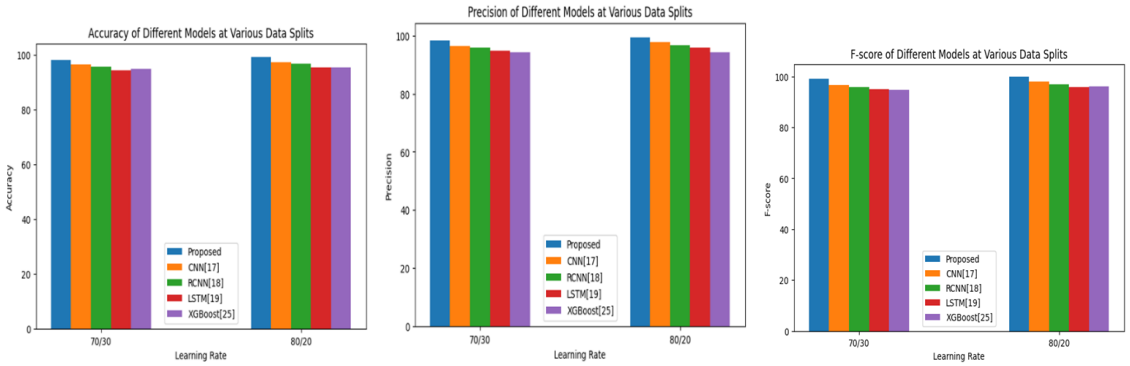
3.2.2 Face Alignment using Self-Supervised Learning

Self-supervised learning is based on the idea of first training a shallow network with sparsely annotated data and then training a network with a pretext task on a large-scale unlabeled dataset. As of right now, our feature extractor is self-supervised in its learning process. A small amount of annotated data is used to derive the final landmark prediction. After freezing, the feature extractor Ψ, a lightweight predictor, is trained over it. The output of Ψ is a landmark heatmap with an interval of $\in R^{\mathrm{H}\times\mathrm{W}\times\mathrm{K}}$, where K is the number of landmarks. The heatmap's predicted location of landmark k, when weighed, yields its final position $(\hat{x}^k, \hat{y}^k)$. With a l_2 loss, it is overseen by the landmark's annotated location (x^k, y^k).

3.3 Facial Emotion Detection Using Faster RCNN

Use Faster R-CNN, a state-of-the-art emotion recognition model known for its speed and accuracy. The fast R-CNN works by effectively proposing Regions of Interest (RoIs) in an image and then classifying these regions into facial emotions. This modeling approach is composed of a classification network to anticipate comments on individuals in this community, an RPN to generate ROIs, and a backbone network, usually a CNN, for feature extraction.

Face recognition using Faster R-CNN requires the weights previously trained in the Faster R-CNN algorithm to be optimized on the recognition dataset Through transfer learning, fig optimizes its properties to better recognize facial emotions. Optimization strategies such as optimizing backbone networks, improving field coefficients, and optimizing non-maximum constraints help improve accuracy and speed. In addition to training and evaluation, develop a model for real-time facial emotion recognition, which enables applications in emotion recognition, human-computer interaction, and emotion to be applied to evaluation. A deep learning model called Faster R-CNN is utilized for facial emotion recognition in images. Its accuracy and speed make a major improvement over earlier R-CNN models and Fast R-CNN. Similar to Selective Search, both methods required an extensive number of candidates bounding boxes or potential product regions and then feeding those bounding boxes to the detection network. FIGURE 2 shows the architecture of Faster R-CNN.



(a) (b) (c)

Figure 2: Comparison of the accuracy metric, Precision, and F-score.

3.4 Overall Model Adjustment Using Attentive-Net

Utilize Attentive-Net to dynamically adjust its focus on different parts of an image based on the outputs from face detection, and head pose estimation, face spoofing detection, and emotion detection components. The Face Detection module receives all of the observed frames as input. Face detection is used to identify the frontal areas in the frames that the camera captures. Using Attentive-Net to recognize the faces in the image and resize them into different scales results in the creation of a scale pyramid. A series of smoothing filters with a radius double that of the previous ones is then used to further smooth the image [32]. Equation mentions the deferrable smoothing filter f, which

is used for the smoothing Eq. (7).

