Journal of Intelligent Systems and Internet of Things
Vol. 10, No. 02, PP. 102-112, 2023
DOI: https://doi.org/10.54216/JISIoT.100209

Real Time E-learning Students Monitoring for Optimization Facial Landmark Recognition Based on Hybrid Deep Learning Techniques

Shahad salh Ali [*1], Jamila Harbi Al'Ameri [1], Thekra Abbas [1]

[1] College of Science / Computer Science Department, Mustansiriyah University, Baghdad, Iraq
Emails: shahad.salh@uomustansiriyah.edu.iq; dr.jameelahharbi@gmail.com; thekra.abbas@uomustansiriyah.edu.iq
*Corresponding Author: shahad.salh@uomustansiriyah.edu.iq

Abstract

The onset of digital education, propelled by the global COVID-19 crisis, has revolutionized the education sector, presenting unique difficulties, including the crucial task of preserving academic honesty. This study explores the possibilities of computer vision technologies, specifically face recognition and detection systems, to deter dishonest practices in online learning contexts. In this article we aim to construct efficacious strategies that leverage these technologies to track student actions in real-time and alert educators about possible cheating instances. This study presents two innovative models addressing cheating in online learning settings using cutting-edge computer vision techniques. Our initial model is an ensemble learning based face recognition system that blends the functionalities of three different deep learning (DL) structures: VGG, MobileNet, and DenseNet. This ensemble learning approach aims to offset the shortcomings of individual models while amplifying the overall effectiveness. The model's efficiency will be gauged by juxtaposing it with other models and testing its performance against renowned benchmark datasets. Following this, we propose a second model designed for real-time face and cheating detection. This model integrates the FaceMesh model, facial landmarks analysis, and head pose estimation to identify possible cheating behaviors, such as significant shifts from a neutral or forward-facing head position. This model's efficiency will be assessed through testing in simulated cheating scenarios and using authentic data from online learning contexts. Upon testing and validation, our proposed models have shown encouraging outcomes. The ensemble learning model outstripped individual models by attaining a remarkable accuracy rate of 91% through soft voting. Furthermore, the face detection system showcased sturdy abilities in recognizing faces under diverse conditions and accurately pinpointed potential cheating behaviors based on head pose estimation.

Received: April 27, 2023 Revised: July 22, 2023 Accepted: October 17, 2023

Keywords: Face Identification; Face Detection; E-learning; Ensemble Learning; Deep Learning.

1. Introduction

The realm of computer vision research focused on deciphering human behavior through computational methods is the detection, tracking, and understanding of human physical movements. This field delves into the examination of a user's head positioning, indications of eye fatigue, and yawning to interpret their behavior patterns. The landscape of teaching and learning has undergone a substantial shift towards digital platforms in the aftermath of the COVID-19 pandemic [1]. This sudden transition from traditional classrooms to online learning environments brought about a myriad of challenges [2]. A notable issue arising from this shift is the lack of physical presence. When learners are unable to connect personally with their instructors, there is a potential decrease in motivation and interest, which in turn impacts their learning outcomes [3]. The outcome of such disconnect often leads students to either prematurely abandon the online course or discontinue it altogether. Comprehending a learner's engagement level in an online learning environment is, therefore, of critical importance. Consequently, scholars are expending considerable efforts to address these emerging issues in the digital education sphere [4]. Technology plays a pivotal role in this effort. For example, the use of a 3D face mesh and media pipe library can help determine the position of the head's landmark and its relative position to the camera [5]. Furthermore, dlib package can be employed to compute the 2D facial landmarks, predicting signs of user fatigue and yawning [6]. The Eye Aspect Ratio is utilized to anticipate if an eye will exhibit signs of drowsiness once a specific threshold is crossed. Similarly, yawning can be inferred when the gap between the upper and lower lips exceeds a certain value [7]. An integral component of the aforementioned approach is the automatic monitoring of changing facial expressions of online learners, which can offer significant insights into their emotional states. With an effective system in place, these metrics can serve as reliable indicators of students' engagement, alertness, and overall wellbeing during their e-learning journey. These advancements in computer vision technologies pave the way for a more nuanced understanding of student behavior in online learning environments and, in turn, offer potential strategies to enhance their learning experiences.

