International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Training Convolutional Neural Network with Logistic Regression Model for Facial Recognition to Monitor Attentiveness in Classrooms

Snehal Chaudhary[1], Shrikala Deshmukh[2], Dr.Manisha M. Kasar[3], Gauri Rao[4], Mrs. Anuradha Sagar Nigade[5], Aparna Shankar Patil[6], Dr. Meena Chavan[7]

Submitted: 03/10/2023 **Revised**: 22/11/2023 **Accepted**: 05/12/2023

Abstract: This paper presents a facial recognition system that relies on spectral analysis. Utilizing this technology can enhance classroom security by preventing irregularities such as falsified attendance records or the use of counterfeit identities. The system employs an image recognition algorithm to extract pertinent details from a photograph, subsequently encoding and comparing it with other facial data stored in a database. This image data comprises attributes that highlight distinctions between the system's facial images and those in the image repository. The Facial Recognition System comprises two distinct processing modules: training and recognition. Its efficacy and precision in recognizing individuals were assessed in a high school classroom setting.

Keywords: *Facial recognition, feature classification, frame extraction, deep neural network*

1. Introduction

Facial recognition technology was initially introduced in the 1960s but experienced a slowdown during the AI winter. It has recently witnessed significant advancements, mainly due to the increased capabilities of deep neural networks. This technology has found applications in various domains, including Face ID devices, unlocking mechanisms, public security services, and smart payment systems. One specific facet of facial recognition is the recognition of human emotions, known as facial expression recognition. While humans naturally excel at this, computational approaches have also been developed.

In facial recognition, application faces are compared to a dataset of available facial data to identify human faces. This perspective views facial recognition as a logical extension of technology-based monitoring practices that have been prevalent in schools since the 1990s. However, the primary aim of this report is to highlight concerns regarding the specific implications and potential consequences of implementing facial recognition technology in educational institutions.

The article delves into several social challenges and concerns, particularly addressing how this technology could reshape the nature of schools and education along divisive, authoritarian, and oppressive lines. These concerns arise from ongoing debates among scholars in communication, media, and surveillance studies. Despite some calls for an outright ban on surveillance and monitoring technologies in schools, it is acknowledged that these tools can be applied in ways that are not necessarily harmful but may also fall short of being genuinely beneficial.

Recent advancements in artificial intelligence have paved the way for the adoption of facial recognition technology on college campuses, making it both cost-effective and valuable. This trend suggests that the traditional practice of carrying a physical photo ID on campus may soon become obsolete. Already, smartphone users can effortlessly unlock their devices using facial recognition technology.

In the near future, facial recognition systems in educational institutions could offer real-time classroom analytics based on students' reactions during lectures, potentially leading to enhanced campus security measures. While some may view facial recognition as an intrusive surveillance tool, when used judiciously by instructors and administrators, it has the potential to personalize and empower student data without compromising security. This technology might be leveraged to improve students' retention of course material, reduce the time professors spend on class monitoring, and enhance the overall quality of education.

2. Related Works

In recent years, researchers have made notable strides in

[1] *Assistant professor,Dept of Information Technology, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune*
[2] *Assistant professor,Dept of Information Technology, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune*
[3] *Assistant professor,Dept of Computer Science Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune*
[4] *Assistant professor,Dept of Computer Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune*
[5] *Assistant Professor, Dept of Electronics and Communication Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune*
[6] *Assistant Professor, Dept of Electronics and Communication Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune*
[7] *Assistant professor, Dept of Electronics and Communication Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune*

developing automatic classifiers for facial expressions [1, 2, 3]. Some systems designed for facial expression recognition categorize emotions into groups like happiness, sadness, and anger [4]. Others aim to pinpoint specific muscle movements observable in the face [5], offering a more objective characterization of facial expressions. The Facial Action Coding System (FACS) [6] stands out as the most widely recognized psychological framework for categorizing nearly all facial movements. FACS employs Action Units (AU) to identify distinct facial movements based on their visual manifestations. Each AU represents one of the 46 fundamental elements contributing to observable facial movements or deformations, with expressions often comprising a combination of several AUs [1, 2].

Moreover, significant advancements have been achieved in facial expression detection systems, which encompass Bayesian Networks, Neural Networks, and the multilevel Hidden Markov Model (HMM) [7, 8]. Some of these methods encounter challenges related to recognition speed and timing. Typically, to achieve precise recognition, two or more approaches are integrated, and relevant features are extracted as needed. The effectiveness of each technique heavily relies on pre-processing the images, particularly addressing issues related to illumination and feature extraction.