$$\hat{s} = \sum_{m=-2}^{2} \sum_{n=-2}^{2} f\,(m,n)\,.s_0(i-m)(j-n) \tag{7}$$

Additionally, down sampling is used to smooth and cut the frame size in half.

$$s_{l+1}\,(i,j) = \sum_{m=-2}^{2} \sum_{n=-2}^{2} f\,(m,n)\,.s_l(2i-m)(2j-n) \tag{8}$$

Given that every image consists of a single face and numerous real-world landmarks, each image is represented by the notation $(x_n,\ y_n,\ z_n)$, where $n = 1, 2, \ldots, N$. In this instance, x_n denotes the n^{th} picture sample, $y_n = c,\ c = 0, \ldots C-1$ is the face label, and $z_n = [z_n^1; ..;\ z_n^m]$. For the n^{th} image sample, T stands for the facial landmarks. The m^{th} facial landmark in the n^{th} image sample is set to $z_n^m = 1$ if it is present and $z_n^m = 0$ otherwise. As a result, the dataset in question is represented as $(X, Y, Z) = \{(x_n, y_n,\ z_n),\ n \in \{1, 2, \ldots, N\}\}$. On a scale pyramid, three unconnected steps are implemented, with the previous level's outputs serving as the input for the subsequent stage.

An online proctoring system can determine whether the examinee is acting or whether another person is assisting them. There is no way to identify the faked face. Every face recognized in the frame is taken as genuine, and the remaining steps are carried out. However, there is a good chance that someone will pose as someone else during an examination. The process of facial spoofing involves expressing one's level of confidence through the liveliness of their face. When two or more faces in an image line up, the Face Detection module locates the face and gathers its features.

In addition to additional resources like computers and systems that are placed in the room, examinees may also ask for assistance from other people in the room. If the examinee is continually looking at an angle that is different from the left or right, the proctor will receive an alarm. Head posture estimation is replaced with gaze tracking. When an examinee wears a spectacle, the gaze-tracking system's eye-ball recognition method produces inaccurate findings. Therefore, our method uses an affine rotation matrix for head-pose estimation.

3.5 RESULT AND DISCUSSION

The projected method's results are compared with those of earlier methods in this section. The study's implementation makes use of the Python platform. Thirty percent are used for testing, and seventy percent are used for training. Two versions of the model must be created to train it on a single subgroup and then assess its ability to generalize. The Online Exam Proctoring Dataset is utilized by the implementation.

3.6 Dataset Description

In the OEP dataset, the portions of all training movies that do not contain any instances of cheating are classified as samples of the negative class, while the remaining segments belong to the positive class. The positive cheating samples are separated into three primary groups. Three distinct

individuals with teaching expertise were given access to all of the testing films utilized in our system, along with a Graphical User Interface (GUI) intended to manually record the instances of cheating [33].

3.7 Overall Comparison by Varying the Learning Rate

The suggested UARN model's performance metrics are compared to those of other methods that are currently in use, such as CNN [17], RCNN [18], LSTM [22], and XGBoost [28]. Here is the comparative study in TABLE 2. The performance metrics that we compared in our proposed approach are-

i. **Accuracy-** Accuracy quantifies the frequency with which the model's predictions match the actual results.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN)$$

TP (True Positives): Instances where the model correctly identifies the positive class.

TN (True Negatives): Instances where the model correctly identifies the negative class.

FP (False Positives): Instances where the model mistakenly classifies a negative case as positive.

FN (False Negatives): Instances where the model fails to detect the positive class and classifies it as negative.

ii. **Precision-** Also known as the positive predictive value, precision reflects how often the model's positive predictions are accurate.

$$Precision = TP/(TP + FP)$$

iii. **F-Score-** It assesses a classifier's performance by striking a balance between recall and precision.

$$\text{F1-Score} = (2*(\text{Precision}*\text{Recall}))/((\text{Precision}+\text{Recall}))$$

iv. **Sensitivity-** It is the sum of true positives (TP) and false negatives (FN), is the percentage of correctly detected positive cases (TP) to all actual positive cases.

$$Recall = TP/(TP + FN)$$

v. **Specificity-** The percentage of cases that are accurately classified as not belonging to a specific class is known as specificity.