In the paper, we provide a comprehensive examination of the role of computer vision technologies, particularly face recognition and face detection systems, in upholding academic integrity within online learning environments. We detail an ensemble learning model that synergistically combines three distinct deep learning architectures: VGG, Densenet, and MobileNet. This model's robustness and accuracy in recognizing faces is analyzed and compared to existing models and benchmark datasets. Furthermore, we introduce a face detection model that is designed to detect cheating behaviors by analyzing changes in head position. The model's effectiveness is evaluated using simulated cheating scenarios and real-world data from digital classrooms.

The primary contribution of this paper is the development and evaluation of novel models leveraging computer vision technologies, specifically face recognition and detection systems, for upholding academic integrity in online learning environments. The first of these is an ensemble learning model that combines the strengths of three distinct deep learning architectures: VGG, Densenet, and MobileNet. The second model is a face detection system designed to identify potential instances of cheating through the analysis of changes in head position.

This research proposal aims to develop innovative and effective models utilizing advanced computer vision technologies, specifically face recognition and detection systems, for promoting academic integrity in online learning environments. Section II provides a comprehensive literature review, highlighting previous works and research efforts related to face recognition and detection in educational settings. In Section III, the materials and methods employed in our proposed models are presented. This section outlines the framework of both the face identification and face detection systems. The results and discussion of our research are presented in Section IV. This section evaluates the performance of the developed face identification and detection model. In the conclusion Section V, emphasizes the significance of utilizing advanced computer vision technologies in ensuring academic integrity and fostering a fair and ethical learning environment in online education.

2. Related Work

In the realm of detecting the level of participation, various approaches have been explored, encompassing video, audio, and learner log data. This section provides an overview of previous related works that leverage computer vision techniques, specifically focusing on facial features.

Sharma et al. (2019) [8] introduced a system for eye tracking and emotion analysis, employing two models to identify student participation. The first Convolutional Neural Network (CNN) model was trained on eye images to distinguish between concentrated and distracted states. The second pretrained model, trained on the FER-2013 dataset consisting of grayscale images portraying various emotional states, aimed to analyze students' emotions. The system achieved 92% accuracy in identifying pupils' concentration levels by analyzing their emotions.

Nezami et al. (2019) [9] developed a CNN model to recognize student participation and facial expressions. They constructed their own dataset, the Participate Recognition (ER) dataset, comprising 4,627 annotated images, with 2,290images denoting engaged participation and 2,337 images denoting disengaged participation. The FER-2013 dataset was utilized for facial expression analysis. The participation model achieved a classification accuracy of 72.38%.

Rodriguez et al. (2015) [10] employed a decision tree method to identify student interest based on nonverbal behavior. The dataset comprised data from five students, encompassing 60 incidents divided into three categories: neutral, yes, and no. Five facial characteristics were considered, including the face, eyes, shoulders, mouth, and interest. By employing a random decision tree, the maximum accuracy achieved was 87%.

Thomas et al. (2017) [11] utilized facial analysis to predict student attention levels during the viewing of specific YouTube motivational videos. Data from 10 students were collected while they watched the videos, and their level of participation was assessed. An open source toolkit was used to analyze facial features. The Radial Basis Function SVM classifier, considering gaze and stance features, achieved a classification accuracy of 86% in differentiating between engaged and distracted states.

3. Material and Methods

A. Framework of Face Identification

Our proposed framework utilizes ensemble learning to enhance face identification accuracy by combining multiple models. The system is designed for real-time applications like surveillance, attendance, and security systems. Pre-processing involves resizing images and converting labels to categorical data using LabelEncoder and OneHotEncoder.

The model employs VGG, MobileNet, and DenseNet architectures in a hard and soft voting ensemble learning model. These architectures are chosen for their effectiveness in image identification tasks. Transfer learning is utilized to retrain the last layers of the pre-trained models on our dataset, focusing on face-specific features [12].

The dataset is split into training and validation sets for model training and evaluation. The trained models are combined using hard or soft voting ensemble techniques, reducing variance and improving accuracy.