In their study, Janez and Andrej [13] utilized Kinect technology combined with Machine Learning to automate the detection of student attention in a university classroom. They collected data from an undergraduate class and also developed an algorithm for this purpose.

According to Narayanan Veliyath et al., the learning process is not only influenced by the content taught by the instructor but also by how well students comprehend that information. In a lecture, a attentive student will absorb more knowledge compared to a student who is bored or frustrated [14].

Zhang et al. devised an algorithm for detecting driver sleepiness, using the Karolinska Sleepiness Scale (KSS) as a measure for drowsiness while driving [8]. Their proposed model combines a Mixed Effect Ordered Logit (MOL) model with a Time Cumulative Effect (TCE). In comparative experiments, the MOL-TCE model demonstrated a 62.84 percent higher accuracy than existing models.

In 2018, McDonald et al. developed a contextual system for identifying tiredness in drivers. They integrated the Dynamic Bayesian Network algorithm (DBN) into their approach, which showed a lower rate of false-positive results compared to the current standard, PERCLOS, used for detecting driver drowsiness [9].

Using the wavelet packet transform [10], Phani Krishna et al. created an automatic classification model for detecting driver tiredness. They extracted data from a driver's single-channel Electro-Encephalogram (EEG) signals. The suggested model is capable of real-time sleep analysis with an impressive precision rate of 94.45.

3. Proposed System

This system possesses the capability to automatically recognize a wide range of universal emotions and movements, encompassing emotions such as disgust, fear, happiness, sadness, neutrality, as well as tracking the frequency of nodding, blinking, yawning, and more. In such a system, the process involves analyzing a facial image and generating a computed expression prediction. Figure 1 shows a bried outline of the system.

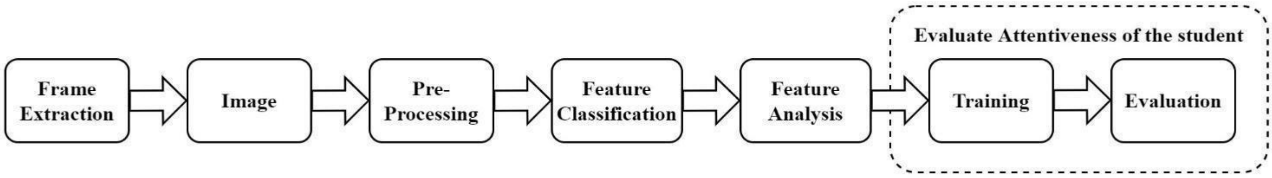


Fig 1: System Architecture for the pro- posed solution

- Frame Extraction: Implement image detection on video content by selecting a set of summary keyframes.
- Pre-Processing: When detecting something in a picture, the first stage is Feature-Based Detection, which resizes and calibrates the raw frames and prepares the raw data for future processing.
- Feature Classification: For effective pattern recognition, choose informative, dis- criminating, and independent features. The purpose of face localization is to use SVM to determine the position of a face in an image.
- Feature Analysis: This module will assist us in examining the feasibility of using human functions as neural network detectors. To recognize the objects, determine the data points and arrange them into their component pieces.
- Training: On the basis of data obtained from internet sources, train the Convolutional Neural Network with Logistic Regression Model.(e.g., Kaggle).
- Evaluation and Output: Displays the ROI that has been identified as well as the categorization result.

The methodology incorporates a module for automated face detection, which is constructed using a training dataset and a neural network. Additionally, neural network architectures can be employed to enhance accuracy. The input image data is fed into the network, which, in turn, generates output layer values serving as the performance metrics for the final model. The highest value from this matrix is computed to indicate the prevailing emotion conveyed by the input data. This process is displayed in the flowchart as seen in figure 2.