$$Specificity = TN/(TN + FP)$$

vi. **MCC-** A balanced measure even for imbalanced datasets. MCC ranges from -1 (worst) to +1 (best).

$$MCC = (TP * TN - FP * FN)/\sqrt{((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))}$$

vii. **NPV-** It is a classification metric that measures the proportion of correctly predicted negative cases among all predicted negatives.

$$NPV = TN/(TN + FN)$$

viii. **FPR-** It measures the proportion of actual negative cases that were incorrectly classified as positive.

$$FPR = FP/(FP + TN)$$

ix. **FNR-** It measures the proportion of actual positive cases that were incorrectly classified as negative.

$$FNR = FN/(FN + TP)$$

Table 2: Comparison of performance metrics for Learning rate 70%

Model	Accuracy	Precision	F-score	Sensitivity	Specificity	MCC	NPV	FPR	FNR
Proposed	98.2174	98.3452	98.9634	98.5732	98.8421	98.1597	98.3751	0.0254	0.0214
CNN	96.5481	96.4721	96.5881	96.3547	96.8421	96.7532	96.8521	0.0465	0.0434
RCNN	95.5874	95.8641	95.8844	95.8547	95.8741	95.2584	95.4721	0.0584	0.0547
LSTM	94.2581	94.8421	94.9671	94.8741	94.2583	94.1258	94.5871	0.0652	0.0684
XGBoost	94.8527	94.2317	94.6896	94.8574	94.8542	94.7532	94.8745	0.0614	0.0624

TABLE 2 shows the performance of machine learning models across a range of parameters through a comparative examination of models trained at a learning rate of 70%. The accuracy of the suggested model is 98.2174%, which is better than other models. It also shows great sensitivity (98.5732%) and accuracy (98.3452%). Furthermore, the suggested model exhibits exceptional specificity (98.8421%). The suggested model's Matthews Correlation Coefficient (MCC) is 98.1597

Table 3: Comparison of performance metrics for Learning rate 80%

Model	Accuracy	Precision	F-score	Sensitivity	Specificity	MCC	NPV	FPR	FNR
Proposed	99.2174	99.4452	99.8604	99.7702	99.4421	99.0597	99.2741	0.0142	0.0104
CNN	97.4401	97.7701	97.8922	97.0512	97.8421	97.0512	97.2011	0.0341	0.0312
RCNN	96.8863	96.6611	96.9735	96.9505	96.7703	96.5541	96.9711	0.0434	0.0417
LSTM	95.4571	95.8801	95.8965	95.7701	95.1473	95.2288	95.4864	0.0541	0.0552
XGBoost	95.5507	95.3305	95.9725	95.6612	95.5511	95.671	95.6032	0.0503	0.0465

A comparison of machine learning models trained at an 80% learning rate is shown in TABLE 2, along with an analysis of the model's performance on a range of criteria. The suggested model stands out thanks to its exceptional 99.2174% accuracy. It demonstrates good sensitivity (99.7702%) and accuracy (99.4452%). Additionally, the model shows remarkable specificity (99.4421%). With a 99.0597% MCC, the suggested model is noteworthy.

Comparing the accuracy, Precision, and F-score values in TABLE 2, and TABLE 3, demonstrates how altering the learning rate from 70% to 80% can impact the performance of various machine learning models. Accuracy gains are also seen in the CNN, RCNN, LSTM, and XGBoost models, which obtain accuracies of 97.4401%, 96.8863%, 95.4571%, and 95.5507%, in that order. The measure of accuracy is compared in **FIGURE 2(a)**. Impressively, the suggested model achieves 99.4452% precision. Likewise, precision improvements are also observed for the CNN, RCNN, LSTM, and XGBoost models, which yield precisions of 97.7701%, 96.6611%, 95.8801%, and 95.3305%, in that order. **FIGURE 2(b)** displays a comparison in precision. The proposed model's F-score rises from 98.65% to 99.65% in TABLE 3, for instance. In this way, the 3D CNN model grows from 95.29% to 97.00%, the Bi-LSTM model increases from 95.99% to 96.76%, the DNN model increases from 95.18% to 96.74%, and the F-scores of the LSTM model increase from 96.87% to 97.80%. A comparison of F-Measures is shown in **FIGURE 2(c)**.

Author Contributions:

S.L. developed the research concept, assisted in designing the methodology, and contributed to drafting the manuscript. N.S. oversaw the project and guided throughout.