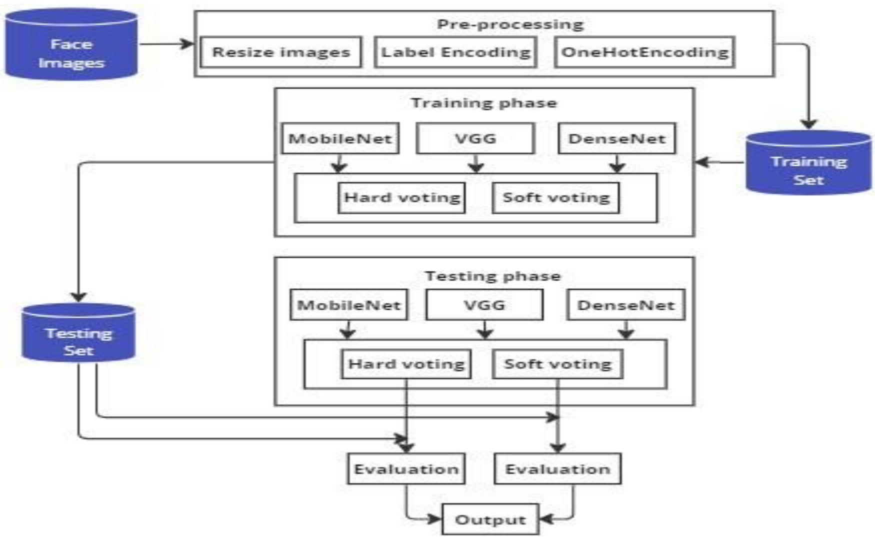

Figure 1: Framework of Face Identification.

1) **Dataset Description:** In this article, we investigate the efficiency of the "Pins Face Recognition" dataset obtained from Kaggle for our study. The dataset comprises 17,534 preprocessed images of 105 celebrities, which were specifically cropped for our research purposes. Figure 1 displays representative facial images from our dataset, providing a visual depiction of the image types utilized in our study. These images showcase ten distinct classes of celebrities that were included in the dataset. To ensure equitable representation across the datasets, we divided the dataset into three sets: a training set, a validation set, and a testing set. Each set contains images from all ten classes, representing different celebrities. The ten classes in our dataset encompass popular and recognizable celebrities, such as Adriana Lima, Alex Lawther, Alexandra Daddario, Alvaro Morte, Amanda Crew, Andy Samberg, Anne Hathaway, Anthony Mackie, Avril Lavigne, and Ben Affleck, who are familiar to students.
Given that the images in the dataset possess varying sizes and resolutions, preprocessing was necessary to make them suitable for our model. It is important to understand the preprocessing steps undertaken to gain insights into how the ML model functions in real-time face recognition.

2) **Data Preprocessing:** In this article, the data preprocessing phase of our e-learning student behavior monitoring system based on ensemble learning for face identification, several crucial steps were implemented. These steps ensured the quality and compatibility of the dataset with our model. We provide an overview of the key pre-processing steps undertaken, including image resizing, label encoding, and onehotencoding techniques.
One of the critical pre-processing steps involved resizing the images in the dataset to a target size of (225,225) pixels [13]. This ensured uniformity and compatibility, facilitating effective model training and evaluation. Resizing the images eliminated variations in dimensions that could introduce bias and hinder the extraction of relevant facial features. The chosen size struck a balance between image clarity and computational efficiency, providing sufficient resolution while maintaining a manageable computational load.

Figure 2: Example of our data file.

Figure 3: Resize images.

The label encoder technique played a significant role in transforming the categorical class labels into numerical representations [14]. This enabled the model to interpret and learn from the labels effectively. Each class label was assigned a unique numerical value, establishing a one-to-one relationship between the original labels and their encoded representations. The label encoder prepared the dataset for subsequent stages such as model training and evaluation, enhancing the compatibility and interpretability of the class labels.

The one-hot encoder technique further transformed the numerical class labels obtained from the label encoding process into a binary format [14]. Each categorical label was converted into a binary vector, representing the presence or absence of a specific class. This encoding scheme allowed the model to effectively handle categorical variables. The resulting one-hot encoded matrix provided a binary representation of the categorical labels, enabling the model to process and analyze the variables accurately during training and prediction phases.

3) **VGG19 Model:** In our article, we adopt the VGG19 architecture, a widely used convolutional neural network (CNN) for image classification tasks. This architecture consists of 16 layers, including 13 convolutional layers and 3 fully connected layers, enabling it to progressively identify intricate features in input images [15]. In our proposed model for face identification in e-learning environments, we leverage the pre-trained VGG19 model with weights obtained from the ImageNet dataset [16]. We modify the model by removing its top layers and introducing custom classification layers tailored to our specific task of recognizing facial landmarks and features. To prevent retraining of the pre-trained layers during our training process, we freeze all layers in the base model except for the last two layers. The output of the VGG19 model is then passed through a Flatten layer, followed by two fully connected Dense layers with 1024 units and ReLU activation functions. To mitigate overfitting, we incorporate a Dropout layer during training. Finally, we add a Dense layer with a softmax activation function to generate the output, which represents the predicted probability distribution for each input image.