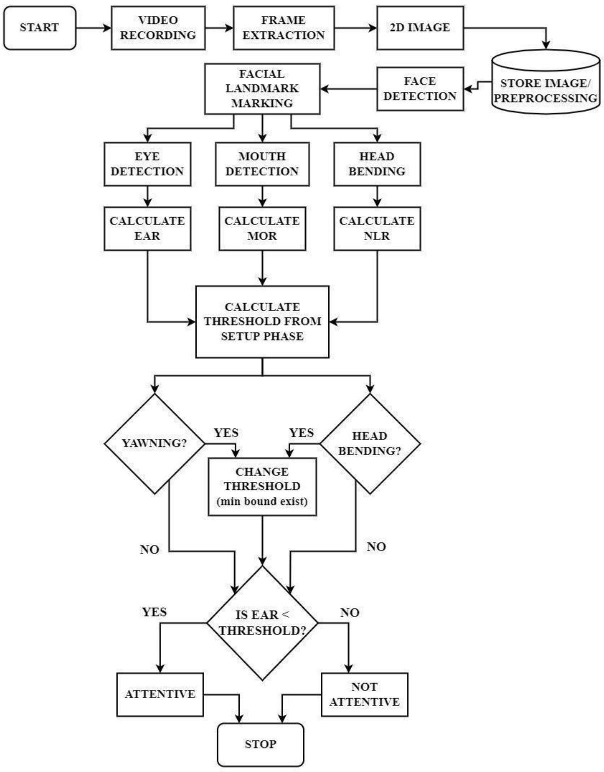


Fig 2: Flowchart of the process that will be followed to design the application

4. Implementation and results

Frame extraction consists of selecting specific keyframes from video content for further analysis. This is achieved by utilizing OpenCV-python to extract particular frames from a video source, which could be either a video file or a live feed from a webcam. These extracted frames are then saved locally for subsequent processing as seen in figure 3a and 3b.

Fig 3a: Sample input file for frame extraction

Fig 3b: Sample frame extracted

Pre-processing has been used to prepare raw data suitable for deep learning and machine learning by assessing its quality, cleaning out noise and inconsistencies, transforming it into suitable formats, and reducing attributes or dimensions as needed.

Feature Classification as shown in figure 4, for the image analysis involved two main approaches: traditional methods that extract image features for classification and advanced methods using deep neural networks (DNNs). OpenCV provides a powerful solution for image classification, utilizing its Deep Neural Network (DNN) module. This module simplified the application of deep learning functions within OpenCV. It enabled transfer learning and the use of pre-trained models, streamlining the process.

By offloading model training to other libraries and utilizing pre-trained models, the implementation became more straightforward. This approach facilitated the adoption of computer vision concepts.

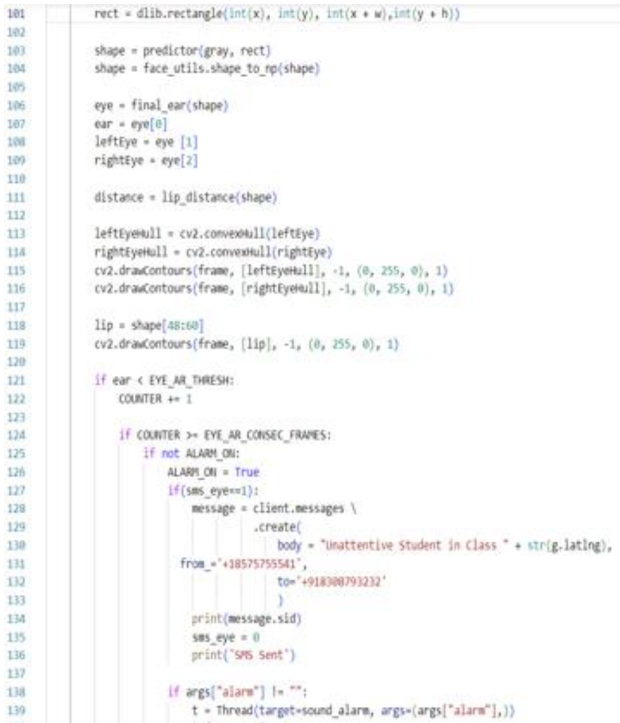


Fig 4: feature classification

Feature Selection involved the technique of streamlining our model's input variables by retaining only pertinent data and eliminating noise as shown in figure 5. This process automatically selected relevant features for our machine learning model by either incorporating or excluding critical features without altering their content. The outcome was a reduction in data noise and a more manageable input dataset. When training the model, we amassed extensive data to enhance the machine's learning process. When a substantial portion of the collected data contained noise, some columns within our dataset did exhibit a limited impact on our model's performance.

```
        if (distance > YAWN_THRESH):
                cv2.putText(frame, "Yawn Alert", (10, 30),
                            cv2.FONT_HERSHEY_SIMPLEX, 0.7, (0, 0, 255), 2)
                if not ALARM_ON:
                    ALARM_ON = True
                    if(sms_yawn==1):
                        message = client.messages \
                                  .create(
                                      body = "Unattentive Student in Class " + str(g.latlng),
                      from_='+18575755541',
                                      to='+918308793232'
                                      )
                        print(message.sid)
                        sms_yawn = 0
                        print('SMS Sent')
                    if args["alarm"] != "":
                        t = Thread(target=sound_alarm, args=(args["alarm"],))
                        t.deamon = True
                        t.start()
        else:
            sms_yawn = 1
            ALARM_ON = False

        cv2.putText(frame, "EAR: {:.2f}".format(ear), (300, 30),
                    cv2.FONT_HERSHEY_SIMPLEX, 0.7, (0, 0, 255), 2)
        cv2.putText(frame, "YAWN: {:.2f}".format(distance), (300, 60),
                    cv2.FONT_HERSHEY_SIMPLEX, 0.7, (0, 0, 255), 2)


    cv2.imshow("Frame", frame)
    key = cv2.waitKey(1) & 0xFF

    if key == ord("q"):
        break

cv2.destroyAllWindows()
vs.stop()
```