Conflict of interest:

The authors confirm there are no conflicts of interest that could have affected the integrity of this research.

Funding:

This study did not receive financial support from any funding organization.

Data Availability Statement:

All data generated or analyzed during this study are included in this published article and its supplementary materials. Additional data can be made available upon reasonable request to the corresponding author.

Research Involving Humans and /or Animals:

This study did not involve human participants or animals. If applicable, ethical guidelines and approval processes would have been adhered to as required by institutional and international standards.

4. CONCLUSION

By combining Attentive-Net for accurate feature learning and YOLOv7-SGCN for real-time detection, the Attentive Proctoring System greatly improves the integrity of online exams. Reliability in remote assessments is strengthened by the system's capacity to detect possible infractions with high accuracy, sensitivity, and specificity through multi-modal liveness detection, head pose estimation and emotion identification. However, issues with interpretability, transparency, and user trust are brought up by AI-based proctoring. To tackle this, we suggest incorporating Explainable AI (XAI) methods, like decision-tree-based explanations, attention visualizations, and saliency maps, to offer more lucid insights into behaviors that have been identified. This will decrease false positives and increase trust by assisting educators and students in better understanding the system's conclusions. To improve flexibility in a variety of testing situations, including accommodations for students with disabilities, future improvements will concentrate on adaptive learning models that improve detection over time. To increase the transparency, equity, and efficacy of AI-driven proctoring, it will also be essential to optimize LMS integration, computational efficiency and ethical compliance.

References

[1] Li J, Wu CH. Determinants of Learners Self-Directed Learning and Online Learning Attitudes in Online Learning. Sustainability. 2023;15:9381.

[2] Ballad CA, Labrague LJ, Cayaban AR, Turingan OM, Al Balushi SM. Self□Directed Learning Readiness and Learning Styles Among Omani Nursing Students: Implications for Online Learning During the Covid□19 Pandemic. Nurs Forum. 2022;57:94-103.

[3] Kemp K, Baxa D, Cortes C. Exploration of a Collaborative Self-Directed Learning Model in Medical Education. Med Sci Educ. 2022;32:195-207.

[4] Yang X, Wu D, Yi X, Lee JH, Lee T. iExam: A Novel Online Exam Monitoring and Analysis System Based on Face Detection and Recognition. 2022. ArXiv preprint https://arxiv.org/pdf/2206.13356.

[5] Nurpeisova A, Shaushenova A, Mutalova Z, Ongarbayeva M, Niyazbekova S, et al. Research on the Development of a Proctoring System for Conducting Online Exams in Kazakhstan. Computation. 2023;11:120.

[6] Kochegurova EA, Zateev RP. Hidden Monitoring Based on Keystroke Dynamics in Online Examination System. Program Comput Softw. 2022;48:385-398.

[7] Ababneh KI, Ahmed K, Dedousis E. Predictors of Cheating in Online Exams Among Business Students During the COVID Pandemic: Testing the Theory of Planned Behavior. Int J Manag Educ. 2022;20:100713.

[8] Roa'a M, Aljazaery IA, Alaidi AH. Automated Cheating Detection Based on Video Surveillance in the Examination Classes. Int J Interact Mob Technol. 2022;16:125.

[9] Johri A, Hingle A. Students Technological Ambivalence Toward Online Proctoring and the Need for Responsible Use of Educational Technologies. J Eng Educ. 2023;112:221-242.

[10] Lee K, Fanguy M. Online Exam Proctoring Technologies: Educational Innovation or Deterioration? Brit J Educational Tech. 2022;53:475-490.

[11] Li S, Xie Z, Chiu DK, Ho KK. Sentiment Analysis and Topic Modeling Regarding Online Classes on the Reddit Platform: Educators Versus Learners. Appl Sci. 2023;13:2250.

[12] Aristeidou M, Cross S, Rossade KD, Wood C, Rees T, et. al. Online Exams in Higher Education: Exploring Distance Learning Students' Acceptance and Satisfaction. J Comput Assist Learn. 2024;40:342-359.

[13] Henderson M, Chung J, Awdry R, Ashford C, Bryant M, et al. The Temptation to Cheat in Online Exams: Moving Beyond the Binary Discourse of Cheating and Not Cheating. Int J Educ Integr. 2023;19:21.