4) **Densenet Model**: In our study, we adopt the pre-trained MobileNet model with weights obtained from the ImageNet dataset. To adapt it to our specific task of face identification in e-learning environments, we remove the top layers of the model [17]. Similar to our VGG19 architecture, we freeze all layers in the base model except for the last two layers to prevent

retraining of the pre-trained layers during the training process. This ensures that the valuable knowledge learned from ImageNet remains intact. To tailor the MobileNet model to our desired output, we incorporate custom classification layers. The output of the MobileNet model is passed through a Flatten layer. Following that, we introduce two fully connected Dense layers with 1024 units and ReLU activation functions. Dropout layers are included to mitigate overfitting during training. Finally, we add a Dense layer with a softmax activation function, which generates the final output representing the predicted probability distribution for each input image.

5) **Densenet Model**: The DenseNet architecture is a CNN renowned for its ability to extract increasingly complex features from input images. In our study, we incorporate the pretrained DenseNet121 model with weights obtained from the ImageNet dataset. To tailor it for our specific task of face identification in e-learning environments, we remove the top layers of the model [18]. Similar to our approach with the VGG19 and MobileNet architectures, we freeze all layers in the base model except for the last two layers. This ensures that the pre-trained layers remain unaltered during the training process. Subsequently, we introduce custom classification layers to the DenseNet121 model to suit our desired output. The output of the DenseNet121 model is first passed through a Flatten layer. Next, we include two fully connected Dense layers with 1024 units and ReLU activation functions. To mitigate overfitting, we integrate Dropout layers during training. Lastly, we append a Dense layer with a softmax activation function to generate the final output, representing the predicted probability distribution for each input image.

6) **Ensemble Learning Model:** Ensemble learning combines multiple models to enhance accuracy and mitigate weaknesses. Hard voting selects the most frequent prediction, suitable when models have similar accuracy. Soft voting considers probability values, useful when models vary in accuracy. Our article presents a hard and soft voting model for e-learning behavior monitoring based on face identification.
    - **Hard voting:** Hard voting is a powerful technique that enhances our model’s accuracy and robustness by mitigating individual model weaknesses and biases [19]. Our ensemble approach incorporates three pre-trained models (MobileNet,VGG, and DenseNet), renowned for their excellence in computer vision tasks and achieving state-of-the-art results in image identification benchmarks. By leveraging the strengths of these models, we aim to improve accuracy and reliability in predicting e-learning students’ behavior through face identification. To execute hard voting, we generate class probabilities for each model on the test images. Subsequently, we select the class with the highest probability for each sample from each model. By performing majority voting using the "stats.mode" function from the "scipy" library, we combine the predictions of all three models. This process yields a final prediction for each test sample that surpasses the accuracy and reliability of individual model predictions.
    - **Soft Voting:** In our thesis, we employed the soft voting technique to enhance our model’s performance. Soft voting involves averaging the predicted class probabilities from multiple models and selecting the class with the highest average probability as the final prediction [20]. Using the same three pre-trained models, we predicted the class probabilities for our test images. We then calculated the average probabilities for each class label by averaging the probabilities from each model, represented by the equation (MobileNet + VGG + DenseNet) / 3. This resulted in an array of average probabilities for each class label for every test sample. To obtain the final prediction based on soft voting, we applied the "np.argmax" function to the array of average probabilities. This identified the class with the highest average probability for each test sample. Incorporating soft voting allowed us to leverage the strengths of individual models, enhancing the accuracy and robustness of our model for e-learning student behavior monitoring based on face identification. Soft voting is particularly valuable when models exhibit varying levels of confidence, facilitating a nuanced decision-making process based on probability values.

B. Framework of Face Identification

In our proposed model architecture, we prioritize real-time face detection and cheating detection through facial landmarks. By utilizing the FaceMesh model and employing precise analysis of facial landmarks and head pose estimation, our model offers an effective solution for face identification and the identification of potential cheating behavior. This approach has wide-ranging applications, including e-learning platforms and video proctoring systems.

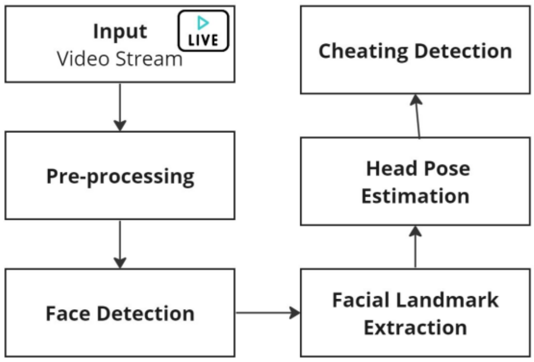

Figure 4: Framework of Face detection.

