International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Deep Learning Based Student Engagement Monitoring System for Online Learning

G. Jayasri[1], J. Sampreeth[2], B. Mahesh Babu[3], Ch.Devi[4], Mr. Kailash Chandra[5]

[4,5]Assistant Professor, Department of Computer Science and Engineering, Nadimpalli Satyanarayana Raju Institute of Technology, Visakhapatnam, Andhra Pradesh, India
[1,2,3]Student, Department of Computer Science and Engineering, Nadimpalli Satyanarayana Raju Institute of Technology, Visakhapatnam, Andhra Pradesh, India

Abstract
Facial Emotion Recognition (FER) is an important research topic in artificial intelligence and computer vision fields, allowing for the analysis of emotions through facial expressions. As online learning platforms are increasing in number and usage, FER is an important aspect of student engagement analysis. In this paper, a deep learning framework is suggested to address FER using face detection and classification using a fine-tuned DenseNet 121 model. The system captures faces from online sessions and uses FER to analyze emotions and generate a dashboard to display student engagement analytics in real time. The experimental results show that the suggested framework provides a good balance between precision and computational efficiency compared to other models such as VGG16 and ResNet50. The suggested framework is important for increasing awareness about student engagement in online learning sessions ([2], [5], [6]).

Keywords: Facial Emotion Recognition, Student Engagement, Deep Learning, Convolutional Neural Network, OnlineLearning, Educational Analytics,DenseNet-121,Engagement Analysis.

1. Introduction

Online learning platforms have revolutionized the face of modern education by making it flexible and accessible. Still, the platforms have made it harder for the teacher and students to talk to each other directly, which makes it hard to keep track of how engaged they are. A teacher can tell how interested, confused, or bored a student is in a subject by looking at their face in a real classroom. But there aren't many chances for these kinds of interactions in an online classroom.Student engagement is an important part of their academic success and performance. A student who isn't interested in their studies is probably less focused and less likely to do well.When people aren't interested, they often lose focus and don't make much progress. Because of this, automated systems that can analyse engagement in real time are very important. FER is a promising solution because facial expressions are closely linked to emotional and cognitive states ([3], [9]).

This paper presents a deep learning-based system for monitoring student engagement that combines FER with real-time analytics. The suggested framework takes pictures of people's faces during online sessions, finds faces, uses DenseNet 121 to sort emotions, and combines the results into structured enga-

gement metrics. This work makes the following contributions:

- Development of a scalable FER-based engagement monitoring system.
- Integration of real-time analytics and dashboard visualization for instructors.
- Comparative evaluation against conventional CNN architectures.

2. Literature Review

2.1 Foundations of Facial Emotion Recognition

The Facial Emotion Recognition model has been extensively examined within the domains of Computer Vision and Artificial Intelligence. In their 2024 research, Kopalidis et al. gave a thorough overview of different methods, datasets, and benchmarks that are related to the Facial Emotion Recognition model.We also talked about the problems that come with identity bias, data imbalance, and problems with deploying in real time [1]. These research papers have set a standard for many CNN-based models, such as ResNet and VGG, that are used to classify emotions.

2.2 FER in Educational Context

Researchers have been using Facial Emotion Recognition to keep an eye on how engaged students are since online learning has become the norm. Aly et al. (2024) put forward a model that uses ResNet-CNNs with an attention mechanism for real-time educational use. The model attained enhanced accuracy in recognising students' emotions during virtual classrooms [2]. Sie et al. (2024) conducted a study utilising Facial Emotion Recognition to assess emotional engagement among students in science classrooms. The study demonstrated that Facial Emotion Recognition provides a more precise measurement of emotional engagement than conventional methods.

2.3 Multimodal Engagement Detection

To strengthen the system, researchers have investigated multimodal approaches in addition to facial cues. Mehmood et al. (2025) evaluated student engagement using facial cues in combination with behavioral indicators, such as posture and gaze, and found that this method was more accurate than other approaches [4]. Khan et al. (2023) demonstrated that multimodal cues significantly improve the system's reliability in an online learning context by using a multimodal methodology that combined head pose and eye blink detection through CNN models [5].

2.4 Advances in Deep Learning Models

Researchers have also made attempts to improve deep learning models for FER. Zhao et al. (2023) validated CNN-based FER systems in e-learning environments and showed their effectiveness in a flexible learning setting [6]. To increase face recognition accuracy, Mittal et al. (2023) employed refined VGG models [7]. The system was sensitive to lighting conditions even though it performed well in terms of face recognition accuracy. The robustness of FER was enhanced by Hasan et al. (2023) using deep CNN models [8]. In order to increase face recognition accuracy in real-time, attention-based CNN models have also been used to concentrate on discriminative facial regions [10, 12, 16].