Fig 5: Feature selction

Testing and Results

The frames were tested in various scenarios considering the factors diverse lighting conditions, the varying postures and facial orientations of students, and the presence of students who wear eyeglasses. Figure 6 shows successful test when there was ambient light

Fig 6: Student's face and eyes are successfully detected in ambient light

Figure 7 shows the Student's face, eyes , eye blinks, and drowsiness was successfully detected when his face was positioned at the Centre

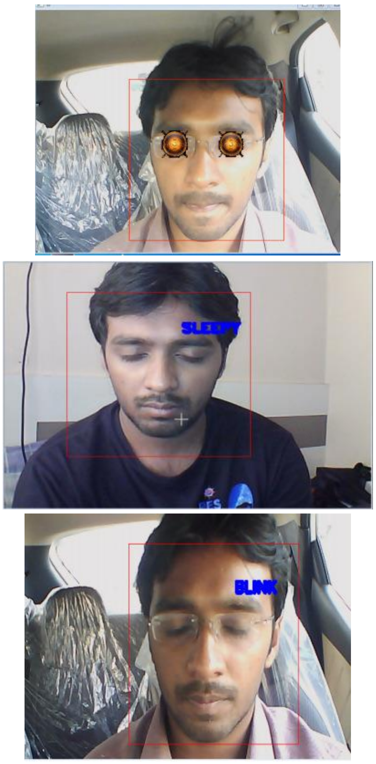


Fig 7: Detecting center positioned face

Figure 8 shows the Student's face, eyes , eye blinks, and drowsiness was successfully detected when his face was positioned at the right side of camera.

Fig 8: Detecting right positioned face

Figure 9 shows the Student's face, eyes , eye blinks, and drowsiness was successfully detected when her face was positioned at the left side of camera.

Fig 9: Detecting left positioned face

We have successfully detected these features even for stduents wearing spectacles. Figure 10 shows the output of this test case

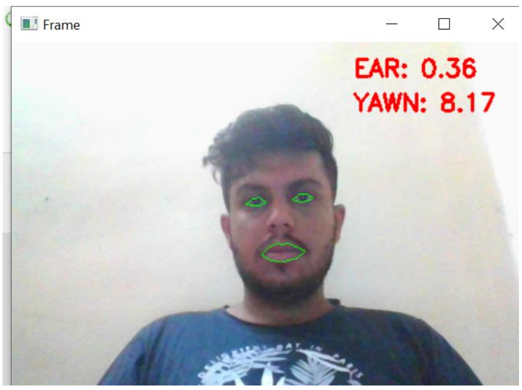
Fig 10 Detecting blink, sleepy and eyes of students with spectacles

The test however has not been entirely successful in identifying the eyes and blink even when the students head is tilted. When face is tilted for more than 30 degrees from vertical plane, it was observed that the detection of face and eyes failed. Figure 11 shows the same.

Fig 11: detection of eyes succesful only for tiklt less than 30degress (middle)

After successful testing in most of the scenarios, the output was incorpoarted with API for text messaging (Via Twilio) so that the detected unattentive students could be notified to concerned faculty.

Figure 12 shows detetcting a drosy or yawning student and its subsequent notification sent as SMS as seen in figure 13.