1) **Video Stream:** In our article, the input to our model is a video stream, which serves as the primary data source for face detection and cheating detection. The video stream consists of sequential frames, representing individual images captured at specific moments in time [21]. These frames are typically obtained from a camera or extracted from a video file.The video stream provides a continuous flow of frames that our model analyzes in real-time. Each frame contains visual information, including images of people's faces captured by the camera [22]. The frame rate determines the frequency at which our model processes the frames. Since our model focuses on face detection and cheating detection, it is crucial for the input video stream to contain individuals' faces. This can be achieved by positioning the camera appropriately or configuring the video source to capture frames with visible faces.
2) **Preprocessing:** In our proposed model, the preprocessing stage is pivotal for preparing input video frames to detect potential cheating behavior. This stage ensures frames are in the appropriate format and orientation required by the model. Initially, real-time frames are captured from a camera, capturing visual information of the monitored individual. Preprocessing steps are then applied to ensure consistency and compatibility. Firstly, frames are converted from BGR to RGB color space using OpenCV's cv2.cvtColor() function, guaranteeing consistent and accurate operations and computations [23]. Next, frames are horizontally flipped using the cv2.flip() function, aligning them with the model's expectations for further processing. This ensures accurate analysis of facial landmarks and head pose by conforming the frame's left and right directions to the model's assumptions. These preprocessing steps are crucial to standardize the input frames in the desired format and orientation for subsequent operations. By converting frames to RGB color space and flipping them horizontally, we ensure compatibility with the FaceMesh model and other analysis techniques employed in our proposed model.
3) **Face Detection:** Accurate and efficient face detection is crucial in our model for real-time identification of potential cheating behavior. We utilize the FaceMesh model from the mediapipe library, which combines DL algorithms and landmark regression techniques to accurately locate and identify faces in video frames [24]. The face detection process begins by capturing real-time video frames, which are then analyzed by the FaceMesh model. To ensure optimal performance, we set a minimum detection confidence threshold of 0.5, filtering out false positives and considering only highly confident face detections. Once a face is detected, the model proceeds to extract 468 facial landmarks, including key points like the nose, eyes, and mouth. These landmarks provide vital spatial information and act as reference points for subsequent analysis, such as facial landmark extraction and head pose estimation. Accurate face detection sets the foundation for reliable and comprehensive analysis of individuals' behavior in real-time video streams.
4) **Facial Landmark Extraction:** In our model, we utilize the FaceMesh model from the mediapipe library for accurate and efficient extraction of facial landmarks. This model detects 468 facial landmarks, including key points like the nose, eyes, mouth, and chin, providing rich information about the face's geometry [25]. By iteratively examining the detected face regions, we select specific landmarks of interest for further analysis, focusing

on those crucial for detecting potential cheating behavior. The 2D coordinates of these selected landmarks are extracted, representing the position within the video frame and providing spatial information about the face's structure. For certain landmarks like the nose, we also computetheir 3D coordinates to capture the 3D aspect of the face. These extracted 2D and 3D facial landmark coordinates are then used for head pose estimation and cheating detection. By analyzing the positions, orientations, and movements of these landmarks, our model can identify potential instances of cheating behavior based on predefined criteria.

5) **Head Pose Estimation:** Head pose estimation is crucial for detecting potential cheating behavior in real-time video streams. It analyzes an individual's head orientation and movement, providing insights into their attention and focus. In our model, head pose estimation is performed using computer vision techniques and the solvePnP function from OpenCV [26]. It calculates rotation and translation vectors based on the 3D coordinates of facial landmarks and their corresponding 2D positions. By mapping the facial landmarks to 3D space, we can determine the head's orientation using rotation angles. Predefined thresholds identify significant deviations, indicating potential cheating behavior [27]. Real-time visualization of head pose results in a line representing the head's direction and textual information overlayed on the frame. This incorporation provides valuable insights into an individual's attention level and potential cheating behavior.
6) **Cheating Detection:** Cheating detection in our model relies on analyzing head pose, particularly rotation angles, to identify significant deviations indicating potential cheating behavior. Threshold values for these angles determine acceptable head movement ranges. Rotation angles (yaw, pitch, roll) are calculated from the rotation matrix obtained using the solvePnP function. If angles exceed thresholds, it suggests excessive head tilt, rotation, or orientation, possibly indicating a lack of attention or engagement. Upon detection, our model classifies the head pose as "Looking Left," "Looking Right," "Looking Up," or "Looking Down," and overlays "CHEATING!!" on the frame for visual alert [28]. Visualization enhances understanding of the detected head pose and cheating behavior, aiding effective monitoring and intervention. Lines on the frame indicate head pose direction based on nose landmark coordinates and rotation angles. Text overlays provide feedback and information about the detected head pose and potential cheating behavior, including the detected direction and numerical values of rotation angles. The output is the real-time visual representation of the frame with overlaid graphical elements and text annotations, facilitating immediate monitoring and intervention.