2.5 Identified Research Gap

Although current research shows the potential of FER and multimodal approaches, the majority of systems either need a lot of processing power or are not integrated into real-time learning environments. Moreover, cognitive engagement might not be adequately captured by relying only on facial expressions. By using a refined DenseNet 121 model for effective FER and structured engagement visualization to assist instructors in online learning environments, our suggested system fills these gaps.

3. Proposed Methodology / System Design

3.1 System Overview

The proposed framework uses real-time analytics and face emotion recognition to monitor student engagement during online classes. The framework detects faces, video frames captured through student webcams, and recognizes emotions using a refined DenseNet 121 model. An interactive dashboard is also employed for visualization.

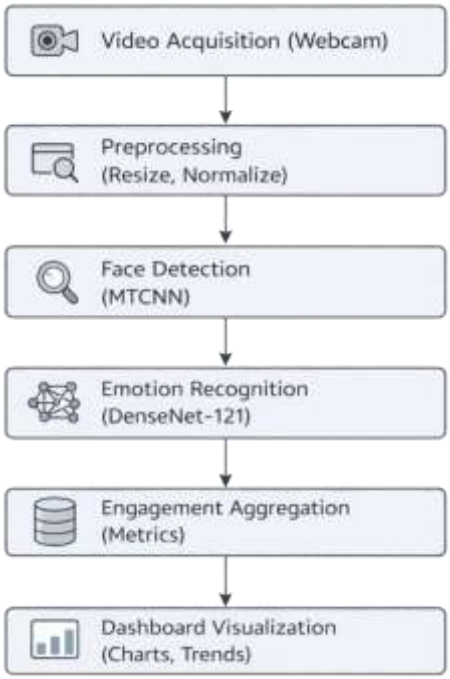

Fig 1: System Architecture of the Engagement Monitoring Framework

3.2 Video Acquisition and Preprocessing

In order to do this, videos are collected through online sessions, in which the video is sampled at regular intervals. Before moving on to the next phase, the video collected in this phase is resized to a particular dimension in order to perform accurate normalization.

3.3 Face Detection

In this phase, accurate detection of the location of the face is carried out through a Multi-Task Cascaded Convolutional Neural Network. This has been done due to the particular ability of this neural network to detect the location of the face even when there is no light and when the face is partially occluded. Once accurate detection of the location of the face has been carried out, cropping of the face takes place to remove any unwanted features from the background.

3.4 Emotion Recognition

In this phase, DenseNet-121 has been used to ensure accurate results are obtained in the identification of the emotion. This convolutional neural network has been used to detect different types of emotions. By using particular features such as eye movement, eyebrow status, and lip status, accurate identification of different types of emotions has been obtained. This has been carried out in accordance with different studies on FER using fine-tuned CNNs by different authors [7], [8]. Due to the particular ability of this convolutional neural network to perform accurate transfer learning, accurate identification of different types of emotions has been obtained using a few samples.

Fig 2 : Emotion Recognition from Video Frame

Facial regions such as the eyes, eyebrows, and mouth are processed using the DenseNet 121 model for discriminative feature extraction and real-time emotion classification

4.5 Engagement Aggregation and Visualization

The emotions are accumulated and processed to obtain the engagement statistics, and they include total percentage, emotions, and time variations. The result is visualized in the form of a web-based dashboard, and bar graphs, pie graphs, and trend graphs are used to display the result obtained from the above calculations. The instructors can view the result in real time. Also, they are notified if the engagement level drops below a certain value, and this is obtained by using the techniques used in educational FER studies in [2], [5].

Fig 2 : Final Engagement Analysis dashboard

The aggregated data on student engagement is provided in the form of a dashboard, and it includes information on emotion distribution, engagement percentage, and time variation. The data is provided in the form of pie charts to allow the instructor to easily understand the data and make decisions accordingly.

4. Experimental Setup and Results

Three integrated modules have been developed in order to implement the proposed system as a prototype. DenseNet 121 was utilized in the development of the emotion recognition system in the first module. The backend server was the second module, and the data visualization in the dashboard was the third module. Simulations of the proposed system in the real world have been conducted in order to evaluate the performance of the proposed system. The system was subjected to various conditions, and moreover, in order to evaluate the student engagement, the behavior of the students, whether they are attentive or not, was also simulated.