[14] Baniamer Z, Muhamed B. Cheating in Online Exams: Motives, Methods, and Ways of Preventing From the Perceptions of Business Students in Bahrain. In: Hamdan A, Hassanien AE, Mescon T, Alareeni B, editors. Technologies, artificial intelligence and the future of learning post-COVID-19: the crucial role of international accreditation. Cham: Springer International Publishing. 2022:267-282.

[15] Komosny D, Rehman SU. A Method for Cheating Indication in Unproctored On-Line Exams. Sensors. 2022;22:654.

[16] Ferdosi BJ, Rahman M, Sakib AM, Helaly T. Modeling and Classification of the Behavioral Patterns of Students Participating in Online Examination. Hum Behav Emerg Technol. 2023;2023:1-19.

[17] Kaddoura S, Gumaei A. Towards Effective and Efficient Online Exam Systems Using Deep Learning-Based Cheating Detection Approach. Intell Syst Appl. 2022;16:200153.

[18] Mahmood F, Arshad J, Ben Othman MT, Hayat MF, Bhatti N, et al. Implementation of an Intelligent Exam Supervision System Using Deep Learning Algorithms. Sensors. 2022;22:6389.

[19] Ndungu JN, Chepsergon AK. Dual Role of AI in Academic Dishonesty and Integrity Management in the Institutions of Higher Learning in Sub-Saharan Africa. J Res Educ Technol. 2024;2:69-79.

[20] Cholakov G, Stoyanova-Doycheva. A Extending Fraud Detection in Students Exams Using AI. TEM J. 2024;13:3068-3078.

[21] Sevnarayan K, Maphoto KB. Exploring the Dark Side of Online Distance Learning: Cheating Behaviours Contributing Factors and Strategies to Enhance the Integrity of Online Assessment. J Acad Ethics. 2024;22:51-70.

[22] Alsabhan W. Student Cheating Detection in Higher Education by Implementing Machine Learning and LSTM Techniques. Sensors. 2023;23:4149.

[23] Yulita IN, Hariz FA, Suryana I, Prabuwono AS. Educational Innovation Faced With COVID-19: Deep Learning for Online Exam Cheating Detection. Educ Sci. 2023;13:194.

[24] Banzon AM, Beever J, Taub M. Facial Expression Recognition in Classrooms: Ethical Considerations and Proposed Guidelines for Affect Detection in Educational Settings. IEEE Trans Affect Comput. 2023;15:93-104.

[25] Nurpeisova A, Shaushenova A, Mutalova Z, Zulpykhar Z, Ongarbayeva M, et al. The Study of Mathematical Models and Algorithms for Face Recognition in Images Using Python in Proctoring System. Computation. 2022;10:136.

[26] Pang D, Wang T, Ge D, Zhang F, Chen J. How to Help Teachers Deal With Students Cheating in Online Examinations: Design and Implementation of International Chinese Online Teaching Test Anti-cheating Monitoring System (OICIE-ACS). Electron Com Res. 2022:1-14.

[27] Saraff S, Tripathi M. Emotional Intelligence: Identifying Emotions From Facial Expressions. J Psychosoc Res. 2022;17:97-106.

[28] Hossen MK, Uddin MS. Attention Monitoring of Students During Online Classes Using XGBoost Classifier. Computers and education. Artif Intell. 2023;5:100191.

[29] Reza AM. Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement. J VLSI Signal Process Syst Signal Image Video Technol. 2004;38:35-44.

[30] Huang JC, Frey BJ. Cumulative Distribution Networks and the Derivative-Sum-Product Algorithm: Models and Inference for Cumulative Distribution Functions on Graphs. J Mach Learn Res. 2011;12:301-348.

[31] Bhatti UA, Tang H, Wu G, Marjan S, Hussain A. Deep Learning With Graph Convolutional Networks: An Overview and Latest Applications in Computational Intelligence. Int J Intell Syst. 2023;2023:8342104.

[32] Healy DM, Rockmore DN, Kostelec PJ, Moore S. FFTs for the 2-Sphere Improvements and Variations. J Fourier Anal Appl. 2003;9:341-385.

[33] http://cvlab.cse.msu.edu/project-OEP.html