3. Results and discussion

A. Evaluation of Face Identification

The evaluation of the face identification model is crucial to assess its performance and effectiveness in real-world applications. In our comparative analysis, the accuracy values of different models are presented in Table I to evaluate the performance of the face identification model. Accuracy is a commonly used metric to measure the model's ability to correctly classify instances.

Table 1: Accuracy of Face Identification Models

Models	Accuracy
VGG19	87%
MobileNet	86%
DenseNet	85%
Ensemble Learning with Hard Voting	90%
Ensemble Learning with Soft Voting	91%

The VGG19 model achieved an accuracy of 87%. VGG19 is a deep convolutional neural network known for its complex architecture and capability to learn intricate features from images. The high accuracy of 87% indicates that VGG19 performed well in accurately classifying the instances in the dataset. MobileNet achieved an accuracy of 86%. Despite being a lightweight convolutional neural network designed for efficient inference on mobile and embedded devices, MobileNet demonstrated its effectiveness in image classification tasks by achieving a high accuracy. With an accuracy of 85%, DenseNet, another deep convolutional neural network with a dense connectivity pattern, showcased its ability to promote feature reuse and gradient flow, resulting in successful classification of instances. The Ensemble Learning with Hard Voting model achieved an accuracy of 90%. By combining the predictions of MobileNet, VGG19, and DenseNet using a majority voting scheme, this ensemble model leverages the diverse predictions of individual models to achieve higher accuracy compared to any single model alone. The Ensemble Learning with Soft Voting model attained the highest accuracy of 91%. By incorporating the confidence or probability scores of each model's predictions and combining them using weighted averaging, the soft voting approach further enhances the ensemble's performance, resulting in improved accuracy compared to the hard voting ensemble.

B. Evaluation of Face Detection

Our face detection system for cheating detection demonstrated promising results during testing and validation. It successfully identified faces in various lighting conditions and angles, showcasing its robust performance in real-time face detection and tracking, even in challenging scenarios. The system's head pose estimation, derived from detected facial landmarks, achieved a notable level of accuracy. By utilizing rotation angles (yaw, pitch, and roll), calculated through the conversion of the rotation vector into a rotation matrix using Rodrigues transformation and subsequent matrix decomposition with RQ decomposition, the system accurately estimated the head's orientation in 3D space. These rotation angles played a crucial role in determining the head's tilt, rotation, and orientation. A key feature of the system was its capability to detect potential cheating behavior. By defining predefined threshold values for the rotation angles, the system identified significant deviations from a neutral or forward-facing head position. Abnormal head movements such as excessive head turning or pronounced tilting were successfully flagged as potential indicators of cheating behavior. The system classified these irregular head movements into categories like "Looking Left," "Looking Right," "Looking Up," or "Looking Down," and promptly displayed a "CHEATING!!" alert on the video frame, as illustrated in Figure 5. This real-time alert system proved valuable by enabling swift intervention and maintaining the integrity of the monitored task or activity. Additionally, the graphical representation of the head pose direction, based on the 2D coordinates of the nose and accompanied by text overlays indicating the calculated rotation angles, facilitated a clear and visual understanding of the head pose and potential cheating behavior.