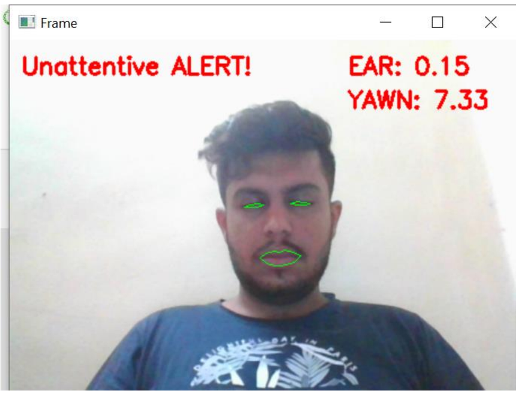


Fig 12 Yawning and drowsy eyes stduent detected

Fig 13 SMS alert for unattentive student

5. Conclusion

In conclusion, we have tried to develop a comprehensive system for enhancing the educational experience in institutions by leveraging facial recognition technology and monitoring facial expressions. The system may be utilized to enhance the overall educational experience, improve attendance tracking, and promote active student involvement while respecting privacy and accessibility for all stakeholders in educational institutions. The system possesses the capability to recognize and match specific faces within a database used for training purposes, which is essential for monitoring individual student attendance and participation. The system is capable of identifying and analyzing a set of facial indicators that signify a student's level of engagement during class. Educational institutions can implement and utilize it without requiring extensive technical proficiency and the system may also be further enhanced to utilize for online meetings to actively encourage student engagement by offering real-time feedback to both students and educators, motivating students to participate and remain attentive during meetings. Our future scope also includes addressing privacy concerns, by incorporating emphasis to safeguard students' personal data and images and include robust privacy protection mechanisms to obtain necessary consent to uphold a secure and respectful learning environment.

References

[1] Y. Tian, T. Kanade, and J. Cohn, “Recog- nizing action units for facial expression analysis”, IEEE Transactions on Pattern Analysis and Ma- chine Intelligence, 23(2), 2001.

[2] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, “Fully automatic facial action recognition in spontaneous behavior”, Proceedings of the IEEE Confer- ence on Automatic Facial and Gesture Recogni tion, 2006.

[3] M. Pantic and J.M. Rothkrantz, “Facial action recognition for facial expression analysis from static face images”, IEEE Transactions on Systems, Man and Cybernetics, 34(3), 2004

[4] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan, “Dynamics of facial expression extracted automatically from video”, Image and Vision Computing, 24(6), 2006.

[5] M.S. Bartlett, G. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, and J.R. Movellan, “Au- tomatic recognition of facial actions in spontaneous expressions”, Journal of Multimedia, 2006.

[6] P. Ekman, W. Friesen, “Facial Action Coding System: A Technique for the Measurement of Facial Movement”, Consulting Psychologists Press, 1978

[7] Cohen, Ira, et al., “Evaluation of expression recognition techniques”, Image and Video Retrieval”, Springer Berlin Heidelberg, 2003. 184- 195.

[8] X. Zhang, X. Wang, X. Yang, C. Xu, X. Zhu, et al., “Driver drowsiness detection using mixed-effect ordered logit model considering time cumulative effect,” Analytic Methods in Accident Research, vol. 26, no. 9, pp. 100114, 2020.

[9] A. D. Mcdonald, J. D. Lee, C. Schwarz, and T. L. Brown, “A contextual and temporal algorithm for drive drowsiness detection,” Accident Analysis Prevention, vol. 113, no. 9, pp. 25–37, 2018.

[10] V. Phani Krishna and S. Chinara, “Au- tomatic classification methods for detecting drowsiness using wavelet packet transform ex- tracted time-domain features from singlechannel EEG signal,” Journal of Neuroscience Methods, vol. 347, no. 3, pp. 108927, 2021.

[11] M. Taherisadr, P. Asnani, S. Galster and O. Dehzangi, “ECG-based driver inatten- tion identification during naturalistic driving us- ing Mel-frequency cepstrum 2-D transform and convolutional neural networks,” Smart Health, vol. 9-10, no. 5, pp. 50–61, 2018.

[12] J. Lee, J. Kim and M. Shin, “Cor- relation analysis

between Electrocardiography (ECG) and Photoplethysmogram (PPG) data for driver’s drowsiness detection using noise replacement method”, Procedia Computer Science, vol. 116, no. 4, pp. 421–426, 2017.

[13] Zaletelj, J., Koˇsir, A., “ Predicting stu- dents’ attention in the classroom from Kinect facial and body features”, Eurasip. J. Image Video Process. 2017(1), 80 (2017)

[14] Veliyath, N., De, P., Allen, A.A., Hodges, C.B., Mitra, A., “Modeling students’ attention in the classroom using eye trackers”, ACMSE 2019 – Proceedings of the 2019 ACM Southeast Conference, pp. 2–9 (2019)

[15] Anand, R., Ahamad, S., Veeraiah, V., Janardan, S.K., Dhabliya, D., Sindhwani, N., Gupta, A. Optimizing 6G wireless network security for effective communication (2023) Innovative Smart Materials Used in Wireless Communication Technology, pp. 1-20.

[16] Soundararajan, R., Stanislaus, P.M., Ramasamy, S.G., Dhabliya, D., Deshpande, V., Sehar, S., Bavirisetti, D.P. Multi-Channel Assessment Policies for Energy-Efficient Data Transmission in Wireless Underground Sensor Networks (2023) Energies, 16 (5), art. no. 2285,