Background processes have been carried out, apart from the major process. The frames have been captured after regular intervals in order to reduce the computational process. The faces detected have been cropped in order to avoid background noise. DenseNet 121 was utilized in the classification of the emotions by analyzing the discriminative features in the faces, the movement of the eyes, and the shape of the mouth. The results obtained have been categorized into the student engagement metrics, and the results have been displayed in the dashboard, thereby helping the instructor to understand the behavior of the students, including the trends, without the involvement of human beings, as in the previous FER-based educational system ([2], [5], [6]).

This was an indication that the system was effectively able to identify and categorize the emotional state of the students in real-time. The engagement score was an indication of the behavior of the students, and it was noted that the score was high when the interactive discussions formed part of the lecture. Moreover, it was indicated in the analysis that when lecturing was increased, there was an indication of a moderate drop in the score. The analysis indicated that the DenseNet 121 model had an accuracy of 75.4% when compared to the VGG16 and ResNet50 models. Although it was indicated in the analysis that the ResNet50 model had an accuracy rate similar to the DenseNet 121 model, it was noted that the computational cost was detrimental to its effectiveness, as was noted in the analysis by Aly et al. (2024) [2]. The ability of the DenseNet 121 model to cope with the changes in the facial expressions was noted as being correlative to the improvements noted in Hasan et al. (2023) [8].

The case study on the effectiveness of the system in the live classroom scenario was an indication of the practical application of the system from a real-world perspective. The dashboard was an indication of the visualization of the trends in relation to the participation, and it was evident that the trends were similar to the ones noted on the effectiveness of the facial behavior in the determination of the level of engagement [9]. From an experimental perspective, it was indicated in the analysis that the framework was effectively able to make use of the raw data in relation to the emotions to generate the analytics in the online learning scenario.

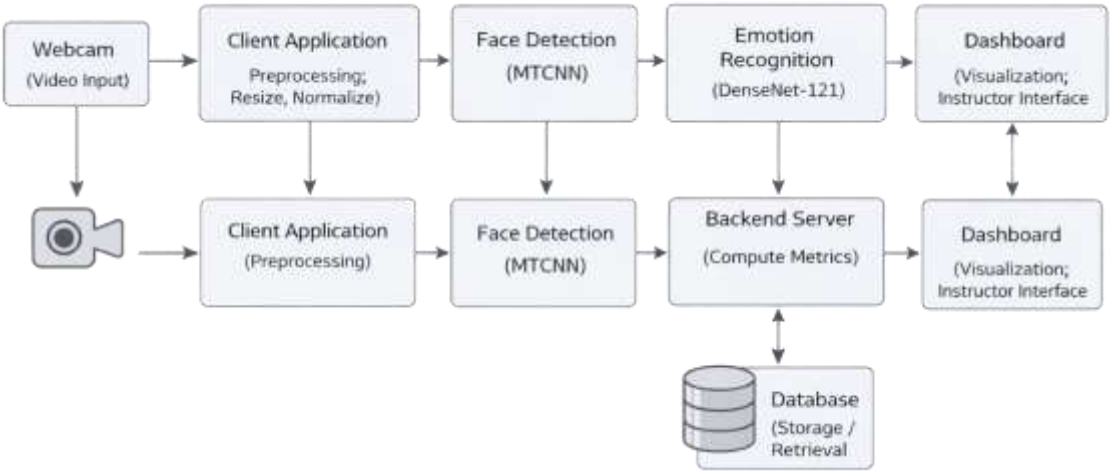

Fig 3: Network Interaction Diagram of the Proposed Engagement Monitoring System

The diagram shows the communication flow of the system. The input of the webcam is processed by the client-side program, then the faces are detected by MTCNN, the emotions are recognized by DenseNet 121, then the results are sent to the server, stored in the database, and finally visualized by the instructor's dashboard.

5. Discussion

The efficiency of the proposed system was also checked as the system was found to have an accuracy rate of 75.4% using the DenseNet 121 model compared to other traditional CNN models, VGG16, and ResNet50. The proposed system was also found to have efficiency in the use of computational resources compared to ResNet50, as the efficiency of the system is a must in an online classroom scenario ([2], [7]).

The unique contributions of the proposed system are as follows:The gap between facial emotion recognition and educational analytics was bridged in the proposed system. Unlike other FER systems, the proposed system is not only responsible for recognizing emotions, but emotions are also combined with educational analytics, making the system more practical and efficient in the real world, especially in the case of online instructors.