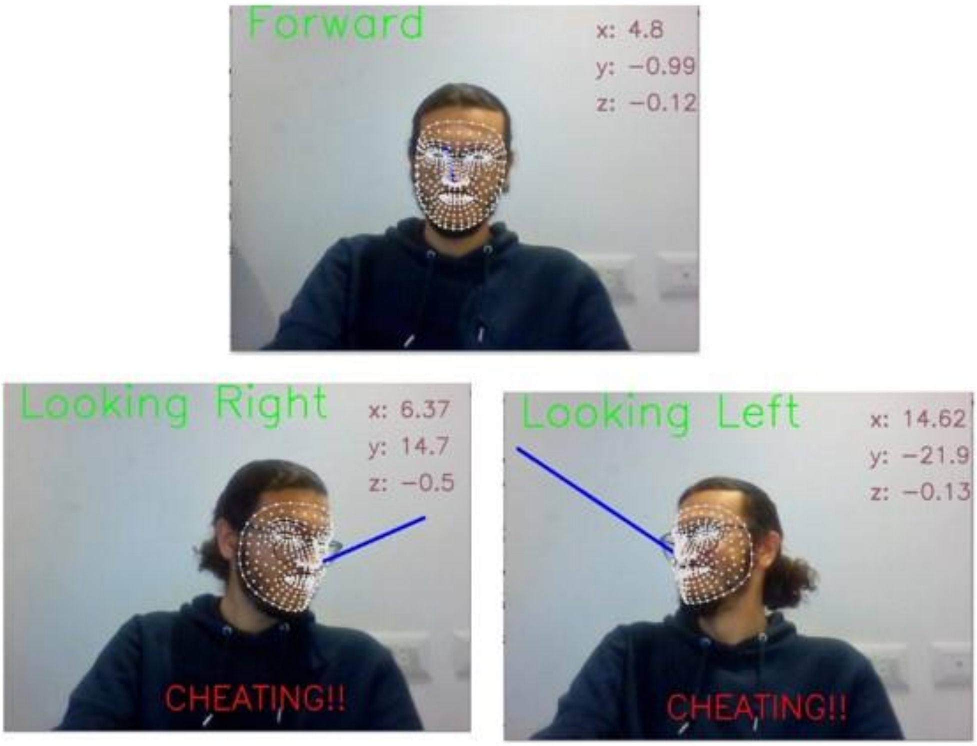

Figure 5: Face cheating detection.

6. Conclusion

This article presents a significant advancement in academic integrity in e-learning environments through the design and implementation of two powerful models using advanced computer vision techniques to prevent cheating. Our ensemble learning model, combining VGG, MobileNet, and DenseNet, improves face identification accuracy and enables effective learner behavior monitoring. Rigorous evaluation against established models and datasets demonstrates the superiority of our approach. Additionally, our real-time face detection and cheating detection model, utilizing FaceMesh, facial landmark analysis, and head pose estimation, excels at identifying suspicious behavior. Evaluation against simulated and real-world cheating scenarios validates its effectiveness in e-learning environments.

Future directions include exploring additional deep learning architectures and techniques to enhance face recognition and cheating detection systems. Integrating other behavior analysis, such as eye movements and body language, can further improve cheating detection. Addressing scalability and efficiency concerns through optimization techniques and distributed computing will enhance the system's applicability in large-scale e-learning platforms.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Arif, Z. H., & Cengiz, K. (2023). Severity Classification for COVID-19 Infections based on Lasso-Logistic Regression Model. International Journal of Mathematics, Statistics, and Computer Science, 1, 25–32. https://doi.org/10.59543/ijmscs.v1i.7715
[2] Colin Conrad, Qi Deng, Isabelle Caron, Oksana Shkurska, Paulette Skerrett, and Binod Sundararajan. How student perceptions about online learning difficulty influenced their satisfaction during canada's covid-19 response. British Journal of Educational Technology, 53(3):534–557, 2022.
[3] Simone CO Conceição. Faculty lived experiences in the online environment. Adult education quarterly, 57(1):26–45, 2006.
[4] Margaret J Cox. Formal to informal learning with it: research challenges and issues for e-learning. Journal of computer assisted learning, 29(1):85–105, 2013.
[5] AL Rohith Reddy, K Anish Gandhi, S Thakrimul Jaffer, and MK Srilekha. Student live behaviour monitoring during virtual class using artificial intelligence. In Journal of Physics: Conference Series, volume 2335, page 012026. IOP Publishing, 2022.
[6] Hao Yang, Li Liu, Weidong Min, Xiaosong Yang, and Xin Xiong. Driver yawning detection based on subtle facial action recognition. IEEE Transactions on Multimedia, 23:572–583, 2020.
[7] Anil Audumbar Pise, Hima Vadapalli, and Ian Sanders. Estimation of learning affects experienced by learners: an approach using relational reasoning and adaptive mapping. Wireless Communications and Mobile Computing, 2022, 2022.
[8] Prabin Sharma, Shubham Joshi, Subash Gautam, Sneha Maharjan, Salik Ram Khanal, Manuel Cabral Reis, João Barroso, and Vítor Manuel de Jesus Filipe. Student engagement detection using emotion analysis, eye tracking and head movement with machine learning. In Technology and Innovation in Learning, Teaching and Education: Third International Conference, TECH-EDU 2022, Lisbon, Portugal, August 31–September 2, 2022, Revised Selected Papers, pages 52–68. Springer, 2023.
[9] Omid Mohamad Nezami, Mark Dras, Len Hamey, Deborah Richards, Stephen Wan, and Cécile Paris. Automatic recognition of student engagement using deep learning and facial expression. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pages 273–289. Springer, 2020.
[10] CA Vázquez Rodríguez, M Mejía Lavalle, and R Pinto Elías. Modeling student engagement by means of nonverbal behavior and decision trees. In 2015 International Conference on Mechatronics, Electronics and Automotive Engineering (ICMEAE), pages 81–85. IEEE, 2015.
[11] Chinchu Thomas and Dinesh Babu Jayagopi. Predicting student engagement in classrooms using facial behavioral cues. In Proceedings of the 1st ACM SIGCHI international workshop on multimodal interaction for education, pages 33–40, 2017.

[12] Brais Martinez, Michel F Valstar, Bihan Jiang, and Maja Pantic. Automatic analysis of facial actions: A survey. IEEE transactions on affective computing, 10(3):325–347, 2017.
[13] Si Si Sann, Swe Swe Win, and Zin Minn Thant. An analysis of various image pre-processing techniques in butterfly image. International Journal for Advance Research and Development, 6(1):1–4, 2021.
[14] John T Hancock and Taghi M Khoshgoftaar. Survey on categorical data for neural networks. Journal of Big Data, 7(1):1–41, 2020.
[15] Anamika Dhillon and Gyanendra K Verma. Convolutional neural network: a review of models, methodologies and applications to object detection. Progress in Artificial Intelligence, 9(2):85–112, 2020.
[16] Miao Tian, Kai Ma, Zhihao Liu, Qinjun Qiu, Yongjian Tan, and Zhong Xie. Recognition of geological legends on a geological profile via an improved deep learning method with augmented data using transfer learning strategies. Ore Geology Reviews, page 105270, 2022.
[17] Andrey V Savchenko, Lyudmila V Savchenko, and Ilya Makarov. Classifying emotions and engagement in online learning based on a single facial expression recognition neural network. IEEE Transactions on Affective Computing, 13(4):2132–2143, 2022.
[18] Yi Zhu and Shawn Newsam. Densenet for dense flow. In 2017 IEEE international conference on image processing (ICIP), pages 790–794. IEEE, 2017.
[19] Eric Bauer and Ron Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Machine learning, 36:105–139, 1999.
[20] Sourya Dipta Das, Ayan Basak, and Saikat Dutta. A heuristic-driven ensemble framework for covid-19 fake news detection. In Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1, pages 164–176. Springer, 2021.
[21] Richard Szeliski. Video mosaics for virtual environments. IEEE computer Graphics and Applications, 16(2):22–30, 1996.
[22] Byoung Chul Ko. A brief review of facial emotion recognition based on visual information. sensors, 18(2):401, 2018.
[23] Michal Podpora, Grzegorz Pawel Korbas, and Aleksandra Kawala-Janik. Yuv vs rgb-choosing a color space for human-machine interaction. In FedCSIS (Position Papers), pages 29–34. Citeseer, 2014.
[24] Yassine Hammadi, François Grondin, François Ferland, and Karina Lebel. Evaluation of various state of the art head pose estimation algorithms for clinical scenarios. Sensors, 22(18):6850, 2022.
[25] Nannan Wang, Xinbo Gao, Dacheng Tao, Heng Yang, and Xuelong Li. Facial feature point detection: A comprehensive survey. Neurocomputing, 275:50–65, 2018.
[26] Nataniel Ruiz, Eunji Chong, and James M Rehg. Fine-grained head pose estimation without keypoints. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pages 2074–2083, 2018.
[27] Ambi Singh and Smita Das. A cheating detection system in online examinations based on the analysis of eye-gaze and head-pose. In THEETAS 2022: Proceedings of The International Conference on Emerging Trends in Artificial Intelligence and Smart Systems, THEETAS 2022, 16-17 April 2022, Jabalpur, India, page 55. European Alliance for Innovation, 2022.
[28] Qiang Ji and Xiaojie Yang. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. Real-time imaging, 8(5):357–377, 2002.