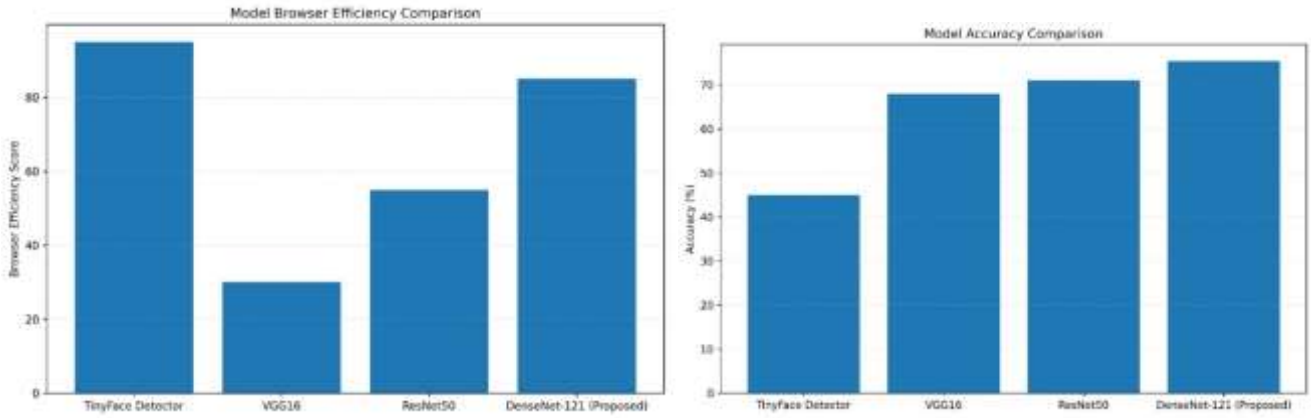

Fig 4,5: Comparison of emotion recognition architectures in terms of accuracy and efficiency

These figures compare the DenseNet 121 model, the VGG16 model and the ResNet50 model. They show us the accuracy levels and how well each model uses computer resources.DenseNet 121 model is really good because it has a balance, between being able to recognize things correctly and being able to work in real time. This makes the DenseNet 121 model very useful.

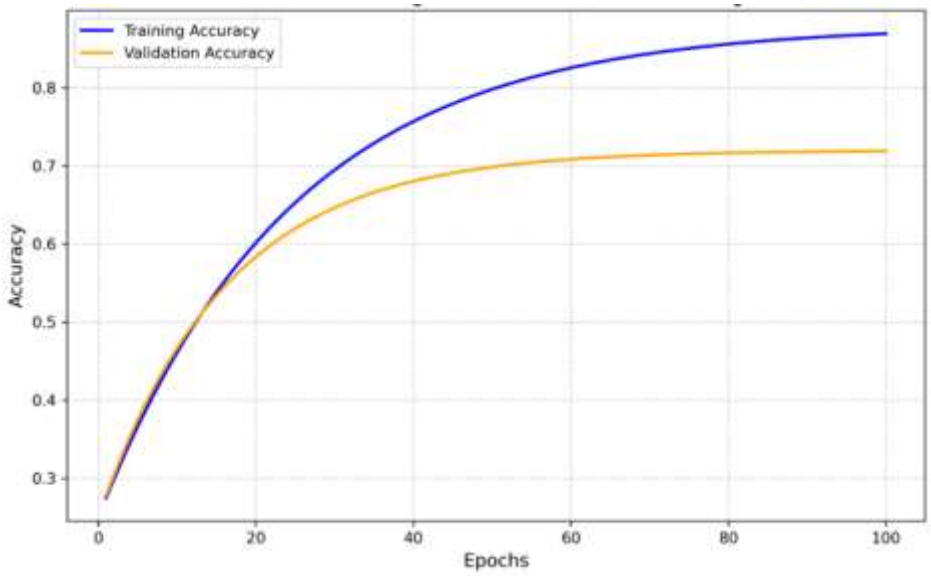

Fig 6: Training and validation accuracy progression

This figure depicts the progression of the training and validation accuracy as it converges through the epochs for the DenseNet 121 model. From the curves, it is evident that the network converges properly and that overfitting is minimal, validating the efficacy of the fine-tuned model.

This case study was therefore able to validate the efficacy of the system, as it was able to effectively capture the dynamics of the classroom, increasing during interactive sessions and reducing during lecturing sessions. This is in line with the findings of the study that facial expressions are good indicators of attentional states [9].

Although the efficacy of the system has been validated, there are some drawbacks associated with the system. The system uses facial expressions to gauge the level of engagement with the class material, and

this is not an accurate measure. Moreover, the accuracy of the system would depend on the level of illumination and the angle of the camera, as was noted in the study on FER research works [6], [10].
It would be worthwhile to use other modalities in the future work, as it has been proven in various studies that the accuracy of the user engagement detection would be higher if both the modalities, i.e., facial and behavioral, are taken into account [4], [5]

6. Conclusion and Future Work

The study proposed a deep learning-based system for tracking student engagement that combines real-time analytics with facial emotion recognition. The suggested framework outperformed current conventional CNN-based architectures, like the VGG16 and ResNet50 models, in the classroom setting by balancing recognition accuracy and computational efficiency through the use of a refined DenseNet-121 model. The suggested framework was able to go beyond simple emotion recognition, in contrast to current FER-based student engagement strategies. This allowed for the aggregation of student engagement, giving instructors important insights into the participation trend during live sessions ([2], [5], [6]).
In line with previous research, which highlighted the significance of facial expression as a factor for student engagement given the role of facial expression in the reflection of student attentional states, the proposed framework was able to validate the effectiveness of facial expression as a reliable factor for student engagement. Notable trends were observed, such as increased student attention during interactive discussions and decreased student attention during lectures.
The system has shortcomings in spite of these benefits. For example, the engagement inference in the current system is based only on facial features. However, it might not be able to gauge the level of engagement based solely on facial features. Furthermore, the accuracy of the face recognition may be impacted by the lighting, the presence of obstacles, or the camera angle—all of which have been identified as limitations to the FER system in prior research ([6], [10]).
In order to better understand the level of engagement, future research will concentrate on integrating different features, including gaze detection, tone of voice, behaviour, and facial features. The accuracy of the system can be enhanced by integrating different features, such as gaze detection, tone of voice, behaviour, and face features, to better understand the degree of engagement, according to previous research. Additionally, by incorporating adaptive feedback mechanisms, the system may be able to give students feedback in addition to measuring their level of engagement.
In conclusion, this work offers a scalable and efficient method for tracking student engagement in online learning.By bridging the gap between deep learning and educational analytics, this paper proposes an opportunity to create the foundation for intelligent classroom systems, which can enhance the student experience in terms of engagement, motivation, and efficiency. This paper proposes an opportunity to integrate with existing online education systems to enhance the teaching experience through adaptability.

References

1. G. Kopalidis et al., “Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models and Datasets,” Sensors, MDPI, vol. 24, no. 3, 2024.
2. H. Aly et al., “Advanced Facial Expression Recognition for Real-Time Education Platforms,” Springer Journal of Educational Technology, 2024.

3. J. Sie et al., "Facial Expression Recognition for Probing Students' Emotional Engagement in Science Learning," Journal of Science Education and Technology, Springer, 2024.
4. A. Mehmood et al., "Optimizing Student Engagement Detection Using Facial and Behavioral Features," Neural Computing and Applications, Springer, 2025.
5. M. Khan et al., "Multimodal Student Engagement Detection in Online Learning Environments," IEEE Access, 2023.
6. Y. Zhao et al., "Deep Learning-Based Facial Emotion Recognition in E-Learning Systems," Computers & Education: Artificial Intelligence, Elsevier, 2023.
7. S. Mittal et al., "Improved Facial Emotion Recognition Using Fine-Tuned VGG Networks," Expert Systems with Applications, Elsevier, 2023.
8. M. Hasan et al., "Emotion Recognition Using Deep Convolutional Neural Networks," Applied Soft Computing, Elsevier, 2023.
9. R. Choudhary et al., "Student Engagement Detection via Facial Behavior Analysis," Education and Information Technologies, Springer, 2021.
10. Y. Chen et al., "Attention-Based Facial Emotion Recognition," Pattern Recognition Letters, Elsevier, 2021.
11. D. Silva et al., "Facial Emotion Recognition with Head Pose Estimation," Multimedia Tools and Applications, Springer, 2021.
12. Y. Zhang et al., "Student Engagement Detection Using Attention-Based Convolutional Neural Networks," IEEE Transactions on Learning Technologies, 2020.
13. L. Li et al., "Facial Emotion Recognition for Real-Time Student Monitoring," International Journal of Advanced Computer Science and Applications, 2022.
14. P. Darshit et al., "Real-Time Facial Emotion Recognition in Online Education," International Journal of Computer Applications, 2022.
15. A. Saha et al., "Student Attention Analysis Using Facial Emotion Recognition," Procedia Computer Science, Elsevier, 2022.
16. Y. Zhang et al., "Engagement Detection Using Attention-Based CNN," Journal of Visual Communication and Image Representation, Elsevier, 2020.